Swinging, Fast and Slow: Multiscale Synchronisation Dynamics Reveals the Impact of an Improvisatory Approach to Performance on Music Experience

Takayuki Nozawa*[1,2], Madalina I. Sas*[3], David Dolan[4], Hardik Rajpal[3], Fernando E. Rosas[3,5], Christopher Timmermann[3], Pedro A.M. Mediano*[3], Keigo Honda[1], Shunnichi Amano[1], Yoshihiro Miyake[1], and Henrik J. Jensen[3,1]

[1]Tokyo Institute of Technology, Tokyo, Japan
[2]University of Toyama, Toyama, Japan
[3]Imperial College London, London, UK
[4]Guildhall School of Music and Drama, London, UK
[5]University of Sussex, Brighton, UK

ABSTRACT

Experiences of collective creative activities play an essential role in human societies, yet these experiences are particularly hard to capture, making their scientific scrutiny extremely challenging. Here we investigate the experience of audience members during a musical concert associated with collective improvisation by analysing the audience's subjective reports and movement patterns. Our results show that performance with improvisational elements affect movement synchronisation dynamics between performers and audience members differently at different timescales, which are predictive of changes in the subjective perception of music. These results provide a first step towards the quantification of some of the fundamental aspects of these collective experiences. Moreover, the reported findings shed new light on the relevance of the often-neglected multiscale coordination between audiences and performers, and explains how this rich tapestry of physical behaviour is connected with the quality of the collective subjective experience.

1 Introduction

Collective activities involving shared experiences are ubiquitous in human culture, and are believed to play crucial roles for strengthening social bonds, sense of group belonging, and social cohesion[1,2]. Empirical investigations have shown that interpersonal synchronization of physical activity between humans is strongly associated with collective subjective experience[3–5], such as a feeling of unity and perceived social bonding[6] and the group experience in shared social and ritual celebrations[7,8], but also with positive objective outcomes in various types of verbal and non-verbal interaction[9,10].

Among collective activities, music making and listening occupies an important place in all known human societies[11,12], and often reveals, even within unique, cultural-specific approaches, universal elements of expressing and perceiving emotional cues[13]. As a group activity, music making requires high levels of empathy[13–15], coordination, and synchrony[16], which support the emergence of leadership[17], improvisation[18,19], and group states of flow[20,21], and moreover, is known to engage audiences in a participatory, reciprocal relationship with the performers[22,23].

Within musical praxis, musical improvisation is a highly complex creative skill as well as social process which requires years of training and special conditions to emerge[24], as it involves risk-taking within given structures and dealing with the unknown in real-time[25]. Improvisation has universal appeal manifested in different forms across cultures and musical genres[26], yet from the early 20th century and until recently, the mainstream of Western classical music performance has been largely dominated by notation-only based performance: following the score strictly and accurately and aiming for the best and most expressive performance while avoiding spontaneous, improvisatory elements[27].

On the other hand, a more improvisatory approach is regaining attention[25,28–30], characterised by spontaneity and risk-taking, which allows performers to deviate from the written text, according to the stylistic language in question, in an unrehearsed coordination with the other ensemble partners, thus emphasising the differences between the notion of music as performance versus music as text[25,31]. Importantly, a number of studies place the omission of improvisation from classical music performance under question, as this practice has been shown to enhance the musical experience of both performers and audiences[32]. We refer to the two performance modes described above as *Strict*, and respectively, *Let-go*[32].

In the context of Western classical music, coordination of physical movements between performers has been investigated[33,34], yet previous studies have rarely investigated the possibility or meaning of physical synchronization between music performers and their audiences, or the synchrony among listeners themselves. Recently, in a wider range of situations, it has been shown that audiences synchronise on physiological markers such as heart and respiration rate[35–39], but in matters of physical activity they are still mostly assumed to be passive and static[40,41].

In this paper, we challenge this assumption and explore the effect of innovative improvisational attitude to classical music performance on the collective motion of a seated audience. For this purpose, we developed a concert-experiment where two classical repertoire pieces where played twice, each in the two performance modes.

Albeit psychological studies of music and improvisation tend to focus on short segments of a few measures being performed by many different musicians[42,43], the current experimental design allows us to address in full depth the improvisational character of performance and study phenomena emerging at the macroscopic musical scale.

During the experiment, we measured the spontaneous movement of the audience. Although maybe subtle, we expect their physical activity to be linked to their experience of the music, as well as to the movements of performers. Thus, we hypothesise the degree of physical synchrony in the audience can differentiate the way they perceive the different performance modes. Specifically, we hypothesise:

1. *Let-go* would be perceived by the audiences as more innovative and improvisatory than *Strict* performances. This is in line with previous work[30,32], and here we aim to confirm the results with a larger group;

2. *Let-go* would induce higher physical synchrony (with performers and within the audience) than *Strict* performances, and also the temporal variability of the degree of physical synchrony would be associated with the audience's innovative experience. This is in line with the indications that meta-stability of synchrony —dynamic in and out of synchronous mode— is a marker of adaptability[44];

3. the effects on the audiences' perception and physical synchrony would be positively associated.

To deepen our understanding of these elements, we also explore the role of psychological absorption in a subject's positive experience. Finally, we examine the role of visual cues by studying a subgroup of listeners who are blindfolded.

2 Results

A concert of classical music was organised where a string quartet performed each of the two pieces (Mozart's string quartet *KV. 421 no. 15* (exposition of the first movement) and Haydn's *Op. 76 no. 1* (third movement)) twice: once in *Strict* mode (characterised by aiming at following the written score strictly and avoiding any expressive gesture not directly indicated by the written text), and once in *Let-go* mode, (a 'beyond text' interpretation with real-time improvisatory elements)[30,32]. Comparing *Let-go* with *Strict* performances of the same repertoire piece by the same performers allows us to experimentally manipulate collective musical experiences while controlling other factors. Both versions of each piece were played one after the other using a randomised ordering. The concert was attended by 42 audience members. Questionnaire responses and movement data were collected in order to investigate how the two performance modes affect the audience's experience. Details of the experimental design can be found in Section 4.

2.1 Performance ratings

As a first step in our analysis, we investigated the subjective experience of audience members as reflected by questionnaire responses given after each pair of performances. Questionnaire scores were analysed via multilevel models that included experimental variables as fixed effects and participant IDs as random effects (see Section 4.4).

Results reveal that the audience was receptive to the performance mode, rating the *Let-go* performances to be significantly more Improvisatory ($p = 0.001$), Innovative ($p = 0.043$), and Risk-taking ($p = 0.004$) than the *Strict*. This is in accordance with our hypothesis 1 and previous work[32]. In contrast, no significant differences were observed regarding how Musically Convincing and Emotionally Engaging both renditions were.

