Audio-visual concert performances synchronize an audience's 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
3350 Mississauga Road, Mississauga
Canada
https://orcid.org/0000-0002-7101-945X

Abstract

Despite the increasing availability of recorded music, people continue to engage in live musical experiences such as multimodal live concerts. However, the dynamics of audience engagement in such contexts are largely understudied. In a classical concert experiment, we presented audiences with audio-only (AO) and audio-visual (AV) piano performances while cardiorespiratory measures were continuously recorded. To investigate engagement, cardiorespiratory synchrony was calculated using both correlation and phase coherence methods. Only correlation measures remained significant in comparison to control (circular-shifted) data. Significant synchrony measures were then assessed between modalities, both across and within music pieces. 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, synchronized deceleration-acceleration heart rate (HR) patterns, typical of an 'orienting response' (an index of directed attention), occurred *within* music pieces at salient events (i.e., at section boundaries). In other words, seeing musicians perform heightened audience engagement at structurally important moments in the music. These results highlight the multimodal effects of music in real-world contexts, calling for future studies to explore wider-ranging genres and contexts to better understand dynamics of audience synchrony and engagement.

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

1. Introduction

People enjoy engaging with music, especially in live concerts[1–3]. However, the dynamics of musical engagement – defined here as a listener's real-time absorption in music[4,5] – outside the lab are largely unknown, particularly in concert settings. One concert component that likely enhances engagement is *seeing* musicians perform[6,7]. The visual component has indeed been shown to have a consistent effect on music performance evaluation[8]. Although we have recently replicated this effect outside laboratory settings, showing that audio-visual (AV) performances led to stronger aesthetic appreciation than audio-only (AO) performances in a concert context[9], these evaluations were recorded at the end of 7-12 minute long musical pieces. An important next step is to explore how the engagement of listeners varies over time, in both AO and AV conditions.

Although musical engagement can be measured with continuous behavioural responses, these ratings might distract from the music experience. A promising alternative to measure engagement in a non-distracting way is to assess neural and peripheral physiological synchrony (e.g., [3,10–13]). This approach is based on the assumption that increased engagement to music is more likely to evoke common time-locked responses across participants[4,14]. Synchrony can be between a participant's response and a representative stimulus feature, like the auditory envelope[4,11] or spectral flux[15,16]. Such synchrony can be assessed using correlation or phase coherence, i.e., rhythmic alignment[4,11,15]. 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. Again, this can be calculated with correlation[11,17] 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[11,18,19], speech[20–22], and music[3,4,14]. In phase coherence measures, higher neural synchrony occurred in engaging group discussions, indicating a potential marker for shared attention mechanisms[23]. Although phase synchrony in speech was attributed to stimulus intelligibility[24], 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 (Dynamic Attending Theory, DAT[25,26]). Indeed, successful coordination in musical ensembles was related to phase synchrony in neural[27] and heart[28] rhythms (see also[29,30]). Recent work shows coupling of cerebral activity occurs when experiencing similar emotional response to live music concerts[10]. 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[3,13,31].

Recent frameworks propose certain neural mechanisms – such as synchrony – might extend to cardiac, respiratory, gut, and pupil rhythms[32–34]. Promising results show cardiorespiratory ISC related to engagement with narratives and instructional videos[12,14,35]. In the phase domain, some research indicates that the alignments of breathing/heartbeats to external rhythms aid the processing of upcoming stimulus events[36,37]. In music, respiration has been shown to entrain to musical beats[38,39], while heartbeats do not[40–42]. Thus, 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[12]). However, music listening is dynamic, where engagement, neural synchrony[43,44], and physiological synchrony[13] change over time, especially at salient moments in the music. We defined such salient moments in music as section boundaries, which are structurally important moments in music involving several music features that change simultaneously[45–48]. Section boundaries have previously been shown to drive peripheral physiological responses[13].

The current study

Audiences might find concerts engaging due to not only listening to music but also seeing a musician perform[6,7]. To test this, concert audiences were presented with AO and AV performances of Western classical piano pieces. 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 the assumption that time-locked responses are more synchronized than control (circular-shifted) data. 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 in AV than AO at salient moments in the music, i.e., at section boundaries. Observing a clear effect of section boundary, allowed studying how these typical responses differed in the different performance modalities.

2. Method

Participants, stimuli, and procedures are identical to Czepiel et al.[9] (see General Methods and Experiment 2). Key details of the procedure are outlined below. Data and code are available on Open Science Framework (OSF)[1].

Participants

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

Stimuli

Participants were presented with AV and AO versions (see Figure 1a) of three Western classical piano pieces: 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

---

[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.

movement, and Olivier Messiaen: *Contemplation of the joyful Spirit* (No. 10 from *Twenty Contemplations on the Infant Jesus*). Both versions were performed by the same professional pianist, on the same piano, in the same 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[9].

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 in AV and AO modalities. The order of modality was counterbalanced across the two concerts. After each piece, participants rated items such as engagement.

