Audio-visual concert performances synchronize an audience's heart rates.

Short title: Audio-visual performances synchronize heart rates.

Anna M. Czepiel*[1,2,3], Lauren K. Fink[2,4,5], Mathias Scharinger[6,7], Christoph Seibert[8], Melanie Wald-Fuhrmann[2,5], & Sonja A. Kotz[3,9]

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

*Corresponding author
Email: a.czepiel@utoronto.ca
Anna M. Czepiel
Department of Psychology
University of Toronto Mississauga
3359 Mississauga Road, Mississauga
Canada
https://orcid.org/0000-0002-7101-945X

Abstract

People enjoy engaging with music. Live music concerts provide an excellent option to investigate real-world music experiences, and at the same time, use neurophysiological synchrony to assess dynamic engagement. In the current study, we assessed engagement in a live concert setting using synchrony of cardiorespiratory measures, comparing inter-subject, stimulus-response, correlation, and phase coherence. As engagement might be enhanced in a concert setting by *seeing* musicians perform, we presented audiences with audio-only (AO) and audio-visual (AV) piano performances. Only correlation measures were above chance level. In comparing time-averaged synchrony across conditions, AV performances evoked higher inter-subject correlation of heart rate (ISC-HR). However, self-reported engagement did not correspond to synchrony when averaged across music pieces. On the other hand, time-resolved analyses show that synchronized deceleration-acceleration heart rate (HR) patterns, typical of an 'orienting response' (an index of directed attention), occurred *within* music pieces at salient events of section boundaries. That is, seeing musicians perform heightened audience engagement at structurally important moments in Western classical music. Overall, we could show that multisensory information shapes dynamic engagement. By comparing different synchrony measures, we further highlight the advantages of timeseries analysis, specifically ISC-HR, as a robust measure of holistic musical listening experiences in naturalistic concert settings.

Keywords: cardiorespiratory synchrony; engagement; inter-subject correlation; stimulus-response correlation; music concerts.

1. Introduction

People enjoy engaging in music, especially in live concerts[1–3]. While concerts provide an excellent option to explore music listening experiences, assessing this experience dynamically is methodologically challenging. One way to explore musical engagement over longer time scales is with synchrony of peripheral responses either across participants or to perceived musical structures. However, many different analysis techniques can be employed. Here, we compared different synchrony analyses to gauge their robustness. Using cardiorespiratory synchrony, the current study assessed engagement, defined as a listener's real-time absorption in music[4–6]. As engagement might be enhanced in a concert setting by *seeing* musicians perform[7–10], we presented participants with audio-visual and audio-only musical performances in a concert context.

Musical engagement can describe either involvement in musical activities (e.g., instrumental practice) or the time-varying mental state when listening to music[11]. Here, we refer to engagement as an absorbed state of heightened attention and affect, which can change on a second-by-second basis while listening to music[5,6,11–13] (for a review, see[5]). Indeed, music is a particular driver of (regular) time-varying attentional focus due to its metrical structure/beat. Although musical engagement can be measured with continuous behavioural responses[12,13], these ratings might distract from directly experiencing music. A promising alternative to measure engagement in a non-distracting way is to assess neural and peripheral physiological responses. One potential index is the orienting response: the engagement of attention, which is reflected in increased skin conductance and deceleration-acceleration patterns of respiration/heart rate[14]. Another method to objectively assess engagement is through physiological synchrony (e.g.,[2,15–18]), which assumes that increased engagement with music is more likely to evoke common time-locked responses across participants[4,19]. Synchrony can be between a participant's response and a stimulus feature that represents how we process/respond to sound, like the auditory envelope (i.e., how sound intensity

changes over time). More recently, spectral flux (spectrum changes over time) has been used as a more faithful representation of how humans respond to auditory stimuli[20–24] as it represents both envelope and rhythmic information[24] and is related to perceived higher activity (e.g., faster music) and fullness of sound[25]. Assessed by calculating the (Euclidean) distance between the spectrum of successive frames, spectral flux can also be assessed within specific frequency bands (sub-bands), which represent different rhythmic elements[26]. For example, in pop music, spectral flux of lower frequency bands can represent kick drum and bass guitar rhythms, while higher frequency bands can represent cymbals and hihat rhythms[27]. Synchronisation of participant responses to stimuli can be assessed using correlation or phase coherence, i.e., rhythmic alignment[4,24]. In this paper, we refer to these as stimulus-response correlation (SRC) and stimulus-response phase coherence (SRPC), respectively. A related synchrony approach is to assess similarity across multiple time-locked participant responses[28]. Similarly to stimulus-response synchrony, inter-subject synchrony can be calculated with correlation[17,29] or using phase coherence; in this paper, we refer to these as inter-subject correlation (ISC) and inter-subject phase coherence (ISPC), respectively.

Growing evidence has related synchrony of neural responses to a participant's engagement with stimuli. In correlational measures, SRC and ISC were related to engagement with movies[17,30,31], speech [32,33], and music[2,4,13,19]. In phase coherence measures, higher neural synchrony occurred in engaging group discussions, indicating a potential marker for shared attention mechanisms[34]. Although phase synchrony in speech was attributed to stimulus intelligibility[35], it was postulated that phase synchrony increases when engaging in stimuli. According to the Dynamic Attending Theory (DAT), internal oscillations adapt to external rhythms so that attending energy is optimised at expected time points[36,37]. Indeed, successful coordination in musical ensembles was related to phase synchrony in neural[38] and heart[39] rhythms (see also[40,41]). Recent work shows coupling of cerebral activity occurs when experiencing similar emotional response to live music concerts[15]. However,

most synchrony research focused exclusively on neural responses, such as those measured via electroencephalography (EEG); more suitable approaches for a live concert audience could be to explore synchrony in peripheral physiological responses[2,16,42].

Recent frameworks propose certain neural mechanisms – such as synchrony – might extend to cardiac, respiratory, gut, and pupil rhythms[43–45]. Promising results show cardiorespiratory ISC related to engagement with narratives and instructional videos[18,46]. In the phase domain, some research indicates that the alignments of breathing/heartbeats to external rhythms aid the processing of upcoming stimulus events[47,48]. In music, respiration has been shown to entrain to musical beats[49,50], although heartbeats do not[51–53]. The current study aimed to further assess cardiorespiratory synchrony as an index of engagement in response to music performances in a concert audience.

Cardiorespiratory synchrony can be time-averaged across stimuli to assess engagement between conditions (e.g., attended/unattended conditions[18]). However, music listening is dynamic, where engagement, neural synchrony[12,13,54], and physiological synchrony[16] change over time. Previous research reported that engagement tends to peak at musical climaxes or structural boundaries[55]. Indeed, it seems beneficial for attentional resources to 'latch onto' structural boundaries: the brain naturally segments incoming information into more meaningful units to process them more efficiently[56,57]. According to the Generative Theory of Tonal Music (GTTM[58]), we 'segment' auditory events based on salient changes when several music features change simultaneously. The ability to segment music at such salient moments has seen consistent empirical support in behavioural ratings[59–62]. However, consciously rating engagement continuously distracts from a natural music listening experience. Recent research showed that salient musical section boundaries evoke neurophysiological synchrony, potentially reflecting heightened engagement[12,13,18]. Therefore, to

complement the current investigation of engagement during audio-only (AO) and audio-visual (AV) music performance on a global level (time-averaged across pieces), we also tested time-varying engagement – indexed by physiological synchrony – between AO and AV music performances during salient moments, defined as section boundaries.

The current study

Audiences might find concerts engaging due to not only listening to music but also seeing a musician perform[8,10]. To test this, concert audiences were presented with AO and AV performances of Western classical piano pieces. Musical pieces were selected as they represent a range of genres (Baroque, Classical-Romantic, 20th century music), and characters, as well as range of tempi, harmonic and musical phrase structures. Presenting such a musical range was thought to evoke high engagement levels. Additionally, pieces were chosen as they may be considered typical pieces heard in live Western classical performances, though not known to the audience, thereby reducing familiarity effects. Based on Kaneshiro et al.[4], dynamic engagement was assessed by calculating cardiorespiratory synchrony using both correlation and phase coherence. For correlation, we assessed heart and respiration rate, e.g., speed of breathing and heartbeats in terms of beats per minute (BPM). For phase coherence, we assessed the phase alignment of heartbeat cycles (i.e., whether heartbeats align in time) and breathing cycles (i.e., whether breathing aligns in time). Synchrony was then assessed both across (i.e., time-averaged synchrony) and within (i.e., time-resolved synchrony) musical pieces. First, we tested that time-locked responses were synchronized above chance level. Second, we hypothesised that time-averaged synchrony would be higher in AV than in AO musical performances. Third, we expected time-resolved synchrony to be higher at section boundaries (based on previous literature[12,13,16]). Further, we expect synchrony to be higher at section boundaries in AV than AO at salient moments in the music (interaction effect), based on literature

148 showing that visual information enhances perceptual responses[63] and that visual imagery heightens
149 engagement[55].
150

151 ## 2. Method

152 Participants, stimuli, and procedures are identical to Czepiel et al.[7] (see General Methods and
153 Experiment 2). Key details of the procedure are outlined below. Data and code are available on Open
154 Science Framework (OSF)[1].
155

156 *Participants*

157 The study was approved by the Ethics Council of the Max Planck Society and in accordance with the
158 Declaration of Helsinki. Participants gave their written informed consent. Twenty-six participants
159 attended the concerts. One participant was excluded due to missing physiological data, thus
160 behavioural and physiological data of twenty-five participants were analysed. Participants included
161 nine females, mean age of 51.08 years (SD = 15.48), who on average had 6.18 (SD = 8.20) years of
162 music lessons and attended on average 13 concerts per year (M = 19.97, SD = 20).
163

164 *Stimuli*

165 Participants were presented with AV and AO versions (see Figure 1a) of three Western classical piano
166 pieces: Johann Sebastian Bach: Prelude and Fugue in D major (Book Two from the Well-Tempered
167 Clavier, BWV 874), Ludwig van Beethoven: Sonata No. 7, Op. 10, No. 3, second movement (Largo e
168 mesto), and Olivier Messiaen: *Regard de l'Esprit de joie* (No. 10 from *Vingt Regards sur l'enfant Jésus*).
169 Both versions were performed by the same professional pianist, on the same piano, in the same

[1] Please note this repository is currently private and anonymous while this manuscript is under review; it will be made public upon acceptance of this manuscript.

concert hall. AO versions were recorded prior to the concert and were presented as recordings through high-quality loudspeakers, while AV versions were performed live by the pianist during the concert. A trained sound engineer checked that the sound level was equal across all stimuli; additional analyses of acoustic features showed these performances were comparable across concerts[7].