To quantify to what extent these ratings reflect either a unified factor or different aspects of the audience's experience, we performed a principal component analysis (PCA) to evaluate how much variance in questionnaire scores can be explained as being part of a single factor. Results show that the first principal component (PC1) — mainly consisting of the Improvisatory, Innovative, Risk-taking, and Emotionally Engaging items — accounts for 43.9% of the variance (see Fig. 1). Furthermore, the value of PC1 is significantly higher for the *Let-go* than the *Strict* mode ($p = 0.018$), supporting the idea that it captures a principal axis that differentiates between performance modes.

To evaluate the potential effect of visual cues on the difference of experience between *Strict* and *Let-go*, 13 audience members were blindfolded. Incorporating the blindfolding factor in our multilevel models did not show a significant main

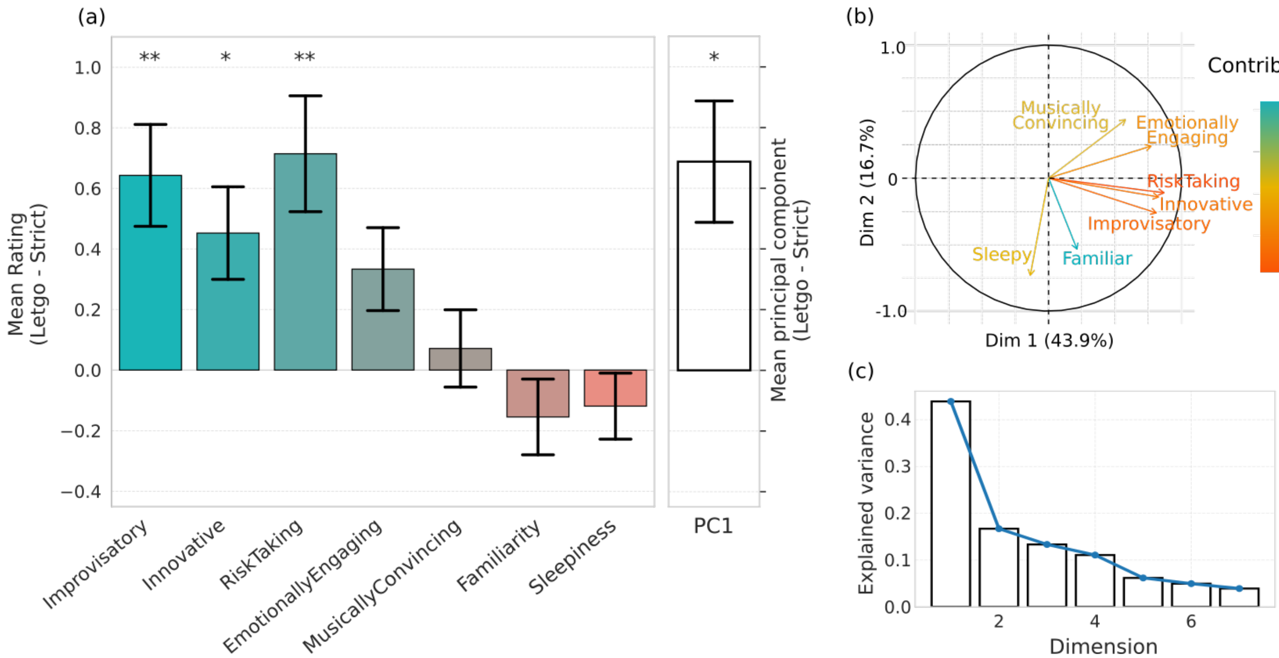

Figure 1. (a) Difference in the audience's perception of the two repertoire works (by Mozart and Haydn) between the *Let-go* and *Strict* performance modes, including all 7 performance ratings and the first principal component (PC1). Error bars show standard error of the mean individual differences between the two modes of performance. Audiences perceived the *Let-go* mode of performance as significantly more Improvisatory, Innovative, and Risk-Taking. (b) Biplot of the contribution of the first two principal components towards the 7 audience ratings. (c) Scree plot showing the amount of explained variance by the 7 principal components.

effect of sight nor significant interactions with performance mode. A significant 3-way interaction was observed for the PC1, Improvisatory, and Risk-taking ratings (See section B.1 in the Supplementary material for more details).

As an additional control, we investigated if the difference in the ratings between performance modes could be related to individual differences in psychological absorption in the audience members, as this trait has previously been linked to higher engagement with music[45]. Results show that absorption has a positive effect on the audience ratings in general, but not on differentiating the mode of performance, and no interaction with the mode of performance. We observe a significant effect on the audience's Improvisatory ($p = 0.010$), Innovative ($p = 0.001$), Risk-Taking ($p = 0.009$) and Emotionally Engaging ($p < 0.001$) ratings, with the strongest effect in the last variable, suggesting that higher absorption is indeed associated with a more positive emotional experience and a higher likelihood to perceive the piece as Improvisatory, regardless of the mode of performance. It is insightful to observe the Musically Convincing rating and absorption are not related, and also that higher absorption subjects are likely to find the piece more Familiar. (See section B.2 in the Supplementary material for more details.)

Finally, to explore the relationship between quantitative and qualitative aspects of musical performance, we also gathered subjective accounts on the performance from the musicians, who consistently reported that they failed to achieve the ideal *Let-go* mode in the first piece (Mozart). Interestingly, we find that the differences in PC1 and the Improvisatory, Innovative and Risk-taking ratings in the Mozart pieces are weaker than those in the Haydn pieces, revealed as significant interactions between the performance mode and composition factors by the multilevel models (see appendix B.1 for details). These results indicate that the audience perceived the differences between the *Let-go* and *Strict* performances of Haydn's composition, but they were not as sensitive to the difference between *Let-go* and *Strict* performances of Mozart's composition. Importantly, this is in accordance with the musicians' report of their own performance.

2.2 Physical synchrony

The second step in our analysis is to investigate the movement patterns of audience members, in particular the synchrony among listeners and with the performers. For this purpose, we carried out quantitative analyses using accelerometer data collected from the audience and performers.