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[15,16,49]. Therefore, the continuous spectral flux signal was obtained using the MIRToolbox[50] in Matlab (2019b), with a frame size of 25ms (as is appropriate for shortterm acoustic features[51]), with a 50% overlap (Figure 1b). 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. Spectral flux was bandpass filtered at heart and respiration frequencies (i.e., 0.98-1.04 and 0.27-0.33 Hz, respectively) and the phase angle 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).

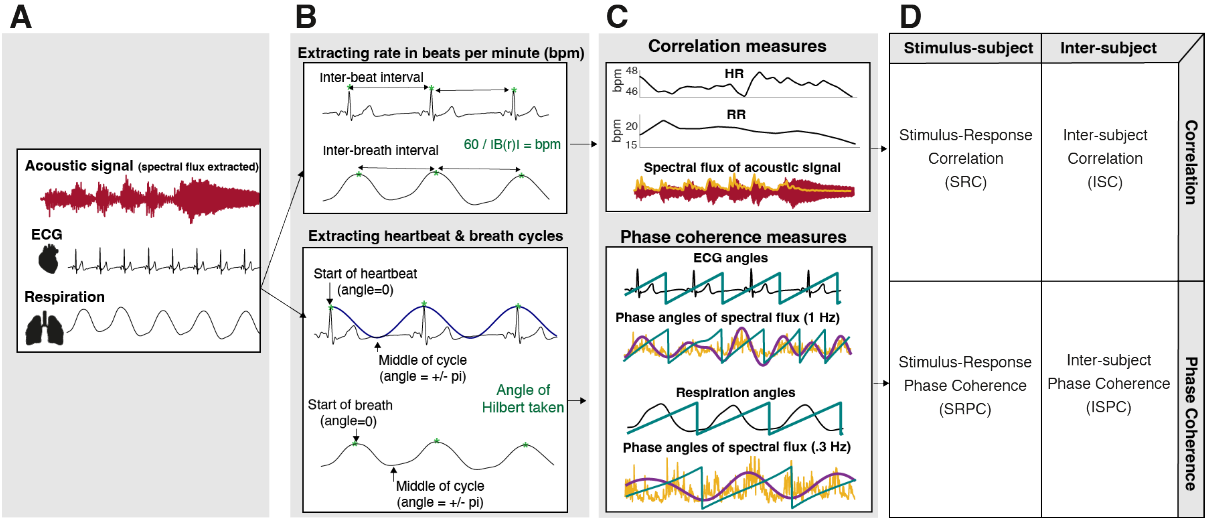


Figure 1. Outline of the experiment and analysis pipelines. **Panel A** shows the study design: audiences were presented with music pieces in audio-visual (AV, orange box) and audio-only (AO, blue box) conditions, while heart (ECG), respiration, and acoustic signal of the music (maroon) were continuously recorded. **Panel B** shows how beats per minute (bpm) were extracted for correlational measures (upper panel), while 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 filtered at the audience's average heart frequency (1 Hz) and respiration frequency (.3 Hz) (from acoustic signal). **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.

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[52] 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 example, if a T wave in the ECG was accidentally identified as a R wave. Data that were too noisy to identify clear R waves and respiration peaks were rejected from further analysis (ECG = 14%, respiration = 7%).

Synchrony analyses

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

Correlational synchrony: ISC and SRC

Heart rate (HR) and respiration rate (RR) were calculated by obtaining the differential timing between peaks, i.e., inter-beat intervals (IBI) for ECG, and inter-breath intervals-(IBrI) for respiration (see Figure 1B). IB(r)Is were then converted to beats per minute (bpm) and interpolated at the original sampling rate to obtain instantaneous HR and RR, which were downsampled to 20 Hz[53]. SRC and ISC were then calculated across sliding windows. Although ISC and SRC measures have previously been calculated within 5 second sliding-windows for EEG[4,43], peripheral responses of heart and respiration are slower, where a typical evoked response can be up to 10 seconds[13,54,55]. Therefore, we chose a 10 second sliding-time window with a one second overlap. 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

To assess whether heartbeats and breaths aligned with the spectral flux rhythms in the music, cardiorespiratory signals were transformed into the phase domain. Although this has been done by assessing cardiorespiratory signals within stationary frequency bands, from which phase angles are calculated[28,56], here, our aim was to assess changes of actual heartbeat and breath cycles that vary over time. To account for this variation, phase was calculated from cycle peaks (see Figure 1B, green dots marking the cycle peaks). First, a continuous sinusoidal wave was fitted to each IB(r)I of detrended and normalised ECG and respiration signals based on[57], 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.

SRPC of heartbeats was calculated by assessing coherence between the phase angles of heartbeat cycles with the spectral flux phase angles corresponding to heart (1.01 Hz), while SRPC of respiration was calculated by assessing coherence between phase angles of respiration cycles with spectral flux phase angles corresponding to respiration (0.30 Hz) frequencies (see Figure 1C). Coherence was calculated based on the following formula based on[58]:

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

where n is the number of phase signals, e^{ik} is Eulers formula (i.e., the complex polar representation of phase angle k for signal r at time point t). ISPC was assessed in a pairwise fashion between heartbeat cycle and respiration cycle phase angles of all possible participant pairs within one concert using the same coherence formula as above.