Procedure

Participants were invited to attend one of two concerts. Electrocardiography (ECG) and respiration were collected using gelled self-adhesive Ag/AgCl electrodes and a respiration belt, respectively, and continuously measured during the concert at a 1000 Hz sampling rate. During the concert, audiences heard the three pieces both in AV and AO modalities. The order of modality was counterbalanced across the two concerts. After each piece, participants rated items such as engagement. We used a German word to reflect engagement as an absorbed state (see Introduction). Although it was translated as 'absorption' in a previous study[7], it can also be used synonymously with engagement (see also[64]).

Data Analysis

Musical analysis: spectral flux and section boundaries

To model how humans respond to music, previous studies have extracted the envelope from the acoustic signal[4]. More recently, spectral flux has shown to be a better predictor of participants' responses to an acoustic signal as it shares information of the acoustic envelope as well as note/speech onset information[23,24,65]. Therefore, the continuous spectral flux signal was obtained using the MIRToolbox[26] in Matlab (2019b), with a frame size of 25ms (as is appropriate for short-term acoustic features[66]), with a 50% overlap (Figure 1b). To allow for comparison between conditions, we confirmed that spectral flux was similar across AV and AO performances, with Pearson

correlations at $r > 0.92$ (see Supplementary Figure 1). To investigate the phase relationship between spectral flux against the cardiac and respiration measures fairly, we assessed spectral flux within the audience members' average heart (1.01 Hz) and respiration (0.30 Hz) frequencies. This approach followed previous research that also selected spectral flux within specific frequencies (sub-bands) to better ascertain rhythm hierarchy levels[25,27]. Although it is possible in the MIRToolbox to specify these sub-bands, the default sub-bands were considered too wide for the current hypotheses of heart/respiration frequencies. Therefore, to obtain rhythmic levels that coincide with heartbeat and respiration rhythms, spectral flux was bandpass filtered at representative ranges of the interquartile range of all heart (0.951 - 1.159 Hz) and respiration (0.268-0.358 Hz) frequencies (see Supplementary Figure 2). The phase angle of filtered spectral flux was calculated from the real part of the Hilbert envelope using MATLAB's *angle* function (Figure 1b, lower panel). The section boundaries in the musical pieces were identified either by a double bar line or end repeat bar line in the score, or by a change/repeat of thematic material (identified in the score by AMC, then confirmed by a music theorist, see Supplementary Table 1).

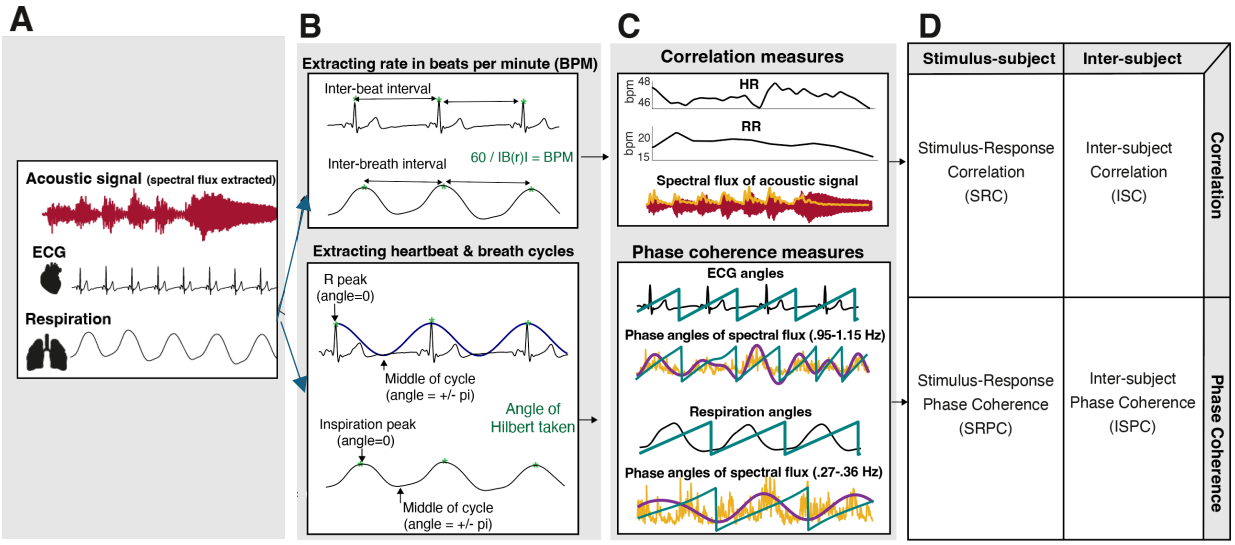

Figure 1. Outline of the experimental design and analysis pipelines. **Panel A** shows the study design: audiences were presented with music pieces in audio-visual (AV, orange box) and audio-only (AO, blue box) conditions, while heart (ECG), respiration, and acoustic signal of the music (maroon) were

continuously recorded. **Panel B** shows how beats per minute (BPM) were extracted from the ECG and respiration signals, which were then used for correlational measures (upper panel). **Panel B** also shows that the angles of heartbeat and respiration cycles were extracted for phase coherence measures, where start of cycle began at peaks of ECG and respiration signals (peaks marked as green dots) (lower panel). **Panel C** shows the correlational measures (upper panel) and the phase-coherence measures (lower panel) we extracted from the continuous measures. For the correlational measures, we extracted heart and respiration rate (HR, RR, from cardiorespiratory signals) and spectral flux (from acoustic signal). For the phase coherence measures, we extracted the phase angle of ECG and respiration cycles (from cardiorespiratory signals), and angles of spectral flux band-pass filtered at the audience's heart frequencies (0.951 - 1.159 Hz) and respiration frequencies (0.268-0.358 Hz). **Panel D** shows the extracted synchrony measures: stimulus-subject (vertical, left) and inter-subject (vertical, right) for correlation (horizontal, top) and phase coherence (horizontal, bottom) measures.

Pre-processing of physiological data.

Raw heart and respiration signals were pre-processed in MatLab 2019b. Missing data from the raw signals were first interpolated at the original sampling rate (all gaps in data were less than 60 ms). Data were cut per piece and further pre-processed using Fieldtrip[67] and the *biosig* toolbox (http://biosig.sourceforge.net). 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. Peaks in ECG and respiration signals were extracted using, respectively, *nqrsdetect* function from biosignal and a custom-made script that located when the respiration signal exceeded a peak threshold. Computationally identified peaks were manually screened to ensure correct identification of R wave (part of the Q-, R- and S-wave [QRS] complex of ECG) and respiration peak locations. Any peaks that were not correctly identified were manually added, while falsely identified peaks were removed, for

238 example, if a T wave in the ECG was accidentally identified as a R wave. In rare cases where
239 participant data within a piece were too noisy to identify clear QRS and inspiration peaks, participant
240 data for this piece were rejected from further analysis (ECG = 14%, respiration = 7%). The pre-
241 processed heart (ECG) and respiration signals were then further analysed in two different ways to
242 calculate synchrony: heart *rate* and respiration *rate* (BPM) were calculated for correlational
243 synchrony, while *phase angle* of ECG (phase of heartbeat cycle) and respiration (phase of
244 inspiration/exhalation cycle) were calculated for phase synchrony (more details outlined below, see
245 Figure 1A-B).
246

247 *Synchrony analyses*

248 Four continuous synchrony measures were extracted: stimulus-response correlation (SRC), inter-
249 subject correlation (ISC), stimulus-responses phase coherence (SRPC), and inter-subject phase
250 coherence (ISPC) for each ECG and respiration (see Fig. 1D).
251

252 *Correlational synchrony: ISC and SRC*

253 Correlational synchrony measures (SRC, ISC) were calculated using heart rate (HR) and respiration
254 rate (RR). To do so, the heart (ECG) and respiration signals converted into rate (beats per minute,
255 BPM) by first obtaining the differential timing between peaks, i.e., inter-beat intervals (IBI) for ECG,
256 and inter-breath intervals-(IBrI) for respiration (see Figure 1B). IB(r)Is were then converted to BPM
257 and interpolated at the original sampling rate to obtain instantaneous HR and RR, which were
258 downsampled to 20 Hz[68]. SRC and ISC were then calculated across sliding windows. Although ISC and
259 SRC measures have previously been calculated within 5 second sliding-windows for EEG[4,12],
260 peripheral responses of heart and respiration are slower, where a typical evoked response can be up
261 to 10 seconds[14,16,69]. Therefore, we chose a 10 second sliding-time window with a one second overlap.
262 For SRC, each participants' HR and RR signal was correlated with the extracted spectral flux. For both

HR and RR signals, $p \times t$ matrices were created, where p was the HR or RR signal for each participant over t timepoints. Signals were correlated in a pairwise fashion, i.e., between all possible participant pairs within one concert. Values underwent Fisher-Z transformation, were averaged, and then transformed back to r values (inverse Fisher-Z).