We start with the degree of synchronisation across the entire spectrum of physical movement, considering the synchrony of movements between audience members (A-A sync) and also between performers and audience (P-A sync) over a wide range of timescales (Fourier periods). For this, we employ the wavelet transform coherence (WTC) on the time-frequency space[46],

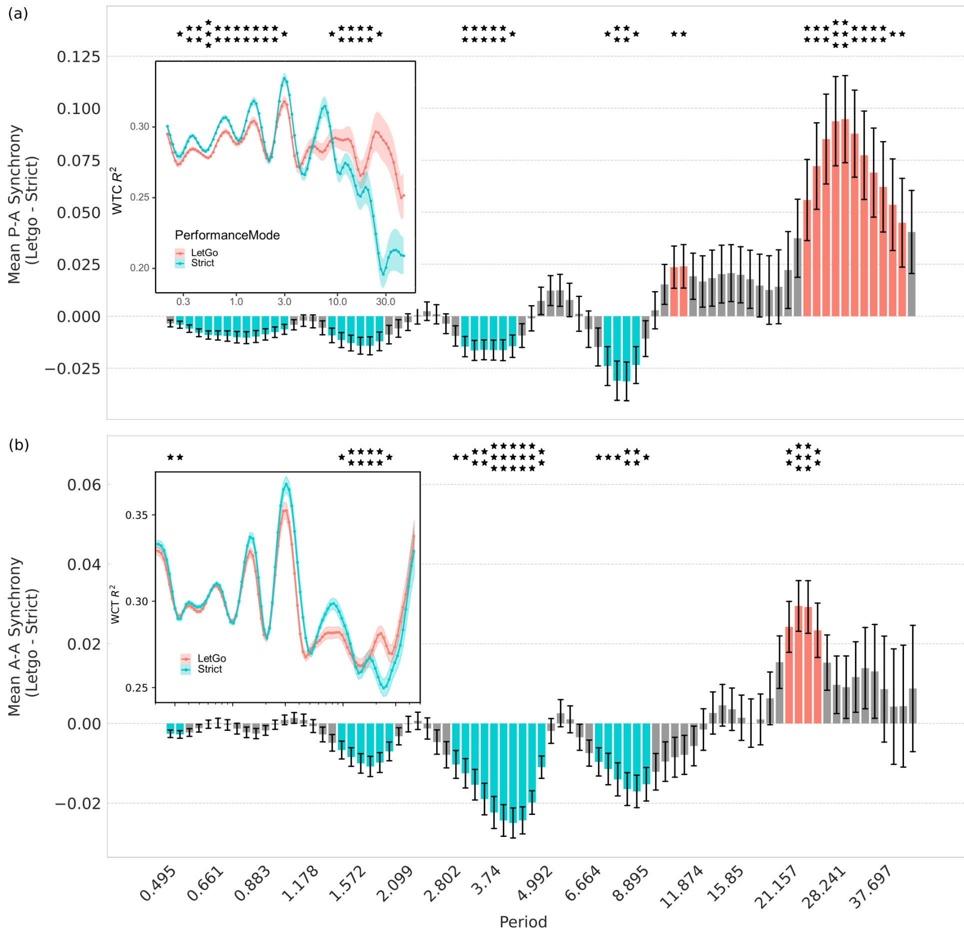

Figure 2. **(a)** Difference in P-A sync at different timescales (Fourier periods) between the modes of performance, *Let-go* and *Strict*, for both repertoire works. Colours reveal the two synchrony regions: scales with negative differences between *Let-go* and *Strict* in blue ('beat-sync'), and scales with positive differences between *Let-go* and *Strict* in red ('music-sync').
(a inset) Mean P-A sync at different timescales for the two modes for the repertoire works.
(b) Difference in mean A-A sync between the two modes for both repertoire works. Colours reveal the two synchrony regions.
(b inset) Mean A-A sync at different timescales for the two modes for the repertoire works.
Error bars in the main plots and shared areas in the insets indicate standard error of the mean (SEM) over 42 listeners. Periods with significant differences between modes are marked by asterisks. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; FDR-corrected.

which has been widely used to evaluate interpersonal physical synchrony in various types of interactions, including in a musical context[18,47–51].

When analysing synchrony at different timescales among audience members and between performers and audience, our results show that in both cases the audience exhibits higher synchrony in the *Strict* mode only at shorter timescales, while during the *Let-go* performances higher synchrony is seen at longer timescales (see Fig. 2).

Short timescales correspond to rhythmic elements of the piece as well as physiological signals such as breathing, and henceforth the synchrony that dominates in *Strict* can be referred to as **'beat-sync'**. In contrast, the longer timescales (more than 10 seconds), that dominate in *Let-go*, correspond to longer musical gestures related to higher-level semantics and musical expression[52], which we therefore describe as **'music-sync'**.

In addition to the average P-A and A-A sync, we studied temporal variability of the P-A and A-A sync at each timescale. Results show that the audience exhibits significantly more variability of synchronisation at longer timescales during the *Let-go* performances (see Fig. 3). We refer to the temporal variability of the synchrony in these longer timescales as **'music-sync variability'**. No significant differences were observed at shorter timescales.

Additional analyses showed no effects of blindfolding on the different types of synchrony and no significant interaction between visibility and performance mode. (See Section C.3 in Supplementary material for details).