Adjusting for lags

To account for lags in responses to the stimuli for SRC and SRPC, we adjusted data with a stable lag to optimally align the stimulus and the corresponding responses[59]. This was done within the first 10 seconds after stimulus onset as this initial onset likely evokes the most reliable response. We calculated SRC and SRPC at lags up to each individual's mean interbeat interval for heart measures (on average 990 ms) or inter-breath interval for respiration measures (on average: 3400 ms), as it might take up to one heartbeat/respiration cycle to begin responding to a stimulus[54] . The positive correlations and phase coherence values were obtained at the optimal lag[60]. 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 stable lags were applied to all corresponding correlation and phase coherence heart and respiration stimulus-response pairings.

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

Continuous synchrony measures for HR (SRC-HR, SRPC-HR, ISC-HR, ISPC-HR) and RR (SRC-RR, SRPC-RR, ISC-RR, ISPC-RR) 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. Nonetheless, 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[54,55], as well as any anticipatory effects at musical events[61].

Significance-testing

Regarding the first hypothesis, we tested the assumption that responses time-locked to a stimulus should evoke similar responses across participants[62]. Therefore, the control condition was time-'unlocked' data, created 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 shifted data sets. Following Harding et al.[63], one true data value and one permuted value (the average) per stimulus and participant was compared by a Wilcoxon signed rank test (data were not normally distributed) across aggregated synchrony values. As we had eight synchrony measures (four synchrony measures each for heart and respiration measures), we corrected the alpha level to .05 / 8 = .006.

AV versus AO: time-averaged synchrony

For the second hypothesis, time-averaged synchrony measures were compared between AV and AO. Self-reported engagement (translated as 'absorption' in a previous study[9], but used synonymously with engagement, see also[20]) 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[64,65]. If these maximal models generated convergence and/or singularity fit errors, models were simplified following recommendations[64,66]. As output from models generating errors should not to be trusted[64], 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[67,68]. Significance values and effect sizes were obtained from the *tab_model* function from *sjPlot* package[69]. As a recommended sanity check, we also ran conventional *t*-tests alongside LMMs[70]. This was done by a one sample *t*-test (or Wilcoxon if data were not normally distributed), run on the difference between AO and AV values.

AV versus AO: epoched

To test the third hypothesis, synchrony trajectories at section boundaries were assessed between modalities across time windows. Synchrony trajectories in the 30-second epochs centred around section boundaries were split into smaller time windows[55,71], yielding five 6-second time windows. For each synchrony measure, linear mixed models (LMMs) with fixed effects of modality and time window were constructed. We added random intercepts for concert, piece, and participants (participants nested within concerts; participant and piece as crossed effects) and a random slope for participants. As above, models generating errors were simplified; error-free models are reported (full models in Supplementary Materials). Estimated marginal means (using *emmeans* package[72]) were used to check pairwise comparisons with Bonferroni adjustments. As we predicted that synchrony would occur at section boundaries, we expected the time window centred at the section boundary to have significantly higher synchrony values than time windows prior to and after a section boundary. In case synchrony was high, we investigated what responses were becoming synchronized. Thus, raw HR and RR data at section boundaries were taken (i.e., HR and RR as in Figure 1, Panel B). Two LMMs were constructed, each for aggregated HR and RR values, with fixed effects of time window and random effects of participant id, piece, and concert, as above.

3. Results

Synchrony compared to circular-shifted control

SRC-HR, ISC-HR, and ISC-RR were significant compared to circular-shifted control (see Table 1) with a large effect size (Cohen 1988). These results were also replicated when values were arbitrarily averaged in 30 second bins (see Supplementary Table 2). We therefore removed SRC-RR and all phase-based measures from further analyses between modalities.

Table 1. Results from Wilcoxon signed-rank tests between time-averaged synchrony across piece sections calculated with original (time-locked) data and control (circular-shifted) data.

	Heart			Respiration		
	Wilcoxon signed rank statistic	*p*	Effect size	Wilcoxon signed rank statistic	*p*	Effect size
SRC	**2**	**< .001**	**0.863**	104	.197	0.268
ISC	**14**	**< .001**	**0.786**	**11**	**< .001**	**0.811**
SRPC	79.5	.078	0.371	174	.509	.140
ISPC	132	.879	0.035	122	.432	0.163

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 2). This effect for ISC-HR occurred in both error-free and full models (see Supplementary Table 6). When checking conventional Wilcoxon tests, 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], $p = .464$, $R^2_{(fixed)} = 0.005$).