Phase coherence synchrony: ISPC and SRPC

Phase coherence measures were used to assess whether heartbeats and breaths aligned with the spectral flux rhythms in music. For this, heart and respiration signals were transformed into the phase domain. Here, we assessed changes of actual heartbeat and breath cycles that vary over time. To account for these variations, phase was calculated from cycle peaks, see Figure 1B, green dots marking the R peaks in the heart signals and inspiration peaks in the respiration signals. For the phase angle of the heartbeat, a value of 0 denoted the R wave (peak of heartbeat cycle) and $\pm\pi$ the midpoint of the cycle (see Figure 1B). For the phase angle of the respiration cycle, 0 denoted the inspiration peak, while $\pm\pi$ represents the midpoint of the cycle or maximum exhalation (see Figure 1B). To obtain such values, a continuous sinusoidal wave was fitted to each IB(r)I of detrended and normalised ECG and respiration signals based on[70] , using the following equation:

$$A\ sin\ (2\ \pi\ f_{(k)} t_{(k)} + \theta)$$

where A is the mean peak amplitude (i.e., average amplitude of signal at time points of QRS/respiration peak), $f_{(k)}$ is the frequency calculated from IBI (i.e., $peak_{k+1}$-$peak_k$), converted to Hz ($\frac{1}{IBI(s)}$). Phase (θ) was optimised so the peak of the sine wave corresponded to the ECG/respiration peak. Next, the phase angle was calculated from the real part of the Hilbert using MATLAB's *angle* function.

286 SRPC of heartbeats was calculated by assessing coherence between the phase angles of heartbeat
287 cycles with the spectral flux phase angles corresponding to heart (bandpass filtered at 0.951 - 1.159
288 Hz), while SRPC of respiration was calculated by assessing coherence between phase angles of
289 respiration cycles with spectral flux phase angles corresponding to respiration (bandpass filtered
290 0.268-0.358 Hz) frequencies (see Figure 1C). Coherence was calculated based on the following
291 formula based on[71]:

292 $$\left| n^{-1} \sum_{r=1}^{n} e^{ik_{tr}} \right|$$

293
294 where n is the number of phase signals, e^{ik} is Eulers formula (i.e., the complex polar representation
295 of phase angle k for signal r at time point t). ISPC was assessed in a pairwise fashion between
296 heartbeat cycle and respiration cycle phase angles of all possible participant pairs within one concert
297 using the same coherence formula as above.
298
299 *Adjusting for lags*
300 To account for lags in responses to the stimuli for SRC and SRPC, we adjusted data with a constant
301 lag to optimally align the stimulus and the corresponding responses. This lag was informed by
302 previous literature that added lags to account for slight delays in stimulus responses[4,72–74]. To do this,
303 we first assessed lags on an individual basis. Lags between a stimulus and each participant's response
304 were calculated within the first 10 seconds after stimulus onset as this initial onset likely evokes the
305 most reliable responses. We calculated SRC and SRPC at lags up to each individual's full cardiac and
306 respiration cycle (based on[72]); that is, their mean inter-beat-interval for heart measures (on average:
307 990 ms) or inter-breath interval for respiration measures (on average: 3400 ms), as it might take up
308 to one heartbeat/respiration cycle to begin responding to a stimulus[14]. The positive correlations and
309 phase coherence values were obtained at the optimal lag[73]. We then averaged the individual optimum
310 lags[74]. This average lag was applied for every participant (i.e., each participant had the same lag for

each stimulus-response synchrony measure). For correlation values, the average optimal lags for HR and RR were 579 ms and 1573 ms, respectively. For phase coherence values, the average optimal lag for heart and respiration phase cycles were 500 ms and 1820 ms, respectively. These lags were applied to all corresponding heart and respiration stimulus-response pairings for the correlation and phase coherence. Therefore, for each stimulus-response synchrony measure, all participants had the same lag.

Synchrony across (time-averaged) and within (time-resolved) pieces

Continuous synchrony measures for correlational measures (SRC-HR, ISC-HR, SRC-RR, ISC-RR)) and phase measures (SRPC and ISPC of heartbeats, and SRPC and ISPC for respiration phase) were calculated in a time-averaged or time-resolved fashion to test the different hypotheses. For the first two hypotheses, we calculated time-averaged synchrony values. Rather than averaging synchrony across the three music pieces - as three musical pieces were thought to yield too few observations - we obtained time-averaged synchrony per piece section. The Bach piece was split into seven sections, the Beethoven piece was split into nine sections, and the Messiaen piece was split into nine sections (for details see Supplementary Table 1). Synchrony was then averaged per piece section, yielding 25 observations per participant, per condition, and per synchrony measure. Averaging across musical sections was to account for naturalistic grouping as defined by the musical composition. As an additional check, we also calculated average synchrony within 30 second bins, thereby controlling for length and section changes (yielding 53 observations per participant, per condition, per synchrony measure, see Supplementary Materials), and ran identical statistics. To address the third hypothesis, we calculated time-resolved synchrony values by cutting epochs ±15 seconds relative to section boundary onsets, to capture event-related respiration and heart responses[14,69], as well as any anticipatory effects at musical events[75].

Significance-testing

Regarding the first hypothesis, we tested the assumption that responses time-locked to a stimulus should evoke similar responses across participants[28]. Therefore, we created time-'unlocked' data by circular shifting ECG and respiration signals. This was done 1000 times and time-averaged synchrony measures were calculated as described above. For each of the permutated data sets, we calculated a *t*-test statistic, creating a null distribution of permutation values. By assessing how many of these permutation values were higher than or equal to the statistic of the true (time-locked) data, we determined the chance of true synchrony (see Supplementary Figure 3). Values were corrected for multiple comparisons using false discovery rate (FDR[76]).

AV versus AO: time-averaged synchrony

For the second hypothesis, time-averaged synchrony measures were compared between AV and AO. Self-reported engagement was also compared between these conditions. Linear mixed models (LMMs) with a fixed effect of modality were constructed for each of the synchrony measures as the outcome variables. As the experimental design meant that data were clustered, the following random effects were added to account for this non-independence: random intercepts for concert, piece, and participants, where participants were nested within concerts, while participant and piece were considered crossed effects. A random slope for participants was also included. This random-effect structure represents maximal models[77,78]. If these maximal models generated convergence and/or singularity fit errors, models were simplified following recommendations[77,79]. As output from models generating errors should not to be trusted[77], we report simplified, error-free models in the results. Maximal model outputs can be nonetheless found in Supplementary Materials, while all LMMS can be found in accompanying code on OSF (see link/Footnote (1) above). LMMs were run using *lmer* from the *lme4* packages[80,81]. Significance values and effect sizes were obtained from the *tab_model* function from *sjPlot* package[82]. As a recommended sanity check, we also ran conventional *t*-tests

361 alongside LMMs[83]. This was done by a one sample *t*-test (or Wilcoxon if data were not normally
362 distributed), run on the difference between AO and AV values.
363
364 *AV versus AO: epoched*
365 To test the third hypothesis, synchrony at section boundaries was assessed between modalities
366 across time windows. Synchrony values in the 30-second epochs centred around section boundaries
367 were split into smaller time windows[69,84], yielding five 6-second time windows. For each synchrony
368 measure, linear mixed models (LMMs) with fixed effects of modality and time window were
369 constructed. We added random intercepts for concert, piece, and participants (participants nested
370 within concerts; participant and piece as crossed effects) and a random slope for participants. As
371 above, models generating errors were simplified; error-free models are reported (full models in
372 Supplementary Materials). Estimated marginal means (using *emmeans* package[85]) were used to
373 check pairwise comparisons with Bonferroni adjustments. As we predicted that synchrony would
374 occur at section boundaries, we expected the time window centred at the section boundary to have
375 significantly higher synchrony values than time windows prior to and after a section boundary. In
376 case synchrony was high, we investigated what responses were becoming synchronized. Thus, HR
377 and RR data at section boundaries were taken (i.e., HR and RR as in Figure 1, Panel B). Two LMMs
378 were constructed, each for aggregated HR and RR values, with fixed effects of time window and
379 random effects of participant id, piece, and concert, as above.
380

381 ## 3. Results

382 **Significance of synchrony measures**

383 In assessing the significance of synchrony measures using a permutation approach and FDR
384 correction, we found that SRC-HR ($p < .001$), ISC-HR ($p = .048$), and ISC-RR ($p = .048$) showed

significant, above-chance synchrony (see Supplementary Figure 3). SRC-RR (p = .408), SRPC for heartbeats (p = .412) and respiration (p = .594), and ISPC for heartbeats (p = .594) and respiration (p = .594) were not above chance level. When synchrony values were averaged in 30 second bins, significant results were replicated for SRC-HR (p < .001), but not for ISC-RR (p = .051) and ISC-HR (p = .065). We therefore carried through SRC-HR, and (tentatively) ISC-RR and ISC-HR measures for further modality comparisons.