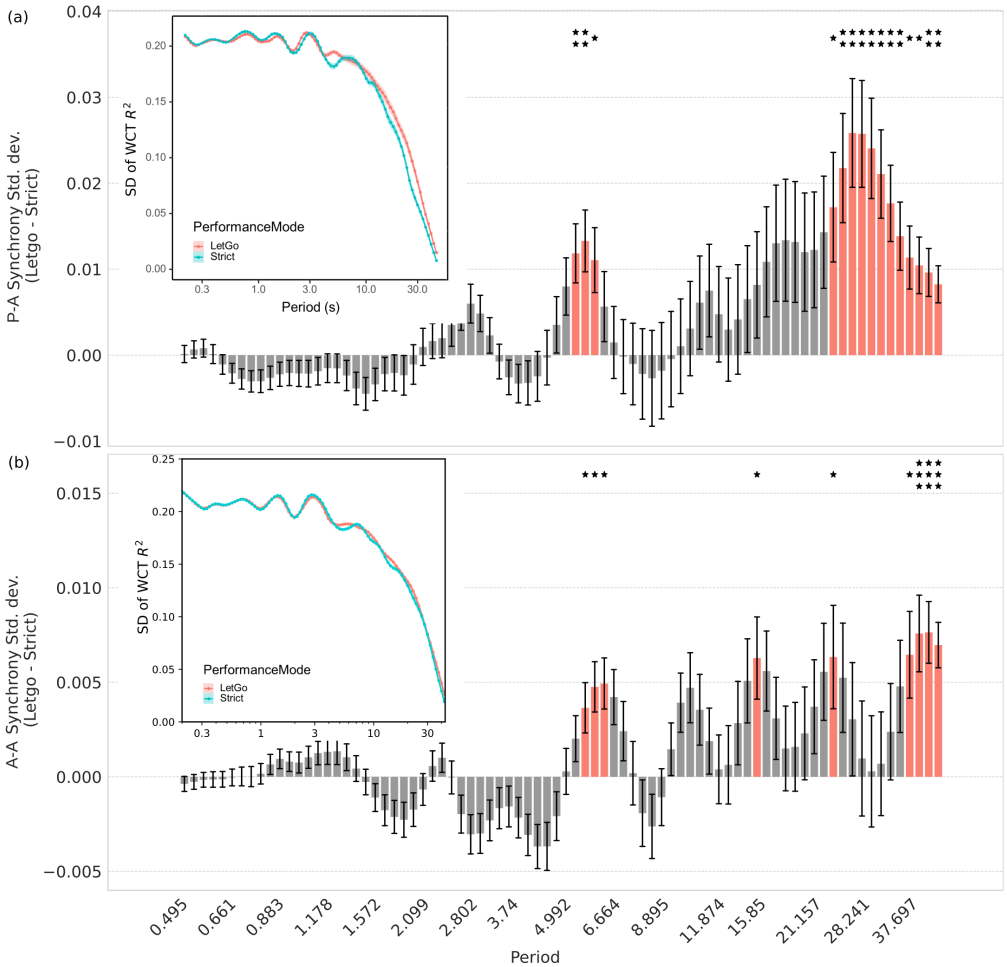

Figure 3. **(a)** Difference in the temporal variability (standard deviation) of P-A sync at different timescales between the two modes of performance, *Let-go* and *Strict*, for the two repertoire works. Red colour reveals the scales with significantly higher variability in *Let-go* than *Strict* mode ('music-sync variability'). **(a inset)** Temporal variability of P-A sync at different timescales for the *Let-go* and *Strict* performances of the two repertoire works. **(b)** Difference in the temporal variability of A-A sync between the two modes of performance for both repertoire works. Red colour reveals the scales with significantly higher variability in *Let-go* than *Strict* mode (music-sync variability). **(b inset)** Temporal variability of A-A sync at different timescales for the *Let-go* and *Strict* performances of the two repertoire works. Periods with significant differences between the two performance modes are marked by asterisks. *: p <0.05; **: p <0.01; ***: p <0.001; FDR-corrected.

The primary driver of the beat-sync, especially at the period range of 3-5 s, is assumed to be the audiences' respiration[53]. In order to deepen our understanding on the nature of this breathing component, we conducted further analysis on breathing patterns extracted from the accelerometer signals of each audience member.

We first investigated the diversity of breathing patterns exhibited by each individual by calculating their entropy rate (ER), which is a well-established information-theoretic metric of pattern diversity[54]. Results reveal an increase in entropy rate (ER) of breathing during the *Let-go* performance (see Fig. 4), suggesting increased variability of breathing patterns, which has been related to increased arousal with positive valence[55].

By studying the level of synchrony between the breathing patterns of pairs of audience members via the phase locking value (PLV)[56], we observe a significantly higher degree of synchrony in *Strict* than *Let-go*. We also investigated synchronisation patterns of higher-order — at the level of triplets — among the audience members, but did not find significant differences beyond the effect observed for pairwise synchrony (see section C.2 in the supplementary material). The larger low-order synchrony observed in *Strict*, which is contrary to our hypothesis 2, can be interpreted as arising from the more regular rhythms that characterise this performance mode, and can be related to the higher synchrony in beat-sync and the higher regularity of tempo in the *Strict* mode.

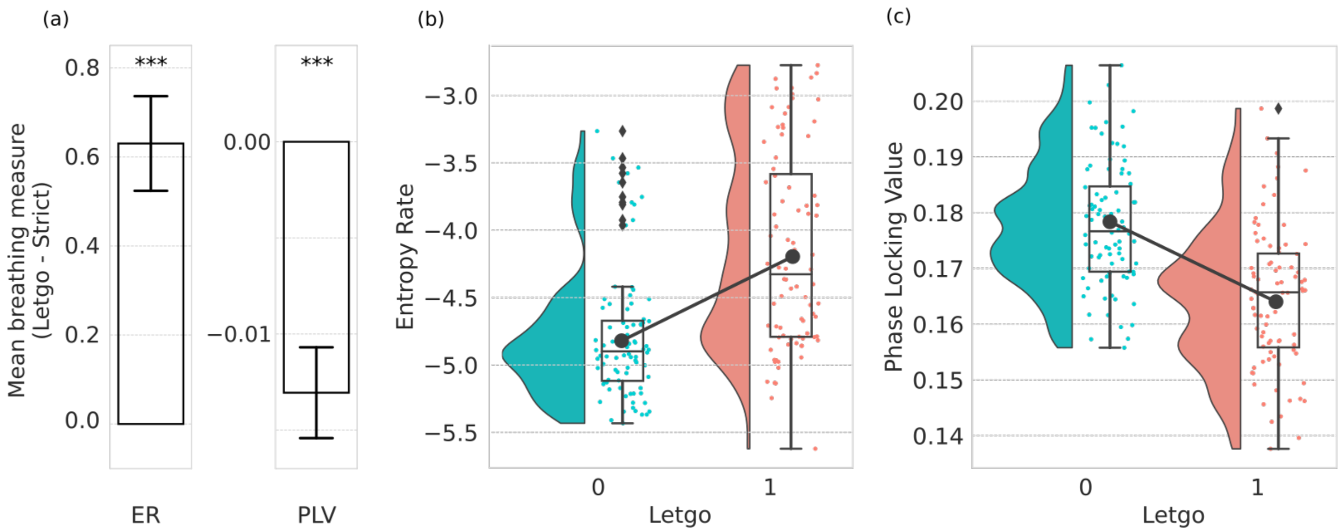

Figure 4. Comparison of breathing rate between the two modes of performance combined over the 2 repertoire pieces. (a) Entropy rate (ER) of individuals' breath is higher on average in *Let-go*, while PLV is lower. (b) Distribution of individual ER of breathing for the two modes. ER is higher in *Let-go*, showing increased variability. (c) Distribution of pairwise synchrony computed as phase locking value (PLV) between all pairs of subjects. On average PLV is higher in strict, corresponding to 'beat-sync'.

2.3 Correlation between ratings and synchrony

As a final step in our investigation, we studied whether the differences in synchrony found in the previous section were predictive of the subjective experience as reported in the questionnaire ratings. For this, we built multilevel models using the various questionnaire items as dependent variables, and mean synchrony (either beat- or music-sync) or music-sync variability as independent variables, while accounting for subject ID using a random intercept (see Section 4 for details).

Results show that higher music-sync variability is most significantly associated with higher subjective scores (see Table 1), in particular the PC1 and items of the audience ratings that constitute perceived innovativeness of the performance. In contrast, increases in the average 'beat-sync' were negatively associated with PC1 and the Improvisatory and Risk-taking ratings. This indicates that higher synchrony in this timescale was linked to lower levels of improvisatory perception. The results are only significant in P-A sync. Changes in average music-sync between audiences or musicians were not associated with differences in the subjective scores.