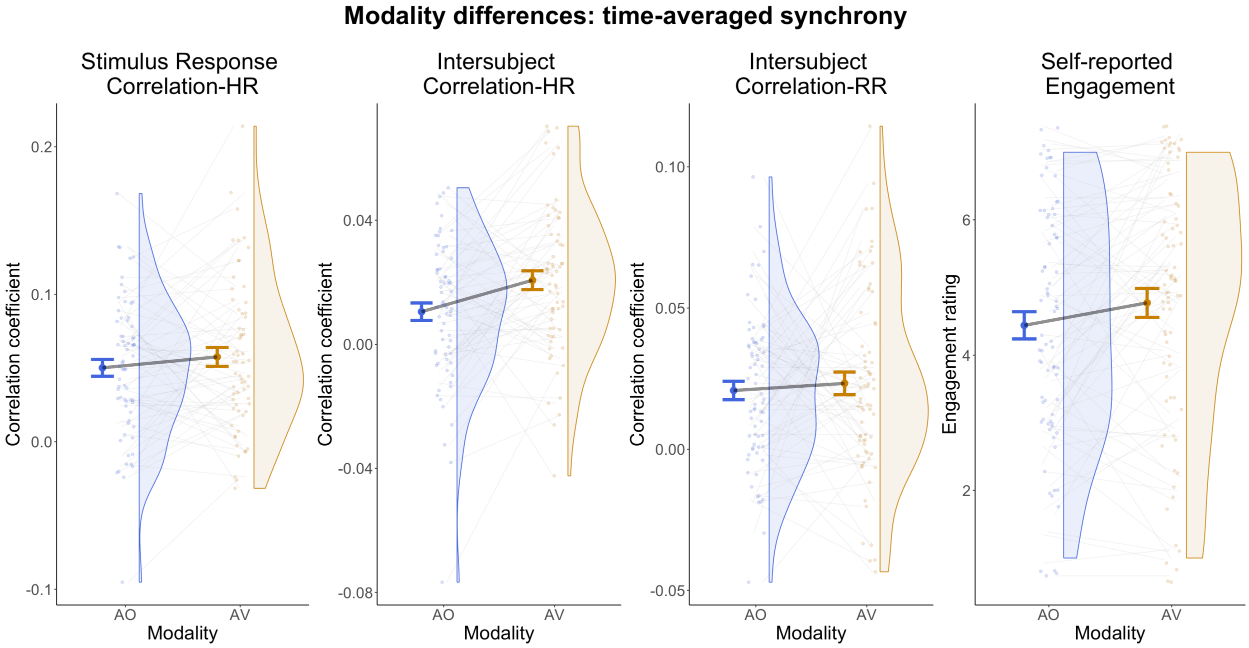


Figure 2. Modality differences for time-averaged synchrony measures.

Table 2. 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, ISC-HR was significantly higher in the AV condition (see Figure 3 and Table 3, see Supplementary Table 7 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 Table and Figure 3).

In assessing the interaction between time window and modality, Figure 3 shows that synchrony was higher just before (window -5) and at section boundaries (window 0) in the AV condition for ISC-HR. This latter effect was further confirmed by LMMs and by significant pairwise comparisons (Bonferroni corrected) for ISC-HR (Table 4).

Upon checking actual HR and RR responses at these time points (cf.[13]), 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 5 and Supplementary Table 8). Although RR decelerated and accelerated similarly to HR (see Figure 4), these were not significant across time windows.