Time-averaged synchrony between modalities

Figure 2 shows that compared to the AO condition, the AV condition evoked generally higher SRC-HR, ISC-HR, and ISC-RR, but only for ISC-HR was this modality difference statistically significant (see Table 1). This effect for ISC-HR occurred in both error-free and full models (see Supplementary Table 6). When checking conventional t-/Wilcoxon tests (depending on normality), results were similar to LMM results, though ISC-HR was no longer significant (SRC-HR, $t(21) = 1.176$, $p = 0.253$; ISC-HR, $t(21) = 1.890$, $p = .073$; ISC-RR, $t(22) = .447$, $p = .659$). Although self-reports of engagement were also higher in AV (see Figure 2), this difference was not significant in the LMM (ß = 0.33, 95% CI [-0.10-0.76], p = .129, $R^2_{(fixed)} = 0.007$), nor in the traditional t-test ($t(24) = 1.901$, $p = 0.069$). In checking the behavioural relevance of synchrony as an engagement index, ISC-HR was not significantly correlated (rho = -0.093, p = .295) nor a significant predictor of self-reported engagement (ß = -0.001, 95% CI [-0.003-0.001], p = .464, $R^2_{(fixed)} = 0.005$).

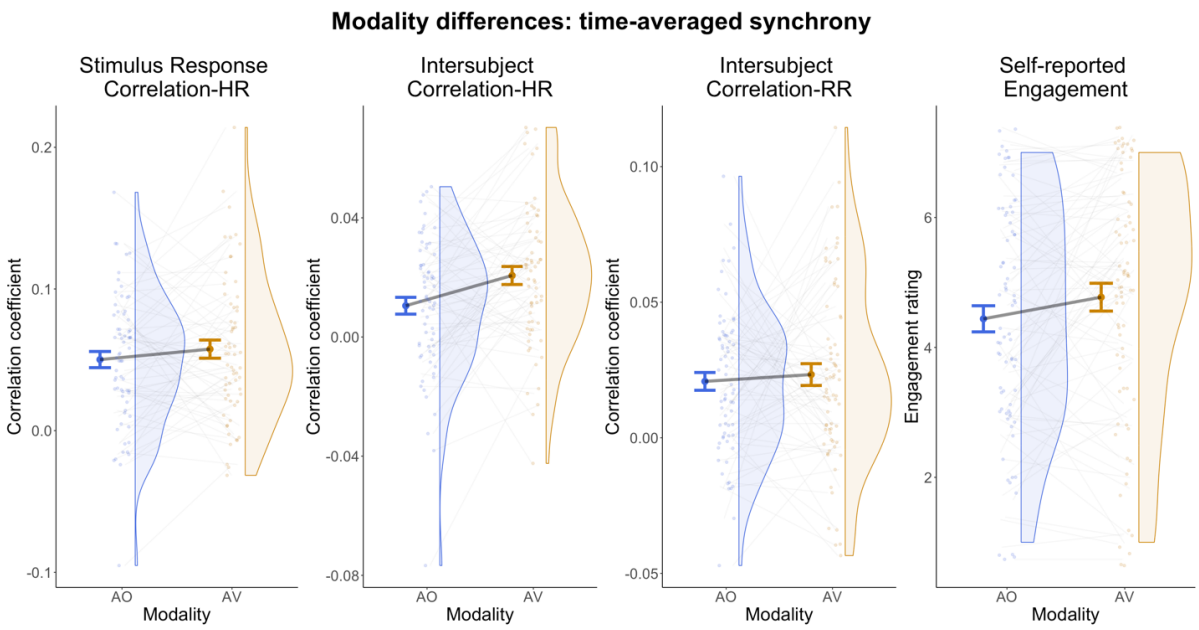

Figure 2. Raincloud plots showing modality differences of audio-only (blue) and audio-visual (orange) conditions for time-averaged synchrony measures of SRC-HR (far right), ISC-HR (right), and ISC-RR (left) as well as self-reported engagement (far left). Smaller points represent each data point, while the larger point represents the mean; error bars represent standard error of the mean. The shape of the distribution is shown by the half violin.

Table 1. Linear mixed models comparing synchrony between modalities

	SRC-HR			**ISC-HR**			**ISC-RR**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.050	0.022 – 0.079	**0.001**	0.012	0.002 – 0.022	**0.020**	0.021	0.010 – 0.032	**<0.001**
mod [AV]	0.009	-0.006 – 0.024	0.235	0.010	0.001 – 0.019	**0.028**	0.002	-0.007 – 0.012	0.633
Random Effects									
σ^2	0.01			0.00			0.01		
τ_{00}	$0.00_{\text{mac:conc}}$			0.00_{piece}			$0.00_{\text{mac:conc}}$		
	0.00_{piece}						0.00_{piece}		
τ_{11}	$0.00_{\text{mac1.modAO}}$			$0.00_{\text{mac1.modAO}}$					
	$0.00_{\text{mac2.modAV}}$			$0.00_{\text{mac2.modAV}}$					
ρ_{01}									
ρ_{01}									
ICC	0.08			0.05			0.01		
N	3_{piece}			3_{piece}			3_{piece}		
	16_{mac}			16_{mac}			16_{mac}		
	2_{conc}						2_{conc}		
Observations	1075			1075			1152		
Marginal R^2 / Conditional R^2	0.002 / 0.079			0.007 / 0.056			0.000 / 0.011		

Synchrony increases at section boundaries, depending on modality

LMMs showed that in time-resolved epochs centred around section boundaries, SRC-HR and ISC-HR were significantly higher in the AV condition, while ISC-RR was higher in AO (see Figure 3 and Table 2, see Supplementary Table 7 and code for maximal models, which replicated these effects).

LMMs additionally showed that time windows predicted synchrony measures at section boundaries. Synchrony measures were significantly higher in the time window centred at the section boundary compared to time windows before (10 seconds before: ISC-HR), and after (5 seconds after: SRC-HR, ISC-RR; 10 seconds after SRC-HR, ISC-HR, and ISC-RR) (see Figure 3 and Table 2).

In assessing the interaction between time window and modality, Figure 3 shows that synchrony was higher before and at section boundaries (window 0) compared to 5 and 10 seconds after section boundary in the AV condition for SRC-HR and ISC-HR. This latter effect was further confirmed by LMMs and by significant pairwise comparisons (Bonferroni corrected) for ISC-HR (Table 3). ISC-RR was higher in AV just after (5 second) section boundary, but this effect was not significant in Bonferroni corrected pairwise comparisons (Table 3).

Upon checking actual HR and RR responses at these time points[16], this synchrony corresponds to typical orienting responses (see Figure 4). HR follows a deceleration-acceleration pattern, where the estimate in HR in window -10 and window 10 is significantly higher than window 0 (intercept, see Table 4 and Supplementary Table 8). Although RR decelerated and accelerated similarly to HR (see Figure 4), these were not significant across time windows.

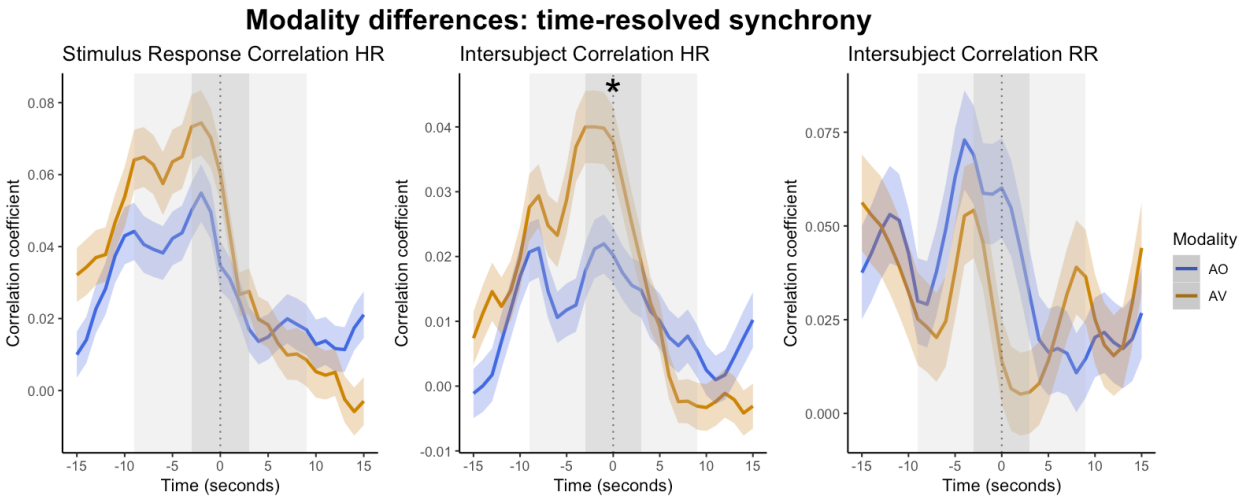

Figure 3. Synchrony values at 15 seconds prior to and after section boundaries for stimulus correlation of HR (left), inter-subject correlation of HR (middle) and inter-subject correlation of RR (right). The zero point represents the section boundary. Thick lines represent the average, and the ribbon represents the standard error of the mean for audio-only (AO, blue) and audio-visual (AV, orange) conditions. * represents a significant modality effect within that time window.

Table 2. Linear mixed-effects models for synchrony with predictors of time window and modality.