Overall, this suggests that having more dynamic synchrony at the scale of the musical discourse is associated with the distinctive experience provided by the *Let-go* performance, while having higher synchrony with musicians in the 'beat-sync' scale is more characteristic of the less risk-taking experience.

It is worth noting that when focusing solely on Haydn's composition (which musicians regarded as more successfully differentiated between *Let-go* and *Strict* modes), the associations of mean 'beat-sync', mean 'music-sync', and 'music-sync variability' with the ratings are stronger. More information can be found in Table 4 in the supplementary material.

3 Discussion

This study reveals different aspects of the impact of improvisatory performance on the audience's collective experience. From a psychological point of view, subjective ratings exhibit significant differences in how the audience experiences performances with and without improvisatory elements. From a physiological perspective, the multi-layered structure of the patterns of collective physical movement show consistent differences between performance modes. Furthermore, results suggest that physical movement is an effective window to look into the internal subjective experience of the audience, as differences in synchronisation patterns are consistently associated with differences in subjective ratings. Overall, these findings suggest that the deviation from expectation at different levels of performance parameters is reflected in intricate interactions involving various physical and subjective dimensions in an emotional dialogue between the performers and audiences.

Consistently with previous studies[32], audiences rated *Let-go* performance higher than *Strict* counterparts in various experiential dimensions, suggesting that the experiment was successful in inducing differentiated musical experiences on the audience. In particular, the audience perceived the higher Improvisatory, Innovative and Risk-Taking character of *Let-*

Table 1. Correlation between audience ratings and mean physical synchrony in shorter and longer timescales, and the temporal variability of synchrony in the longer timescale.

Rating	Beat-sync			Music-sync			Music-sync variability		
	t_{125}	p		t_{125}	p		t_{125}	p	
P-A sync									
PC1	−2.04	**0.043**	*	1.13	0.260		2.82	**0.006**	**
Improvisatory	−2.03	**0.044**	*	1.14	0.256		2.36	**0.020**	*
Innovative	−1.42	0.158		0.75	0.453		2.86	**0.005**	**
RiskTaking	−3.21	**0.002**	**	1.52	0.130		2.93	**0.004**	**
Engaging	−1.55	0.124		0.94	0.347		0.97	0.336	
Convincing	0.32	0.747		−0.23	0.821		1.81	0.072	
Familiar	0.95	0.344		−1.17	0.245		−1.66	0.099	
Sleepy	1.40	0.164		−1.05	0.295		−1.71	0.090	
A-A sync									
PC1	−0.85	0.396		0.64	0.525		3.43	**0.001**	***
Improvisatory	−1.58	0.116		0.28	0.782		3.98	**0.000**	***
Innovative	−1.02	0.307		1.28	0.202		3.35	**0.001**	**
RiskTaking	−1.82	0.071		1.31	0.191		3.41	**0.001**	***
Engaging	−0.04	0.965		−0.05	0.958		0.55	0.586	
Convincing	1.36	0.175		0.12	0.906		2.63	**0.010**	**
Familiar	0.47	0.637		0.79	0.432		−0.24	0.809	
Sleepy	0.88	0.383		2.37	**0.019**	*	−1.85	0.067	

go performances, while considering both performances as Musically Convincing. Additional analyses show no effects of blindfolding on ratings, suggesting that the music itself — rather than visual cues — acted as a driver for the collective subjective experience. Moreover, results also show that performance ratings are also related to the psychological trait of absorption, but this does not explain away the effect of the performance mode. Absorption has been previously linked to the enjoyment of music[57], yet it does not seem to affect collective engagement in the *Let-go* performance.

Our results reveal that improvisatory elements affect movement synchrony of audiences in opposite directions, depending on the timescale. In effect, *Let-go* performances reduce synchrony comparing with *Strict* in shorter timescales (below 10 seconds), while they enhance synchrony on longer timescales (above 10 seconds). Short timescales can be associated with the rhythmic pulse and physiological responses to it, which are more clear in the Strict rendition of the music, and longer timescales with longer structures and musical gestures[52].

Our findings, therefore, suggest that collective music experience is embodied in a multiscale adaptive interaction between the performers and audiences, with these spanning a longer temporal horizon in improvised renditions than in strict ones. Similar time-scale dependency of the physical synchrony has also been observed in different forms of social interactions, including collaborative team problem solving[50] and joke telling[47].

It is worth noticing that the fact that synchrony was observed both for blindfolded and sighted audiences suggest that, in terms of mechanisms, audience modulated their physical synchrony with the performers mainly via auditory rather than visual information, which is in line with previous results[32]. This suggests, in turn, that performance-to-audience synchronisation was primary, and that audience-to-audience synchronisation emerged mainly indirectly, mediated by the former interactions — rather than by the direct interaction between audience members.

Our analysis of synchrony in movement patterns was not restricted to the average degree of synchrony, but also considered the variance of synchrony during the performance. Results show that improvisatory elements increase the temporal variability of synchronisation on longer timescales. Combined with the results of the average sync, this means that the improvisatory performance increased longer-timescale synchrony at specific timings rather than evenly throughout the performances. In other words, it enhanced the shift between convergent (in-sync) and divergent (out-of-sync) phases. This could be interpreted as promoting meta-stable dynamics, which could in turn be a marker of adaptive states[44,58]. In contrast, decreases of synchrony in the shorter-timescale and increase of diversity in breathing pattern by the improvisatory performance took place more evenly over the whole performances, as shown by the less significant changes in temporal variability. This confirms the idea of different origin of the shorter- and longer-scale sync, and further suggests that the shorter-scale sync corresponds to the low-level musical components (shorter beats and metres) and autonomic responses to them, which exist throughout the performances, while the longer-scale sync corresponds to the temporally organized higher-level hierarchical musical structures.

Interpretation in terms of music performance is consistent with our data-driven results. Short timescales are associated with the pulsation of rhythm, which is more pronounced, at times rigidly so, in the *Strict* musical performances. The rhythm of the music is known to act as a driver of physiological rhythms such as breathing[37,41], thus enhancing 'beat-sync'. In contrast, the longer timescales are associated with freer musical gestures, based on deeper, structural pulses in the music that allow more possibilities in terms of phrasing, articulating and ability to deviate from expectations in *Let-go*[52,59]. We can further associate the higher 'music-sync' in *Let-go*, as well as the higher temporal 'music-sync variability', with the audience's synchronised response to the spontaneous and unplanned arrival of the ensemble at the same point in the music, crafting moments of peak emotional expression. Previous work has also shown that the audience shows higher physiological synchrony during important structural moments in the music[37].