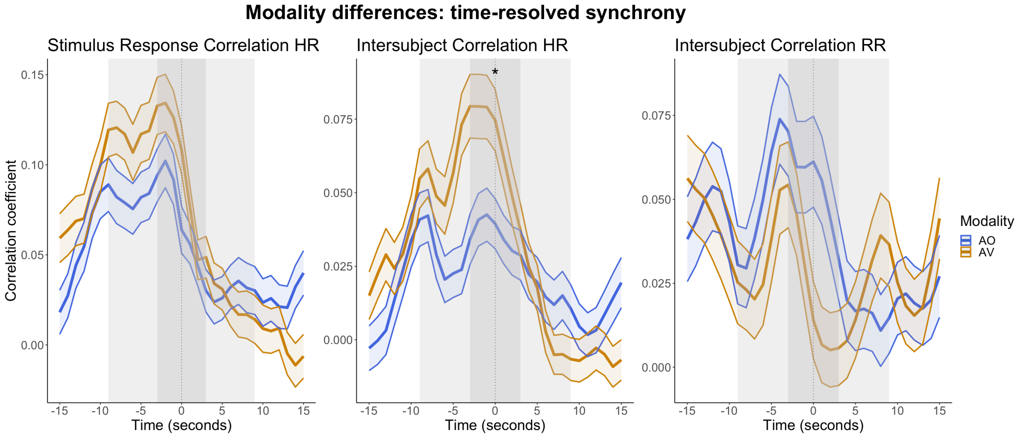


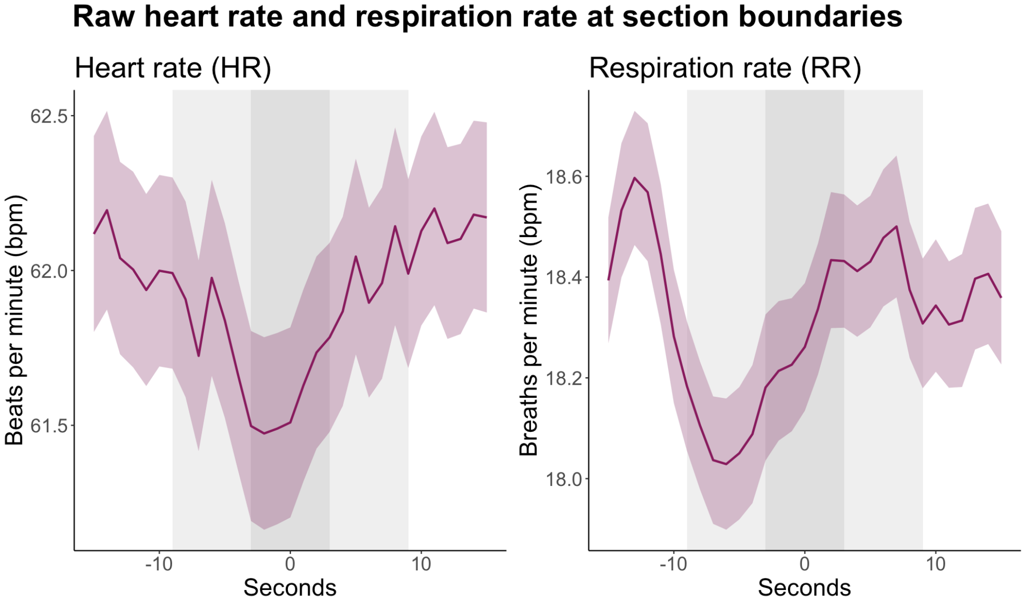
Figure 3. Synchrony values across time centred at section boundaries. *represents a significant modality effect within that time window.

Table 3. Linear mixed-effects models for synchrony with predictors of time window and modality.

	SRC-HR			ISC-HR			ISC-RR		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	0.07	0.04 – 0.09	**<0.001**	0.04	0.02 – 0.05	**<0.001**	0.05	0.03 – 0.07	**<0.001**
mod [AV]	0.03	-0.00 – 0.06	0.093	0.03	0.01 – 0.05	**0.002**	-0.03	-0.06 – -0.00	**0.029**
win-10	-0.01	-0.05 – 0.02	0.339	-0.02	-0.04 – -0.00	**0.018**	-0.00	-0.03 – 0.02	0.784
win-5	0.01	-0.02 – 0.04	0.363	-0.01	-0.02 – 0.01	0.574	-0.00	-0.03 – 0.02	0.793
win [5]	-0.04	-0.07 – -0.01	**0.016**	-0.02	-0.04 – 0.00	0.072	-0.04	-0.06 – -0.01	**0.009**
win [10]	-0.04	-0.07 – -0.01	**0.009**	-0.03	-0.04 – -0.01	**0.008**	-0.03	-0.06 – -0.01	**0.015**
modAV:win-10	-0.00	-0.05 – 0.04	0.917	-0.02	-0.04 – 0.01	0.229	0.03	-0.01 – 0.07	0.155
modAV:win-5	0.01	-0.04 – 0.05	0.811	-0.01	-0.03 – 0.02	0.596	0.01	-0.03 – 0.05	0.499
mod [AV] × win [5]	-0.04	-0.08 – 0.01	0.102	-0.04	-0.07 – -0.02	**0.002**	0.04	0.00 – 0.08	**0.035**
mod [AV] × win [10]	-0.05	-0.10 – -0.01	**0.019**	-0.05	-0.07 – -0.02	**0.001**	0.04	-0.00 – 0.08	0.055
Random Effects									
σ^2	0.01			0.00			0.01		
τ_{00}	$0.00_{\text{mac:conc}}$			$0.00_{\text{mac:conc}}$			$0.00_{\text{mac:conc}}$		
	0.00_{conc}			0.00_{conc}					
τ_{11}	$0.00_{\text{mac.modAV}}$			$0.00_{\text{mac1.modAO}}$					
	$0.00_{\text{mac1.modAO}}$			$0.00_{\text{mac2.modAV}}$					
	$0.00_{\text{mac2.modAV}}$								
ρ_{01}									
ρ_{01}									
ICC	0.06			0.06			0.03		
N	2_{conc}			2_{conc}			16_{mac}		
	16_{mac}			16_{mac}			2_{conc}		
Observations	645			645			690		
Marginal R^2 / Conditional R^2	0.120 / 0.177			0.116 / 0.167			0.025 / 0.056		