Linear mixed models

	SRC-HR			**ISC-HR**			**ISC-RR**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.04	0.02 – 0.06	**<0.001**	0.02	0.01 – 0.03	**<0.001**	0.05	0.03 – 0.07	**<0.001**
mod [AV]	0.02	0.00 – 0.04	**0.028**	0.02	0.01 – 0.03	**0.002**	-0.03	-0.06 – -0.00	**0.030**
win-10	-0.01	-0.03 – 0.01	0.401	-0.01	-0.02 – -0.00	**0.014**	-0.00	-0.03 – 0.02	0.786
win-5	0.01	-0.01 – 0.02	0.487	-0.00	-0.01 – 0.01	0.544	-0.00	-0.03 – 0.02	0.788
win [5]	-0.02	-0.04 – -0.00	**0.021**	-0.01	-0.02 – 0.00	0.063	-0.04	-0.06 – -0.01	**0.009**
win [10]	-0.02	-0.04 – -0.00	**0.021**	-0.01	-0.02 – -0.00	**0.007**	-0.03	-0.06 – -0.01	**0.018**
modAV:win-10	-0.00	-0.03 – 0.02	0.816	-0.01	-0.02 – 0.01	0.236	0.03	-0.01 – 0.07	0.153
modAV:win-5	0.00	-0.02 – 0.03	0.746	-0.00	-0.02 – 0.01	0.612	0.01	-0.02 – 0.05	0.493
mod [AV] × win [5]	-0.02	-0.05 – 0.00	0.089	-0.02	-0.04 – -0.01	**0.002**	0.04	0.00 – 0.08	**0.034**
mod [AV] × win [10]	-0.03	-0.06 – -0.01	**0.008**	-0.02	-0.04 – -0.01	**0.001**	0.04	-0.00 – 0.08	0.059
Random Effects									
σ^2	0.00			0.00			0.01		
τ_{00}	0.00_{piece}			$0.00_{\text{mac:conc}}$			$0.00_{\text{mac:conc}}$		
	0.00_{conc}								
τ_{11}	$0.00_{\text{mac.modAV}}$			$0.00_{\text{mac.modAV}}$					
	$0.00_{\text{mac1.modAO}}$			$0.00_{\text{mac1.modAO}}$					
	$0.00_{\text{mac2.modAV}}$			$0.00_{\text{mac2.modAV}}$					
ρ_{01}									
ρ_{01}									
ICC	0.08			0.06			0.03		
N	2_{conc}			16_{mac}			16_{mac}		
	3_{piece}			2_{conc}			2_{conc}		
	16_{mac}								
Observations	645			645			690		
Marginal R^2 / Conditional R^2	0.120 / 0.191			0.117 / 0.167			0.024 / 0.055		

Table 3. Pairwise comparison (Bonferroni corrected) between modality (AO-AV) in each time window.

	Contrast AO-AV time window	estimate	SE	df	T ratio	*p*
SRC-HR	Window -10	-0.018	0.010	154	-1.882	.308
	Window -5	**-0.025**	**0.010**	**154**	**-2.631**	**.047**
	Window 0	-0.021	0.010	154	-2.196	.148
	Window 5	0.001	0.010	154	0.094	1.000
	Window 10	0.013	0.010	154	1.402	.814
ISC-HR	Window -10	-0.007	0.005	216	-1.467	0.720
	Window -5	-0.012	0.005	216	-2.414	0.083
	Window 0	**-0.016**	**0.005**	**216**	**-3.123**	**0.010**
	Window 5	0.006	0.005	216	1.200	1.000
	Window 10	0.007	0.005	216	1.413	0.796
ISC-RR	Window -10	0.002	0.014	659	0.154	1.000
	Window -5	0.017	0.014	659	1.207	1.000
	Window 0	0.030	0.014	659	2.175	0.150
	Window 5	-0.012	0.014	659	-0.832	1.000
	Window 10	-0.007	0.014	659	-0.500	1.000

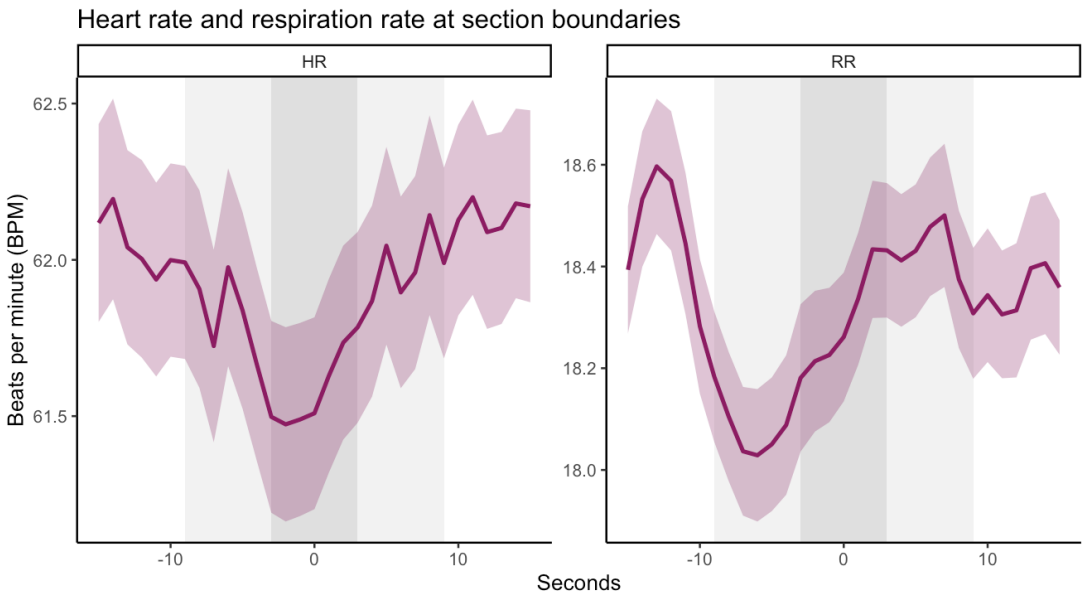

Figure 4. Heart rate (HR, left) and respiration rate (RR, right) BPM values at 15 seconds prior to and after section boundaries. The zero point represents the section boundary. Thick lines denote the average BPM, while ribbons around the average line represent the standard error.

Table 4. Linear mixed models for raw heart and respiration rate at epochs centred around section boundaries

	Heart Rate			**Respiration Rate**		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	61.84	52.52 – 71.17	**<0.001**	18.48	17.12 – 19.83	**<0.001**
win-10	0.46	0.07 – 0.85	**0.022**	0.15	-0.14 – 0.45	0.300
win-5	0.25	-0.14 – 0.64	0.207	-0.24	-0.53 – 0.06	0.115
win [5]	0.37	-0.02 – 0.76	0.061	0.10	-0.19 – 0.40	0.488
win [10]	0.53	0.14 – 0.92	**0.008**	0.07	-0.22 – 0.36	0.638
Random Effects						
σ^2	2.54			1.54		
τ_{00}	$50.88_{\text{mac:conc}}$			$7.24_{\text{mac:conc}}$		
	0.15_{piece}			0.42_{piece}		
	39.81_{conc}			0.04_{conc}		
τ_{11}	$10.74_{\text{mac1.modAO}}$					
	$20.67_{\text{mac2.modAV}}$					
ρ_{01}						
ρ_{01}						
ICC	0.97			0.83		
N	2_{conc}			2_{conc}		
	3_{piece}			3_{piece}		
	16_{mac}			16_{mac}		
Observations	645			690		
Marginal R^2 / Conditional R^2	0.000 / 0.973			0.002 / 0.833		

4. Discussion

An important aspect of (live) musical experiences is the listeners' engagement. Despite the methodological challenges in assessing such musical experiences, dynamic engagement is increasingly assessed with neurophysiological synchrony measures. By extending these techniques to a naturalistic setting in a concert hall, the current study pushed the boundaries in how to measure and analyse musical experiences. Importantly, we also compared approaches to gauge the robustness of these cardiorespiratory synchrony analyses. As engagement might be enhanced by seeing musicians perform[8,10], participants were presented with AO and AV performances of Western classical music in a concert setting while cardiorespiratory measures were continuously measured. We show that heart rate (i.e., *speed* of heartbeats in BPM) correlated with spectral flux of music above chance. Additionally, respiration rate and (tentatively) heart rate correlated across audience members above chance. Musical rhythms did not align with heartbeat rhythms nor breathing rhythms above chance. We also find that compared to the AO condition, AV music performances evoked significantly higher ISC-HR; this effect was more robust on a time-resolved, compared to time-averaged, level.

Audience heart rate and respiration rate correlation

HR correlated with spectral flux (SRC-HR) above chance level, indicating that the acceleration/deceleration of audiences' HR correlated with increases/decreases of spectral flux in the music. ISC-HR and ISC-RR measures were also above chance level, indicating that acceleration/deceleration of HR and RR correlated across audience members experiencing music simultaneously, supporting previous research showing that ISC of HR[18,46] and EEG[4] is evoked when participants simultaneously experience the same auditory stimuli. We note, however, that although

ISC-HR was significant in the main analysis of time-averaged synchrony within musical sections (see Supplementary Table 1), when ISC-HR was time-averaged across 30-second bins this measure did not meet the significance threshold after correcting for multiple comparisons. Therefore, it might be important to assess synchrony related to musical structure and on a time-resolved level; see below discussion on 'Time-resolved synchrony'. That ISC-RR was above chance level somewhat contrasts findings of Madsen and Parra[46], who found ISC-RR not to be significant. One explanation could be due to stimuli differences. Here, we used music stimuli, whereas Madsen and Parra used instructional video (speech). More importantly, the setting of a concert hall might have increased the engagement of the listeners compared to the laboratory setting. Our findings suggest that HR and RR (i.e., speed of heartbeats/breaths) may become synchronized across groups of people experiencing auditory stimuli (like music) simultaneously, but may be context dependent.