Here we must revisit the distinction between the structural design of a composition, and the micro- and macrostructural patterns emerging in performance[25,29]. We argue that performers who apply an improvisational state of mind[32] use the same kind of generative processes inherent in composition[60] in the spontaneous creative processes of performance, whether they are performing a repertoire work or freely improvising[59]. Further explorations of music performance parameters such as tempo and dynamics in important structural moments in both text and performance, and how they are linked to the subjective experience of musicians and audience, are an important avenue for future work.

The statistical associations found between changes in psychological ratings and patterns of collective movement suggest that these may be reflecting different angles of the same underlying phenomenon. Interestingly, results show that higher synchrony in the shorter timescale was negatively associated with the audience's perception of the innovativeness of performances, which further supports the idea that the shorter-scale synchrony may reflect rather automatic and unconscious alignment to low-level structural/syntactic aspects of the music. That is, the more standard and predictable a performance was (especially in the *Strict* mode), the easier it may have been for the audiences to physically and automatically get entrained into it. At the same time, the high predictability may have led to below the optimal zone of uncertainty for music pleasure[61–63], giving the audience the impression the performance was less innovative. On the contrary, higher synchrony and its temporal variability in the longer timescale was positively associated with the audience's innovative experience. Thus, the longer-scale synchrony may reflect the audience's absorption to the dynamics of higher-level musical expression or semantics, which is enriched by the *Let-go* performance mode.

In conclusion, this research uncovers the relevance of the often-neglected multiscale coordination between audiences and performers, and reveals its deep connections with the quality of the collective subjective experience. Our results provide quantitative evidence that illuminates how a collective music experience is embodied in a multiscale dynamical interaction which expands the group flow aspects of the relationships between the improvising musicians[19,20] to a complex dialogue with audiences that is enhanced by the innovative, risk-taking and unexpected qualities of improvisatory performance.

The evaluation of collective creative activities that are particularly difficult to verbalize and share usually requires experts' intuitions. The current results provide a first step towards the quantification of some aspects of these ephemeral experiences, opening the possibility for sensing technologies to evaluate these elusive yet important aspects of collective experience — and even potentially enriching them via personalised real-time feedback. Last but not least, the reported results highlight the importance of regarding collective creative activities as physically embodied experiences, suggesting a rich tapestry of physical behaviour underlying the shared experience even in audiences that could be seen as passive.

4 Methods

4.1 Experimental procedure

The concert/experiment involved the Portorius String Quartet, who performed movements from Mozart (String Quartet No. 15 in D Minor K. 421 – first movement: Allegro moderato) and Haydn (String Quartet in G Major, Hob.III:75, Op. 76, No. 1 – third movement: Menuetto: Presto) as well as improvised pieces in different performance modes (Table 2). Specifically, for the repertoire works, the same piece was performed twice, in each of the two modes, *Strict* and *Let-go*, varying the order, allowing us to better isolate the effect of performance mode on the audience. The two repertoire pieces were chosen as they are both from the classical period and their phrase structure lends itself to more straightforward creative work when performed in *Let-go*, but they contrast each other in mood and musical energy. Mozart's piece is more introverted and in complex from a contrapuntal point of view, Haydn's is more extroverted and varied from a rhythmic point view.

Prior to the concert, all members of the quartet took part in Profesor David Dolan's course *Interpretation through Improvisation* at the Guildhall School of Music and Drama in London[64]. The method applied involves a creative approach to studying and performing repertoire works, engaging with structural, harmonic, rhythmic and motivic reductions with improvisational state of mind.

The concert experiment was conducted in a recital room in the Guildhall School of Music and Drama (see Fig. 5).

Audiences were recruited via posters on bulletin boards and online call for participation. Fifty adult volunteers attended the concert experiment as audience. They were mainly graduate students and staff of the Imperial College London or their families and friends, with a wide range of experience with classical music. Out of them, 8 subjects encountered issues with the physical motion recording or failed in giving the subjective ratings on the performances. Therefore, the data from the remaining 42 subjects were subjected to the analyses. In order to investigate the role of audience's vision, 13 out of the 42 audience members listened to the performances wearing blindfolds.

4.2 Measurements

4.2.1 Body motion acceleration

The performers' head motions were measured with inertial measurement units (IMUs; TSND151; ATR-Promotions, Japan) placed on the middle of their forehead, attached to the fNIRS brain activity measurement device (HOT-1000; NeU, Japan). The audience members' body motion fluctuations were measured with IMUs contained in the smartphones (Zenfone 3 Laser; ASUSTek, Taiwan) that they wore around their necks[65]. The sampling frequency was 100Hz for both sensors, and then downsampled to 50 Hz.

4.2.2 Questionnaires

Before the study, audience members filled a psychometric questionnaire to assess their psychological trait of absorption[66] as this has been previously related to the enjoyment of music[67], as well as susceptibility to altered states of mind and even psychedelic experiences[45].

After each pair of successive performances, the audiences rated their subjective evaluation of each performance on seven items: how they felt each performance to be (1) Improvisatory, (2) Innovative, (3) Emotionally Engaging, (4) Musically Convincing, and (5) Risk-taking. These items were identical to the ones used in the previous studies[30,32]. Two additional items were added, where the audiences were asked to rate their degree of (6) familiarity with the piece and (7) sleepiness. The rating for each item was given on a six-level Likert scale (0—5).

The collected rating data contained small amount of missing values; in the 168 samples consisting of 42 audiences and 4 pieces, ”Improvisatory”, ”Convincing”, ”Familiar”, and ”Sleepy” items had one missing value each, ”Risk-taking” item had two missing values (no observation had more than one missing values). These missing values were imputed using the missForest algorithm, a random forest-based multiple imputation scheme[68].

4.3 Analysis

4.3.1 Wavelet synchrony analysis

To evaluate synchrony between physical activity, triaxial head acceleration data of the musicians (from IMU sensors) and body acceleration data of the audience (from smartphones) was converted to a one-dimensional time series of acceleration Euclidean norm.

$$a(t) = \sqrt{a_x^2(t) + a_y^2(t) + a_z^2(t)}$$

Then, we evaluated physical synchrony of each pair of signals by using the wavelet transform coherence (WTC)[46] of their acceleration norm time series. WTC finds regions in time-frequency space where two time series covary, but do not necessarily have high power. WTC has been used to evaluate interpersonal physical synchrony in various types of interactions[18,48,50] and is defined as[69]:

$$R^2(t,s) = \frac{|S(s^{-1}W^X(t,s)W^Y(t,s))|^2}{S(s^{-1}|W^X(t,s)|^2)S(s^{-1}|W^Y(t,s)|^2)}$$

where W^X and W^Y refer to the wavelet transforms of the two signals and t and s refer to time sample and wavelet scale. Wavelet scale s is directly associated with a Fourier period[69], which is used to discuss scales of synchrony. Results were computed using the open-source `wavelet-coherence` Matlab package[70] and the mother wavelet and initial parameters are the same as in[46].