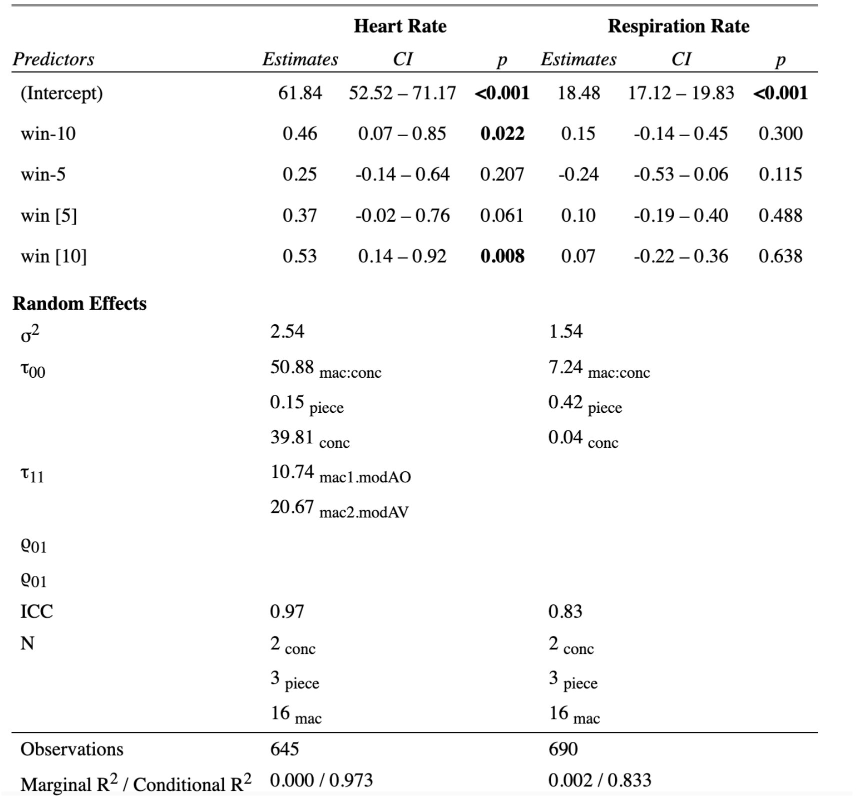
388 **Table 4.** Pairwise comparison (Bonferroni corrected) between modality (AO-AV) in each time
389 window.

	Contrast AO-AV time window	estimate	SE	df	T ratio	*p*
SRC-HR	Window -10	-0.025	0.016	190	-1.533	0.635
	Window -5	-0.033	0.016	190	-2.006	0.232
	Window 0	-0.027	0.016	190	-1.676	0.477
	Window 5	0.009	0.016	190	0.576	1.000
	Window 10	0.025	0.016	190	1.554	0.609
ISC-HR	Window -10	-0.0147	0.010	219	-1.483	0.698
	Window -5	-0.0241	0.010	219	-2.421	0.081
	Window 0	**-0.0315**	**0.010**	**219**	**-3.163**	**0.009**
	Window 5	0.0118	0.010	219	1.187	1.000
	Window 10	0.0142	0.010	219	1.424	0.778
ISC-RR	Window -10	0.002	0.014	659	0.172	1.000
	Window -5	0.017	0.014	659	1.228	1.000
	Window 0	0.030	0.014	659	2.185	0.146
	Window 5	-0.011	0.014	659	-0.805	1.000
	Window 10	-0.007	0.014	659	-0.537	1.000



390
391 **Figure 4.** Trajectories of actual HR and RR data at section boundaries.

Table 5. 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 $_{\text{piece}}$			0.42 $_{\text{piece}}$		
	39.81 $_{\text{conc}}$			0.04 $_{\text{conc}}$		
τ_{11}	10.74 $_{\text{mac1.modAO}}$					
	20.67 $_{\text{mac2.modAV}}$					
ρ_{01}						
ρ_{01}						
ICC	0.97			0.83		
N	2 $_{\text{conc}}$			2 $_{\text{conc}}$		
	3 $_{\text{piece}}$			3 $_{\text{piece}}$		
	16 $_{\text{mac}}$			16 $_{\text{mac}}$		
Observations	645			690		
Marginal R^2 / Conditional R^2	0.000 / 0.973			0.002 / 0.833		

4. Discussion

An important aspect that makes music concerts engaging might be seeing musicians perform[6,7]. The current study investigated whether concert audiences' engagement – here assessed with self-reports and cardiorespiratory synchrony measures – is higher during audio-visual (AV) than audio-only (AO) music performances. Participants were presented with AO and AV performances of Western classical music in a concert setting while cardiorespiratory measures were continuously measured. In comparing control (circular-shifted) data to true synchrony measures, we show that heart and respiration rate (*speed* of heartbeats and breathing in bpm) correlate with music. However, musical rhythms did not align heartbeat rhythms nor breathing rhythms. 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 correlates with spectral flux

Assuming that non-time locked responses would not be correlated, as nothing 'couples' them[12,14,73], the significance of cardiorespiratory synchrony measures in general were assessed against a time 'unlocked' (circular shifted) control data set[14]. SRC-HR was significant, 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 significant, indicating that acceleration/deceleration of HR and RR correlated across audience members experiencing music simultaneously. The ISC significance found in this current study supports previous research showing that ISC of HR[12,73] and EEG[4] is evoked when participants simultaneously experience the same auditory stimuli, such as speech and music. However, the current results that ISC-RR was also significant contrasts findings of Madsen and Parra[73], 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[73] 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 significant. 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[40–42], this contrasts previous findings that breathing aligns with musical beats[38,39]. There are a few potential reasons for this discrepancy between results. First, Etzel et al.[38] and Haas et al.[39] 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.[39] Table 1, Figure 1 and 2; Etzel et al.[38] 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[15]) 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[38,39], 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 stimulus length; the current study has music stimuli at least seven minutes long. In Haas et al.[39], silent and metronome conditions are described as being five minutes long, thus we assume the music condition was also five minutes long (musical excerpts length is unclear, as there are no details of clip or music recording length). Etzel et al.[38] used stimuli between 1 and 3 minutes. Taken together, perhaps better synchronization occurs during shorter musical clips (as is reported in Etzel et al.[38], and as is assumed in Haas et al.[39]), but breathing alignment might not last for longer time frames of over six minutes. Relatedly, a fourth 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[7,74], where attention may fluctuate a lot more between the music and other features. Thus, although neural[59,63] and respiratory rhythms[38,39] 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[29]. Indeed, phase synchrony between participants has previously been found in dancing[75,76] and musical ensembles[27,28,56,77] where there is a clear shared goal of a coordinated performance (see [29,30]). Typical Western classical concert etiquette calls for enhanced focus on the performance, thus reducing body movements and interaction between audience members[7]. Indeed, ratings for 'urge to move to the music' (see[9]) were heavily skewed to low ratings; even testing significance of phase coherence in audience members who only had high 'urge to move' ratings did not yield significant results (see Supplementary Table 5). Additionally, videos showed that the audience members respected this etiquette and generally did not move too much (see[9]). Therefore, we suggest that phase synchrony – typically seen in interactive settings[29] – may not occur in non-interactive settings.