Heartbeat and breathing cycles do not rhythmically align to music

Both musical rhythm–heartbeat phase synchrony and musical rhythm–breath phase synchrony (SRPC measures) were not significantly above chance level. In other words, musical rhythms aligned with neither audiences' breathing cycles nor their heartbeat cycles. Although this lack of phase synchrony was expected for heart rhythms[51–53], this contrasts previous findings by Etzel et al[49] and Haas et al[50] that breathing aligns with musical beats. There are a few potential reasons for this discrepancy between results. First, Etzel et al. and Haas et al. compared respiratory signals to the beat of music, while the current study assesses spectral flux filtered at respiration frequencies. Although our method is therefore not comparable with this previous research, we nonetheless suggest that our method is more generalisable for future naturalistic research. It seemed like the previous research chose music to roughly correspond to healthy breathing frequencies (see Haas et al, Table 1, Figure 1 and 2; Etzel et al. Table 1). However, much music might not correspond to such specific frequencies. Indeed, the musical choice for the current study was motivated by the need to

present a typical musical program, where the beat did not fall into the range of healthy heartbeats/breathing (see Supplementary Table 1). It therefore seemed problematic to assess synchrony of the actual beat with heartbeats/breaths. Assessing musical features that strongly correlate with beat onsets (spectral flux[24]) nonetheless yields a somewhat comparable approach. Additionally, extracting the energy at the natural frequency of heartbeats/breathing allows assessing music that has beats falling outside of the natural breathing/heartbeat range, providing a useful measure for future research. A second explanation for not replicating musical beat-respiration synchrony, is that previous studies used music with relatively regular rhythmic and metric structures[49,50], while the stimuli in the current study included one contemporary piece with ambiguous rhythms. However, there were no respiration synchrony differences between pieces with regular (Bach, Beethoven) and irregular (Messiaen) metric structures (see Supplementary Table 3). A third explanation could be that respiration is more likely to align with beats in laboratory experiments as the attention (per instruction) is mainly on music. Beyond complexities in musical structure, a real-world setting presents several additional factors, such as the visual and social aspects[10,86] , where attention may fluctuate a lot more between the music and other features. Thus, although neural[72,74] and respiratory[49,50] rhythms may align to auditory beats in a laboratory setting, the current results suggest that this may not occur to music with broader tempi in a music concert setting, at least for this choice of Western Classical music.

Heartbeats and breaths likewise did not significantly align in phase between participants. One explanation could be the lack of social interaction, which is a crucial aspect in higher-order phase coupling[40]. Indeed, phase synchrony between participants has previously been found in dancing[87,88] and musical ensembles [38,39,89,90] where there is a clear shared goal of a coordinated performance (see[40,41]). Typical Western classical concert etiquette calls for enhanced focus on the performance, thus reducing body movements and interaction between audience members[10]. Indeed, ratings for

533 'urge to move to the music' (see[7]) were heavily skewed to low ratings; even testing significance of
534 phase coherence in audience members who only had high 'urge to move' ratings did not yield
535 significant results (see Supplementary Table 5). Additionally, videos showed that the audience
536 members respected this etiquette and generally did not move too much[7]. Therefore, we suggest that
537 phase synchrony – typically seen in interactive settings[40] – may not occur in non-interactive settings.
538

539 **Time-averaged synchrony may not relate to naturalistic engagement**

540 Next, synchrony measures were compared between AV and AO conditions. ISC-HR was significantly
541 greater in AV performances. However, this result was not shown to be robust with a sanity check
542 using conventional *t*-test. Although our data are more appropriate for LMMs due to naturalistic
543 paradigm and physiological data[78], the fact that the significance is not replicated in a simpler *t*-test
544 thus allows only cautious interpretation[83]. Another reason to be cautious is that – contrary to what
545 was expected – ISC-HR was not significantly related to engagement that was self-reported after each
546 piece. The current findings contrast findings from Ki et al.[31] who found that neural ISC was stronger
547 in attend than non-attend conditions, especially in audiovisual stimuli. The current finding also
548 contrasts studies showing that HR[18] and neural[19] synchrony can reflect engagement. One explanation
549 for the discrepancy in results is that these previous studies had explicit attend/non-attend
550 conditions, where participants were instructed to count backwards in the non-attend conditions.
551 Here, participants experienced concert performances naturally with no formal instruction. This
552 suggests that while ISC (of HR) could be a marker in clearly defined engagement conditions, it might
553 not necessarily be generalisable to less controlled settings. Additionally, inter- and intra-individual
554 differences may influence synchrony and engagement[55,91,92] , and could be a potential venture for
555 future studies to explore individual versus group synchrony. Overall, we show time-averaged
556 synchrony does not reliably reflect overall self-reported engagement measurements. With this in
557 mind, we anticipated results for time-resolved synchrony might yield clearer results.

558 **Time-resolved synchrony reveals synchronized heart rate orienting response**

559 Engagement with music typically fluctuates across time[4,12,54]. Thus, we assessed synchrony changes
560 between modalities on a time-resolved scale. We 'zoomed in' on synchronized responses at salient
561 moments in the music, here defined as section boundaries in music (see Supplementary Table 1),
562 which are important structural locations in music. The current results show that both ISC-HR and
563 SRC-HR increased at such section boundaries, supporting and extending previous work on ISC[12,16] to
564 SRC. Although we did not have time-resolved self-reports to assess the relation of synchrony to
565 engagement, we assessed the continuous HR and RR. We show that synchronized responses reflected
566 significant deceleration-acceleration patterns in HR and RR. Such patterns have been long associated
567 with a typical orienting response, which marks an increase in attentiveness[93] to anticipate and
568 perceive events[14,94–96]. Thus, such increases of cardiorespiratory synchrony at section boundaries
569 suggest increased engagement with music at these time points, related to auditory cues indicating a
570 section is ending, such as a cadential ending (i.e., harmonic closure in Bach and Beethoven) and a
571 slight slowing down or change in tempo and/or material (see Supplementary Table 1). This
572 physiological synchrony at salient music events suggests a bottom-up capturing of engagement
573 during a shared musical experience[40]. We acknowledge that the definition of salient moments to
574 explore engagement are limited in the current study to section boundaries. However, humans also
575 respond to segmentation not only at section boundaries[4,12,61,97], but also at smaller phrasal[98] and
576 larger, movement[68] levels. Thus, the current work provides a starting point for exploring engagement
577 to salient moments; future work could compare 'non' salient moments, transitions, and segments at
578 lower and higher hierarchical musical levels.

579

580 Importantly, Figure 3 shows that ISC-HR significantly increased at section boundaries more so in the
581 AV condition. This increase suggests that visual information further increased engagement at section
582 boundaries. Indeed, previous performance studies have found that certain gestures cue and

583 anticipate structurally important locations, such as leaning forward towards a cadence(harmonic
584 closure)[99] , increased movement amplitude[100] , or finishing 'flourish' gestures[101] as well as deep
585 breaths and sweeping motions prior to onset of new phrases[102]. Indeed, videos show that the pianist
586 tended to lean forward at section boundaries (though see limitations regarding analysis of visual
587 features). Therefore, a combination of audio and visual aspects likely increases engagement with
588 music at a local level, as shown most robustly with ISC-HR.
589

590 ## 5. Limitations

591 Although we show greater synchrony in the AV (compared to AO) condition, it is unclear whether
592 this synchrony increase might simply be related to increased information processing of both visual
593 and audio information, compared to simply audio information. For example, RR of audiences did not
594 synchronise to the spectral flux of the music (insignificant SRC-RR result), but RR did significantly
595 synchronise across audience members (ISC-RR). This suggests that the audiences' RR synchrony was
596 rather driven by something else than the acoustic signal, perhaps the musician movements (see[7]).
597 Indeed, we saw that SRC-RR was more significant in the AV than the AO condition (see
598 Supplementary Table 4). As it was not possible to extract reliable visual features in the current study
599 due to the camera location in the concert hall (see angle in Figure 1A), future work would therefore
600 need to assess responses not only to auditory features, but also visual features, such as facial
601 expressions[103], quantity of motion, and speed of movements[104] to shed more light on how much
602 variance in the audience experience can be explained by auditory and by visual features. Additionally,
603 we did not test a visual only condition which would allow us to disentangle the difference between
604 added information processing and actual engagement. Finally, we acknowledge that aspects of
605 'liveness' may have likewise contributed to the current results. For example, previous evidence
606 suggested that musicians have different perceived expressiveness when performing with no
607 audience (rehearsal condition) compared to performing in front of an audience (recital condition[105]).

608 However, as the literature is mixed as to whether liveness influences emotional audience
609 experience[106] (for further discussion see[7,107]), but the literature on audio-visual influences on music
610 appreciation is relatively robust[9], it seems more plausible that the current results are explained by
611 the modality difference rather than liveness.
612

613 ## 6. Conclusion

614 The current study pushed the boundaries of how to measure and analyse music listening experiences.
615 One crucial aspect of such a musical experience is the listener's engagement, which can be assessed
616 by measuring neurophysiological synchrony. As engagement could be affected by modality, we
617 explored engagement differences between audio-visual and audio-only music performances in a
618 Western classical concert setting, using cardiorespiratory synchrony as a measure of engagement.
619 We extended studies investigating neural synchrony[4,17], by comparing the robustness of different
620 analyses to calculate peripheral physiological synchrony[18,46,54] in music concert settings[2,15,16]. We
621 showed that seeing musicians perform can heighten an audience's engagement, though this might be
622 context-dependent, and is best assessed by time-resolved (rather than time-averaged) inter-subject
623 correlation of HR (i.e., synchrony of heartbeat speed). Rhythmic alignment (phase synchrony) of
624 heartbeats/respiration with other audience members/music) were below chance level. However,
625 questions arose whether phase coherence might become more relevant in contexts where listeners
626 can move and interact, for example, at pop and rock concerts. More studies, involving a wider range
627 of genres and contexts are needed to further understand synchrony dynamics in real-world music
628 contexts. Importantly, the current study highlights the advantages of timeseries analysis in
629 quantifying dynamic engagement, which could be applied not only to data collected during concert
630 music listening, but in different naturalistic settings (e.g., open air concerts).