By averaging the R^2 coefficients for the performer-audience pairs over the duration of performance and over the four performers, we obtained a measure of how much each listener was in sync with the performers on average, at each timescale, for each performance. Furthermore, the subject-average coefficients are then averaged across subjects for each timescale, in order to infer overall synchrony in the audience.

To obtain an overall degree of variability in the synchrony, we also compute the variance of the wavelet coefficients across listeners in each timescale. Bessel's correction is used when computing the standard error of the means over the whole audience.

Similarly, to evaluate synchrony between audience members, the average measure of synchrony for a given subject S_i in a given timescale s_k was obtained by taking the mean of all pairwise values between S_i and all other audience members $S_j \neq S_i$ in the same timescale.

Due to the similar duration of the repertoire pieces (between 120 and 140 seconds), the same wavelet scales (or Fourier period) can be used to discuss all pieces. We choose a range of relevant periods to be <0.5 s, as the timescales below it have no musical meaning.

The synchrony analysis is conducted in order to identify ranges of frequencies (or bands) where there are significant differences in the audience's degree of synchrony between the performance modes. Averaging the per-subject wavelet coefficients in these bands provides a measure of synchrony in that band, which can be used further to test interactions between different factors affecting different bands. This method can further allow us to incorporate the post-hoc difference in the expected performance modes of the Mozart piece.

4.3.2 Breathing rate analysis

To further investigate A-A sync, the breathing rate of participants was extracted from the front (z-axis) of the triaxial acceleration data by using a continuous wavelet transform[71]. The wavelet coefficients in the relevant scales for breathing (3-5s) were then used to reconstruct the respiration signals[53], producing a time series that can be analysed with stationary methods, due to the oscillatory nature of breathing.

To investigate synchrony of breathing, average pairwise phase locking value (PLV)[56] was computed and averaged for each subject. PLV is a measure of phase synchrony between a pair oscillatory signals calculated using their average phase difference. To obtain mean synchrony for a subject, their mean PLV with all other audience members is computed.

To investigate variability in breathing, entropy rate was computed on each listener's reconstructed breathing signals using the state space estimator[54]. This measure uses vector auto-regressive model to estimate entropy rate of continuous signals and is shown to be data-efficient and calibrated against other measures like Lempel-Ziv complexity (LZc).

4.4 Statistical tests

To study the differences in ratings, as well as in the average synchrony and temporal variability of synchrony at each period (timescale) between performances, we estimated a three-way mixed-effect multilevel model that includes the performance mode and composition as within-subject factors and blindfolding as a between-group factor with fixed effects, and each subject as random effects. We primarily focused on the main effect of the performance mode, as the composition was a factor of little interest. Using the `lme4` package[72] in R statistical software, the multilevel model is expressed as

```
DV ~ Blindfold * Composition * Mode + (1|Subject) +
           (1|Composition:Subject) + (1|Mode:Subject)
```

where DV represents the dependent (target) variable. All the fixed-effect independent variables are zero-centered before estimation[73]. Statistical significance of the variables were tested using the `lmerTest` package[74].

In the analyses of physical synchrony, to correct for multiple testing over many timescales, false discovery rate (FDR) control via the Benjamini-Hochberg procedure[75] was applied to the p-values.

For each subject and performance, the mean physical synchrony values in the timescales with significant *Let-go* $<$ *Strict* difference were averaged into the 'beat-sync' measure, those with significant *Let-go* $>$ *Strict* difference were averaged into the 'music-sync' measure. Similarly, synchrony variability values in the timescales with significant *Let-go* $>$ *Strict* were averaged into the 'music-sync variability' measure.

To investigate how these measures were predictive of the subjective evaluations by the audience, multilevel models of the form `rating ~ sync + (1|subject)` were tested. Here, `sync` represents either the beat-sync, music-sync, or music-sync variability after centering-within-cluster. This analysis is equivalent to the within-subject repeated measures correlations[76], and evaluates how the within-subject variances in the sync and rating are consistently correlated over the four performances (*Let-go* and *Strict* performance mode for the both pieces) or over the two performances with the different modes for each piece, separately.

For studying the relationship between absorption and other variables, since absorption is a between-subject factor, we average the other variables across the performances and use simple linear models of the form `mean_rating ~ mean_sync * absorption`, adding further interactions with `Blindfolded` or `Composition` where relevant.

Acknowledgments

We thank the Guildhall School of Music and Drama for allowing us to use their facilities and for the help of their IT team support. This work was partially supported by the Center of Innovation Program (Grant Number JPMJCE1309) from JST, Japan, and also by KAKENHI (Grant Numbers JP17H01753, JP20H03553, and JP21K19787) from JSPS/MEXT, Japan. M.S. acknowledges a scholarship by Splunk.

Declarations

Ethical approval

This study was approved by the Human Subjects Research Ethics Review Committee, Tokyo Institute of Technology (Approval No. 2019101), and was conducted according to the Declaration of Helsinki. Written informed consent was obtained from all participants.

A Experimental setup

The programme of performances is shown in Table 2, and the seating layout for the audiences is shown in Fig. 5.