Time-averaged synchrony may not relate to naturalistic engagement

Next, synchrony measures were compared between AV and AO conditions. ISC-HR was significantly greater in AV performances. However, this result was not shown to be robust with an ANOVA sanity check. Although our data are more appropriate for LMMs due to naturalistic paradigm and physiological data[65], the fact that the significance is not replicated in a simpler Wilcoxon test thus allows only cautious interpretation[70]. Another reason to be cautious is that – contrary to what was expected – ISC-HR was not significantly related to engagement that was self-reported after each piece. Our current findings contrast findings from Ki et al.[19] who found that engagement and neural ISC were stronger in AV compared to AO stimuli. Our finding also contrasts studies showing that HR[12] and neural[14] synchrony can reflect engagement. One explanation for the discrepancy in results is that these previous studies had explicit attend/non-attend conditions, where participants were instructed to count backwards in the non-attend conditions. Here, participants experienced concert performances naturally with no formal instruction. This suggests that while ISC (of HR) could be a marker in clearly defined engagement conditions, it might not necessarily be generalisable to less controlled settings. As mentioned above, another explanation for differences in results could be the stimuli length. Compared to the music pieces presented here (at least 7 minutes long), stimuli were shorter for Ki et al.[19] (6 minutes), Pérez et al.[12] (1-minute excerpts), and Madsen et al.[14] (90 and 60 seconds). With longer stimuli, engagement is not as sustainable and likely to shift. Inter- and intra-individual differences may influence synchrony and engagement[78,79], and could be a potential venture for future studies. Overall, we show time-averaged synchrony does not reliably reflect overall self-reported engagement measurements. With this in mind, we anticipated results for time-resolved synchrony might yield clearer results.

Time-resolved synchrony reveals synchronized heart rate orienting response

Engagement with music typically fluctuates across time[43,44]. Thus, we assessed synchrony changes between modalities on a time-resolved scale. We 'zoomed in' on synchronized responses at salient moments in the music, here defined as section boundaries in music (see Supplementary Table 1), which are important structural locations in music. The current results show that both ISC-HR and SRC-HR increased at such section boundaries, supporting and extending previous work on ISC[13,43] to SRC. Although we did not have time-resolved self-reports to assess the relation of synchrony to engagement, we assessed the continuous raw HR and RR. We show that synchronized responses reflected significant deceleration-acceleration patterns in HR and RR. Such patterns have been long associated with a typical orienting response, which marks an increase in attentiveness[80] to anticipate and perceive events[54,81–83]. Thus, such increases of cardiorespiratory synchrony at section boundaries suggest increased engagement with music at these time points, related to auditory cues indicating a section is ending, such as a cadential ending (i.e., harmonic closure in Bach and Beethoven) and a slight slowing down or change in tempo and/or material (see Supplementary Table 1). This physiological synchrony at salient music events suggests a bottom-up capturing of engagement during a shared musical experience[29].

Importantly, Figure 3 shows that ISC-HR significantly increased at section boundaries more so in the AV condition. This increase suggests that visual information further increased engagement at section boundaries. Indeed, previous performance studies have found that certain gestures cue and anticipate structurally important locations, such as leaning forward towards a cadence (harmonic closure)[84], increased movement amplitude[85], or finishing 'flourish' gestures[86] as well as deep breaths and sweeping motions prior to onset of new phrases[87]. Indeed, videos show that the pianist tended to lean forward at section boundaries. Therefore, a combination of audio and visual aspects likely increases engagement with music at a local level, as shown most robustly with ISC-HR.

5. Limitations

Although we show greater synchrony in the AV (compared to AO) condition, it is unclear whether this synchrony increase might simply be related to increased information processing of both visual and audio information, compared to simply audio information. For example, synchrony between stimulus features and RR (SRC-RR) was not significant, but synchrony of RR across audiences (ISC-RR) was significant. This suggests that the audiences' respiration synchrony was rather driven by something else than the acoustic signal, perhaps the musician movements (see[9]). Indeed, we saw a slightly larger effect size of SRC-RR between true and control data in the AO than the AV condition (see Supplementary Table 2). Future work would therefore need to assess responses not only to auditory features, but also visual features, such as the quantity and speed of movements[88] to shed more light on how much variance in the audience experience can be explained by auditory versus visual features. Additionally, we did not test a visual only condition which would allow us to disentangle the difference between added information processing and actual engagement.