Acknowledgements

The authors would like to thank Lea Fink, our music theorist who advised on the musical sections. Thanks also go to Klaus Frieler and Örjan De Manzano for discussions about statistics. Many thanks also to the ArtLab team, who assisted in concert preparations.

Competing interests. The authors declare no competing interests.

7. References

1. Gabrielsson, A. & Wik, S. L. Strong Experiences Related to Music: A Descriptive System. *Music. Sci.* **7**, 157–217 (2003).
2. Kragness, H. E. *et al.* An itsy bitsy audience: Live performance facilitates infants' attention and heart rate synchronization. *Psychol. Aesthet. Creat. Arts* (2023) doi:10.1037/aca0000597.
3. Lamont, A. University students' strong experiences of music: Pleasure, engagement, and meaning. *Music. Sci.* **15**, 229–249 (2011).
4. Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P. & Berger, J. Natural music evokes correlated EEG responses reflecting temporal structure and beat. *NeuroImage* **214**, 116559 (2020).
5. Omigie, D. & Mencke, I. A model of time-varying music engagement. *Philos. Trans. R. Soc. B Biol. Sci.* **379**, 20220421 (2024).
6. Schubert, E., Vincs, K. & Stevens, C. J. Identifying Regions of Good Agreement among Responders in Engagement with a Piece of Live Dance. *Empir. Stud. Arts* **31**, 1–20 (2013).
7. Czepiel, A., Fink, L. K., Seibert, C., Scharinger, M. & Kotz, S. A. Aesthetic and physiological effects of naturalistic multimodal music listening. *Cognition* **239**, 105537 (2023).
8. Dobson, M. C. New Audiences for Classical Music: The Experiences of Non-attenders at Live Orchestral Concerts. *J. New Music Res.* **39**, 111–124 (2010).

9. Platz, F. & Kopiez, R. When the Eye Listens: A Meta-analysis of How Audio-visual Presentation Enhances the Appreciation of Music Performance. *Music Percept.* **30**, 71–83 (2012).
10. Wald-Fuhrmann, M. *et al.* Music Listening in Classical Concerts: Theory, Literature Review, and Research Program. *Front. Psychol.* **12**, (2021).
11. Olsen, K. N., Dean, R. T. & Stevens, C. J. A continuous measure of musical engagement contributes to prediction of perceived arousal and valence. *Psychomusicology Music Mind Brain* **24**, 147–156 (2014).
12. Dauer, T. *et al.* Inter-subject Correlation While Listening to Minimalist Music: A Study of Electrophysiological and Behavioral Responses to Steve Reich's Piano Phase. *Front. Neurosci.* **15**, (2021).
13. Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P. & Berger, J. Inter-subject correlation of electroencephalographic and behavioural responses reflects time-varying engagement with natural music. *Eur. J. Neurosci.* ejn.16324 (2024) doi:10.1111/ejn.16324.
14. Bradley, M. M. & Lang, P. J. Affective reactions to acoustic stimuli. *Psychophysiology* **37**, 204–215 (2000).
15. Chabin, T. *et al.* Interbrain emotional connection during music performances is driven by physical proximity and individual traits. *Ann. N. Y. Acad. Sci.* **1508**, 178–195 (2022).
16. Czepiel, A. *et al.* Synchrony in the periphery: inter-subject correlation of physiological responses during live music concerts. *Sci. Rep.* **11**, 22457 (2021).
17. Dmochowski, J. P., Sajda, P., Dias, J. & Parra, L. C. Correlated Components of Ongoing EEG Point to Emotionally Laden Attention – A Possible Marker of Engagement? *Front. Hum. Neurosci.* **6**, (2012).
18. Pérez, P. *et al.* Conscious processing of narrative stimuli synchronizes heart rate between individuals. *Cell Rep.* **36**, 109692 (2021).
19. Madsen, J., Margulis, E. H., Simchy-Gross, R. & Parra, L. C. Music synchronizes brainwaves across listeners with strong effects of repetition, familiarity and training. *Sci. Rep.* **9**, 3576 (2019).
20. Albouy, P., Benjamin, L., Morillon, B. & Zatorre, R. J. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* **367**, 1043–1047 (2020).
21. Albouy, P. *et al.* Spectro-temporal acoustical markers differentiate speech from song across cultures. *Nat. Commun.* **15**, 4835 (2024).
22. Obleser, J., Herrmann, B. & Henry, M. J. Neural Oscillations in Speech: Don't be Enslaved by the Envelope. *Front. Hum. Neurosci.* **6**, (2012).

23. Schultz, B. G., Brown, R. M. & Kotz, S. A. Dynamic acoustic salience evokes motor responses. *Cortex* **134**, 320–332 (2021).

24. Weineck, K., Wen, O. X. & Henry, M. J. Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience. *eLife* **11**, e75515 (2022).

25. Alluri, V. *et al.* Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage* **59**, 3677–3689 (2012).

26. Lartillot, O. & Toiviainen, P. A Matlab toolbox for musical feature extraction from audio. in *International conference on digital audio effects* (Bordeaux, 2007).

27. Burger, B., Saarikallio, S., Luck, G., Thompson, M. R. & Toiviainen, P. Relationships Between Perceived Emotions in Music and Music-induced Movement. *Music Percept.* **30**, 517–533 (2013).

28. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 667–685 (2019).

29. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject Synchronization of Cortical Activity During Natural Vision. *Science* **303**, 1634–1640 (2004).

30. Kang, O. & Wheatley, T. Pupil dilation patterns spontaneously synchronize across individuals during shared attention. *J. Exp. Psychol. Gen.* **146**, 569–576 (2017).

31. Ki, J. J., Kelly, S. P. & Parra, L. C. Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. *J. Neurosci.* **36**, 3092–3101 (2016).

32. Irsik, V. C., Johnsrude, I. S. & Herrmann, B. Neural Activity during Story Listening Is Synchronized across Individuals Despite Acoustic Masking. *J. Cogn. Neurosci.* **34**, 933–950 (2022).

33. Schmälzle, R., Häcker, F. E. K., Honey, C. J. & Hasson, U. Engaged listeners: shared neural processing of powerful political speeches. *Soc. Cogn. Affect. Neurosci.* **10**, 1137–1143 (2015).

34. Dikker, S. *et al.* Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Curr. Biol.* **27**, 1375–1380 (2017).

35. Peelle, J. E., Gross, J. & Davis, M. H. Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cereb. Cortex* **23**, 1378–1387 (2013).

36. Jones, M. R. & Boltz, M. Dynamic attending and responses to time. *Psychol. Rev.* **96**, 459–491 (1989).

37. Large, E. W. & Snyder, J. S. Pulse and Meter as Neural Resonance. *Ann. N. Y. Acad. Sci.* **1169**, 46–57 (2009).

38. Gugnowska, K. *et al.* Endogenous sources of interbrain synchrony in duetting pianists. *Cereb. Cortex* **32**, 4110–4127 (2022).
39. Müller, V. & Lindenberger, U. Cardiac and Respiratory Patterns Synchronize between Persons during Choir Singing. *PLOS ONE* **6**, e24893 (2011).
40. Dumas, G. & Fairhurst, M. T. Reciprocity and alignment: quantifying coupling in dynamic interactions. *R. Soc. Open Sci.* **8**, 210138 (2021).
41. Fairhurst, M. T., Janata, P. & Keller, P. E. Being and Feeling in Sync with an Adaptive Virtual Partner: Brain Mechanisms Underlying Dynamic Cooperativity. *Cereb. Cortex* **23**, 2592–2600 (2013).
42. Tschacher, W. *et al.* Physiological synchrony in audiences of live concerts. *Psychol. Aesthet. Creat. Arts* **17**, 152–162 (2021).
43. Criscuolo, A., Schwartze, M. & Kotz, S. A. Cognition through the lens of a body–brain dynamic system. *Trends Neurosci.* S0166223622001229 (2022) doi:10.1016/j.tins.2022.06.004.
44. Klimesch, W. The frequency architecture of brain and brain body oscillations: an analysis. *Eur. J. Neurosci.* **48**, 2431–2453 (2018).
45. Parviainen, T., Lyyra, P. & Nokia, M. S. Cardiorespiratory rhythms, brain oscillatory activity and cognition: review of evidence and proposal for significance. *Neurosci. Biobehav. Rev.* **142**, 104908 (2022).
46. Madsen, J. & Parra, L. C. Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus* **1**, pgac020 (2022).
47. Al, E. *et al.* Heart–brain interactions shape somatosensory perception and evoked potentials. *Proc. Natl. Acad. Sci.* **117**, 10575–10584 (2020).
48. Grund, M. *et al.* Respiration, Heartbeat, and Conscious Tactile Perception. *J. Neurosci.* **42**, 643–656 (2022).
49. Etzel, J. A., Johnsen, E. L., Dickerson, J., Tranel, D. & Adolphs, R. Cardiovascular and respiratory responses during musical mood induction. *Int. J. Psychophysiol.* **61**, 57–69 (2006).
50. Haas, F., Distenfeld, S. & Axen, K. Effects of perceived musical rhythm on respiratory pattern. *J. Appl. Physiol.* **61**, 1185–1191 (1986).
51. Ellis, R. J. & Thayer, J. F. Music and Autonomic Nervous System (Dys)Function. *Music Percept.* **27**, 317–326 (2010).
52. Koelsch, S. & Jäncke, L. Music and the heart. *Eur. Heart J.* **36**, 3043–3049 (2015).