6. Conclusion

The current study explored engagement between audio-visual and audio-only music performances in a Western classical concert setting, using cardiorespiratory synchrony as a measure of engagement. Extending studies investigating neural synchrony[4,11], this study contributes to a growing literature on measures of peripheral physiological synchrony[12,44,73] in music concert settings[3,10,13]. We show that seeing musicians perform can heighten audience engagement, though this might be context-dependent, and is best assessed by time-resolved (rather than time-averaged) inter-subject correlation of HR (i.e., synchrony of heartbeat speed). We found no significance of phase synchrony (i.e., rhythmic alignment of heartbeats/respiration with other audience members/music). However, questions arose whether phase coherence might become more relevant in contexts where listeners can move and interact, for example in pop and rock concerts. More studies from wider-ranging genres and contexts are needed to further understand synchrony dynamics in real-world music contexts.

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. Lamont, A. University students' strong experiences of music: Pleasure, engagement, and meaning. *Music. Sci.* **15**, 229–249 (2011).
3. 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.
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. 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).
6. 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).
7. Wald-Fuhrmann, M. *et al.* Music Listening in Classical Concerts: Theory, Literature Review, and Research Program. *Front. Psychol.* **12**, (2021).
8. 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).

9. 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).

10. 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).

11. 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).

12. Pérez, P. *et al.* Conscious processing of narrative stimuli synchronizes heart rate between individuals. *Cell Rep.* **36**, 109692 (2021).

13. Czepiel, A. *et al.* Synchrony in the periphery: inter-subject correlation of physiological responses during live music concerts. *Sci. Rep.* **11**, 22457 (2021).

14. 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).

15. 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).

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

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

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

19. 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).

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

21. 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).
22. 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).
23. Dikker, S. *et al.* Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Curr. Biol.* **27**, 1375–1380 (2017).
24. 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).
25. Jones, M. R. & Boltz, M. Dynamic attending and responses to time. *Psychol. Rev.* **96**, 459–491 (1989).
26. Large, E. W. & Snyder, J. S. Pulse and Meter as Neural Resonance. *Ann. N. Y. Acad. Sci.* **1169**, 46–57 (2009).
27. Gugnowska, K. *et al.* Endogenous sources of interbrain synchrony in duetting pianists. *Cereb. Cortex* **32**, 4110–4127 (2022).
28. Müller, V. & Lindenberger, U. Cardiac and Respiratory Patterns Synchronize between Persons during Choir Singing. *PLOS ONE* **6**, e24893 (2011).
29. Dumas, G. & Fairhurst, M. T. Reciprocity and alignment: quantifying coupling in dynamic interactions. *R. Soc. Open Sci.* **8**, 210138 (2021).
30. 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).
31. Tschacher, W. *et al.* Physiological synchrony in audiences of live concerts. *Psychol. Aesthet. Creat. Arts* **17**, 152–162 (2021).
32. 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.
33. Klimesch, W. The frequency architecture of brain and brain body oscillations: an analysis. *Eur. J. Neurosci.* **48**, 2431–2453 (2018).

34. 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).

35. Palumbo, R. V. *et al.* Interpersonal Autonomic Physiology: A Systematic Review of the Literature. *Personal. Soc. Psychol. Rev.* **21**, 99–141 (2017).

36. Al, E. *et al.* Heart–brain interactions shape somatosensory perception and evoked potentials. *Proc. Natl. Acad. Sci.* **117**, 10575–10584 (2020).

37. Grund, M. *et al.* Respiration, Heartbeat, and Conscious Tactile Perception. *J. Neurosci.* **42**, 643–656 (2022).

38. 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).

39. Haas, F., Distenfeld, S. & Axen, K. Effects of perceived musical rhythm on respiratory pattern. *J. Appl. Physiol.* **61**, 1185–1191 (1986).

40. Ellis, R. J. & Thayer, J. F. Music and Autonomic Nervous System (Dys)Function. *Music Percept.* **27**, 317–326 (2010).

41. Koelsch, S. & Jäncke, L. Music and the heart. *Eur. Heart J.* **36**, 3043–3049 (2015).

42. 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).

43. 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).

44. 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).

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

46. 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).

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

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

49. 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.

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

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

52. 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).

53. 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).

54. Bradley, M. M. & Lang, P. J. Affective reactions to acoustic stimuli. *Psychophysiology* **37**, 204–215 (2000).

55. 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).

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

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

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

59. 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).

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

61. 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).

62. 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).

63. 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).

64. 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).

65. 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.

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

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

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

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

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

71. 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).

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

73. Madsen, J. & Parra, L. C. Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus* **1**, pgac020 (2022).

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

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

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

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

78. 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).

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

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

81. 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).

82. 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).

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

84. 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).

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

86. 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).

87. 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.

88. 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).