53. Mütze, H., Kopiez, R. & Wolf, A. The effect of a rhythmic pulse on the heart rate: Little evidence for rhythmical 'entrainment' and 'synchronization'. *Music. Sci.* **24**, 377–400 (2020).

54. Fink, L. K., Hurley, B. K., Geng, J. J. & Janata, P. A linear oscillator model predicts dynamic temporal attention and pupillary entrainment to rhythmic patterns. *J. Eye Mov. Res.* **11**, 10.16910/jemr.11.2.12 (2018).

55. Presicce, G. & Bailes, F. Engagement and visual imagery in music listening: An exploratory study. *Psychomusicology Music Mind Brain* **29**, 136–155 (2019).

56. Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J. & Menon, V. Neural Dynamics of Event Segmentation in Music: Converging Evidence for Dissociable Ventral and Dorsal Networks. *Neuron* **55**, 521–532 (2007).

57. Burunat, I., Levitin, D. J. & Toiviainen, P. Breaking (musical) boundaries by investigating brain dynamics of event segmentation during real-life music-listening. *Proc. Natl. Acad. Sci.* **121**, e2319459121 (2024).

58. Lerdahl, F. & Jackendoff, R. An Overview of Hierarchical Structure in Music. *Music Percept. Interdiscip. J.* **1**, 229–252 (1983).

59. Clarke, E. F. & Krumhansl, C. L. Perceiving Musical Time. *Music Percept. Interdiscip. J.* **7**, 213–251 (1990).

60. Deliege, I. Grouping Conditions in Listening to Music: An Approach to Lerdahl & Jackendoff's Grouping Preference Rules. *Music Percept. Interdiscip. J.* **4**, 325–359 (1987).

61. Phillips, M. *et al.* What Determines the Perception of Segmentation in Contemporary Music? *Front. Psychol.* **11**, 1001 (2020).

62. Popescu, T., Widdess, R. & Rohrmeier, M. Western listeners detect boundary hierarchy in Indian music: a segmentation study. *Sci. Rep.* **11**, 3112 (2021).

63. Davidson, J. W. Visual Perception of Performance Manner in the Movements of Solo Musicians. *Psychol. Music* **21**, 103–113 (1993).

64. Herrmann, B. & Johnsrude, I. S. Absorption and Enjoyment During Listening to Acoustically Masked Stories. *Trends Hear.* **24**, 233121652096785 (2020).

65. Kandylaki, K. D., Utroša-Škerjanec, M. & Kotz, S. Rhythmic Regularity in Speech is Modulated by Amplitude, Pitch, Duration, and the Spectral Center of Gravity. *SSRN Electron. J.* (2022) doi:10.2139/ssrn.4213992.

66. Tzanetakis, G. & Cook, P. Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.* **10**, 293–302 (2002).

67. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput. Intell. Neurosci.* **2011**, 1–9 (2011).

68. Merrill, J., Czepiel, A., Fink, L. T., Toelle, J. & Wald-Fuhrmann, M. The aesthetic experience of live concerts: Self-reports and psychophysiology. *Psychol. Aesthet. Creat. Arts* **17**, 134–151 (2021).

69. Sammler, D., Grigutsch, M., Fritz, T. & Koelsch, S. Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* **44**, 293–304 (2007).

70. Assaneo, M. F., Rimmele, J. M., Sanz Perl, Y. & Poeppel, D. Speaking rhythmically can shape hearing. *Nat. Hum. Behav.* **5**, 71–82 (2021).

71. Cohen. *Analyzing Neural Time Series Data: Theory and Practice*. (The MIT Press, Cambridge, Massachusetts, 2014).

72. Harding, E. E., Sammler, D., Henry, M. J., Large, E. W. & Kotz, S. A. Cortical tracking of rhythm in music and speech. *NeuroImage* **185**, 96–101 (2019).

73. Luck, G., Toiviainen, P. & Thompson, M. R. Perception of Expression in Conductors' Gestures: A Continuous Response Study. *Music Percept.* **28**, 47–57 (2010).

74. Vanden Bosch der Nederlanden, C. M., Joanisse, M. F. & Grahn, J. A. Music as a scaffold for listening to speech: Better neural phase-locking to song than speech. *NeuroImage* **214**, 116767 (2020).

75. Wassiliwizky, E., Koelsch, S., Wagner, V., Jacobsen, T. & Menninghaus, W. The emotional power of poetry: neural circuitry, psychophysiology and compositional principles. *Soc. Cogn. Affect. Neurosci.* **12**, 1229–1240 (2017).

76. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, (2001).

77. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).

78. Page-Gould, E. Multilevel Modeling. in *Handbook of Psychophysiology* (eds. Cacioppo, J. T., Tassinary, L. G. & Berntson, G. G.) 662–678 (Cambridge University Press, 2016). doi:10.1017/9781107415782.030.

79. Barr, D. J. *Learning Statistical Models through Simulation in R: An Interactive Textbook*. (2021).

80. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using {lme4}. *J. Stat. Softw.* **67**, 1–48 (2015).

81. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. {lmerTest} Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* **82**, 1–26 (2017).

82. Lüdecke, D. sjPlot: Data Visualization for Statistics in Social Science. (2023).

83. Arnqvist, G. Mixed Models Offer No Freedom from Degrees of Freedom. *Trends Ecol. Evol.* **35**, 329–335 (2020).

84. Dellacherie, D., Roy, M., Hugueville, L., Peretz, I. & Samson, S. The effect of musical experience on emotional self-reports and psychophysiological responses to dissonance: Psychophysiology of musical emotion. *Psychophysiology* **48**, 337–349 (2011).

85. Lenth, R. V. emmeans: Estimated Marginal Means, aka Least-Squares Means. (2021).

86. Tervaniemi, M. The neuroscience of music – towards ecological validity. *Trends Neurosci.* **46**, 355–364 (2023).

87. Hartmann, M. *et al.* Kinematics of perceived dyadic coordination in dance. *Sci. Rep.* **9**, 15594 (2019).

88. Toiviainen, P. & Carlson, E. Embodied Meter Revisited. *Music Percept.* **39**, 249–267 (2022).

89. Lange, E. B. *et al.* In touch: Cardiac and respiratory patterns synchronize during ensemble singing with physical contact. *Front. Hum. Neurosci.* **16**, (2022).

90. Vickhoff, B. *et al.* Music structure determines heart rate variability of singers. *Front. Psychol.* **4**, (2013).

91. Forster, S. & Lavie, N. Distracted by your mind? Individual differences in distractibility predict mind wandering. *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 251–260 (2014).

92. Wohltjen, S., Toth, B., Boncz, A. & Wheatley, T. Synchrony to a beat predicts synchrony with other minds. *Sci. Rep.* **13**, 3591 (2023).

93. Barry, R. J. & Sokolov, E. N. Habituation of phasic and tonic components of the orienting reflex. *Int. J. Psychophysiol.* **15**, 39–42 (1993).

94. Park, H.-D., Correia, S., Ducorps, A. & Tallon-Baudry, C. Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nat. Neurosci.* **17**, 612–618 (2014).

95. Proverbio, A. M. *et al.* Non-expert listeners show decreased heart rate and increased blood pressure (fear bradycardia) in response to atonal music. *Front. Psychol.* **6**, (2015).

96. Stekelenburg, J. & Boxtel, A. V. Inhibition of pericranial muscle activity, respiration, and heart rate enhances auditory sensitivity. *Psychophysiology* **38**, 629–641 (2001).

97. Czepiel, A. Multimodal_piano_periphys: Data and analysis code. (2023).

98. Teng, X., Larrouy-Maestri, P. & Poeppel, D. Segmenting and Predicting Musical Phrase Structure Exploits Neural Gain Modulation and Phase Precession. *J. Neurosci.* e1331232024 (2024) doi:10.1523/JNEUROSCI.1331-23.2024.

99. Davidson, J. W. Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychol. Music* **40**, 595–633 (2012).

100. Thompson & Luck, G. Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Music. Sci.* **16**, 19–40 (2012).

101. Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C. & Hatch, W. The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *J. New Music Res.* **34**, 97–113 (2005).

102. Vines, B. W., Wanderley, M. M., Krumhansl, C. L., Nuzzo, R. L. & Levitin, D. J. Performance Gestures of Musicians: What Structural and Emotional Information Do They Convey? in *Gesture-Based Communication in Human-Computer Interaction* (eds. Camurri, A. & Volpe, G.) 468–478 (Springer, Berlin, Heidelberg, 2004). doi:10.1007/978-3-540-24598-8_43.

103. Waddell, G. & Williamon, A. Eye of the Beholder: Stage Entrance Behavior and Facial Expression Affect Continuous Quality Ratings in Music Performance. *Front. Psychol.* **8**, 513 (2017).

104. Lillywhite, A. *et al.* A functional magnetic resonance imaging examination of audiovisual observation of a point-light string quartet using intersubject correlation and physical feature analysis. *Front. Neurosci.* **16**, 921489 (2022).

105. Chang-Arana, Á. M., Mavrolampados, A., Pokki, N. & Thompson, M. R. When nerves hit: The effect of trait anxiety, situational stress, and task mastery on the perception and interpersonal accuracy of musical expressiveness. *Front. Psychol.* **13**, 1061922 (2023).

106. Belfi, A. M., Samson, D. W., Crane, J. & Schmidt, N. L. Aesthetic Judgments of Live and Recorded Music: Effects of Congruence Between Musical Artist and Piece. *Front. Psychol.* **12**, (2021).

107. Czepiel, A. M. Assessing real-world music listening in concerts: Aesthetic experiences and peripheral physiological responses. (maastricht university, 2023). doi:10.26481/dis.20231002ac.