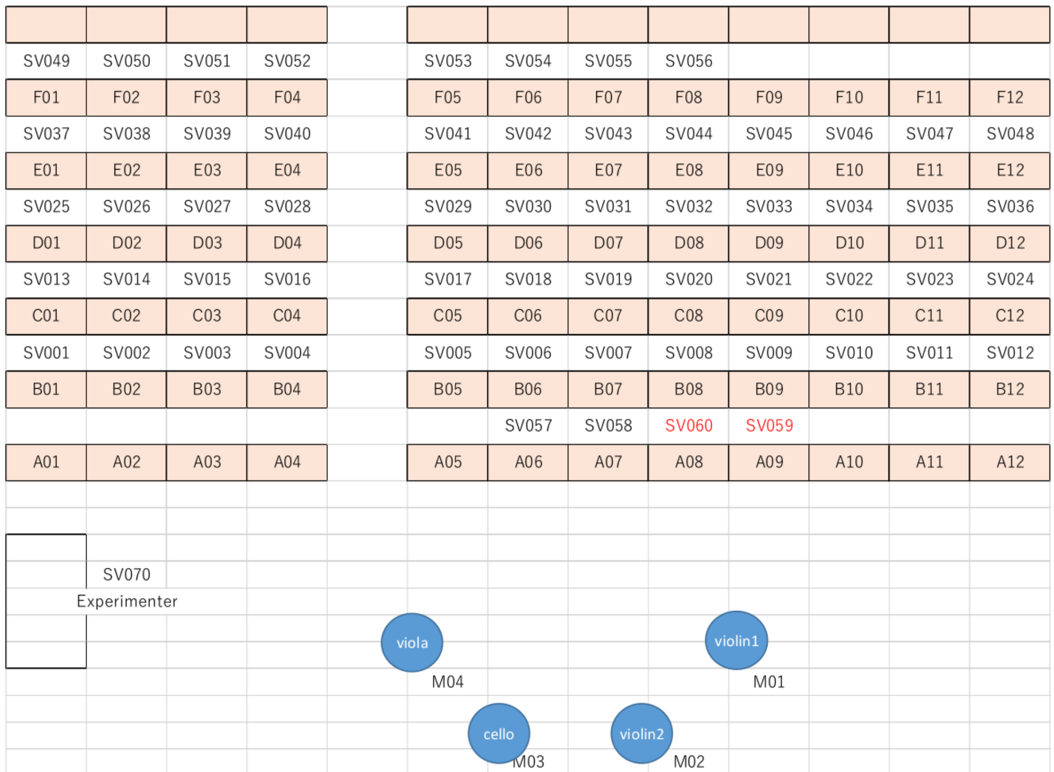

Figure 5. Recital room layout for the concert experiment. Audiences in seats labelled 01 to 04 wore blindfolds during the repertoire pieces.

Table 2. Performance programme

Piece	Type	Composer/Leadership	Performance mode	Blindfolding
1	repertoire	Mozart	*Let-go*	1
2	repertoire	Mozart	*Strict*	1
3	improvisation	single lead	*Strict*	0
4	improvisation	dynamic lead	*Strict*	0
5	improvisation	dynamic lead	*Let-go*	0
6	improvisation	single lead	*Let-go*	0
7	repertoire	Haydn	*Strict*	1
8	repertoire	Haydn	*Let-go*	1

B Psychology

B.1 Combined effects of mode, composition, and blindfolding on performance ratings

Results for the effects of composition and visibility are shown in Fig.6 and Table 3. They revealed significant interaction between the performance mode and composition factors, and main effects of performance mode and composition for the Improvisatory, Innovative, and Risk-taking ratings, and a marginal main effect of the performance mode for the Emotionally Engaging rating.

Visibility (sighted vs. blindfolded) had no significant main effects nor interaction with performance modes on the ratings. Although not significant, blindfolded audiences tended to show less sensitivity to the performance modes than sighted ones. There were also significant 3-way interactions between visibility, performance mode and compositions for some rating items. Composition-wise, the blindfolded audience tended to feel more improvisatory and risk-taking toward the *Let-go* mode performance of the Haydn's composition, but they felt oppositely to the Mozart's composition. We surmise that visual perception can affect the music listening experience to some extent.

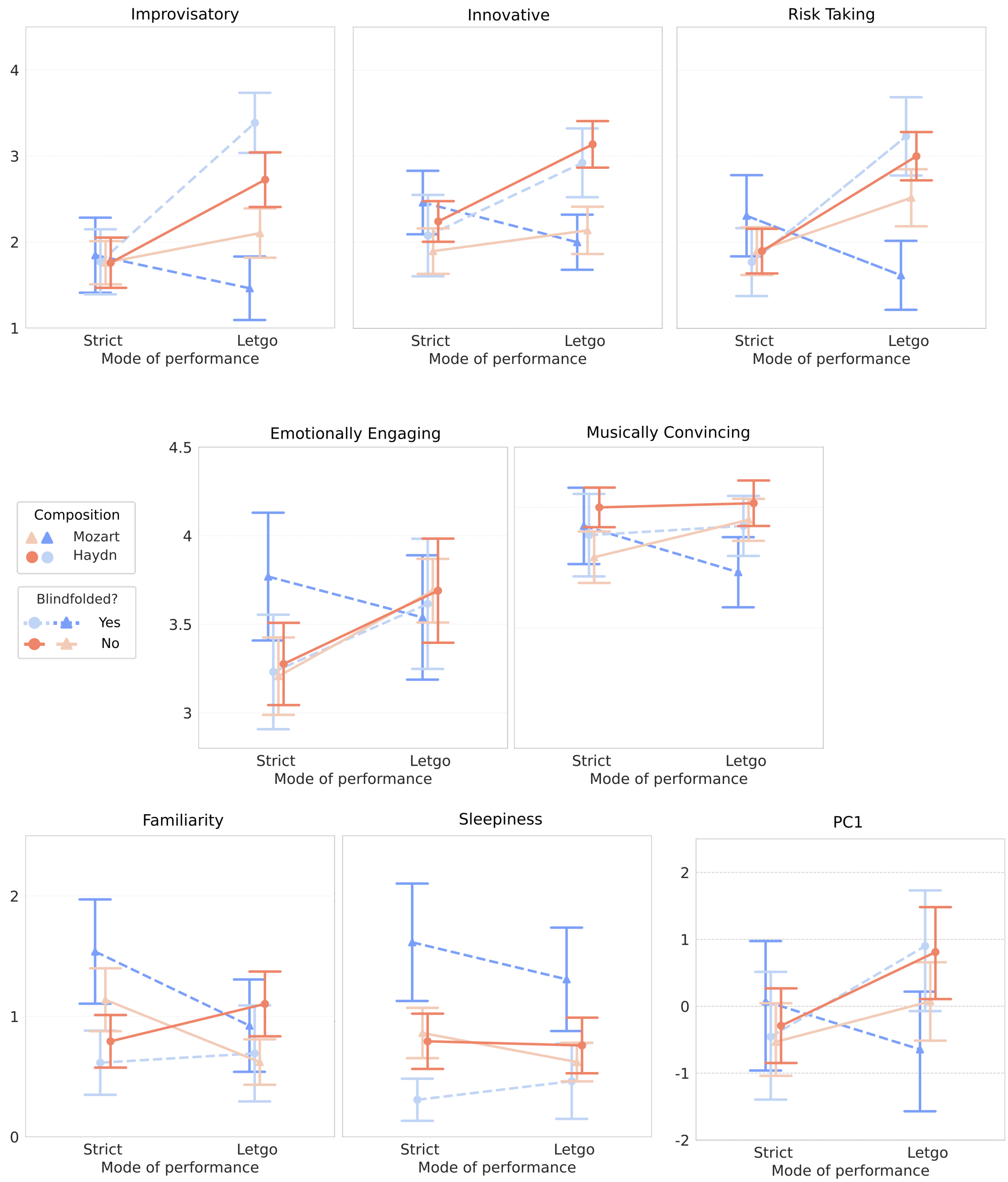

Figure 6. Comparing the audience's perception of modes of performance, separately in each of the two repertoire pieces (by Mozart and Haydn), and grouping by whether the audience was blindfolded. Error bars show standard error of mean (SEM). The audience perceives the Haydn *Let-go* performance as much more Improvisatory, Innovative, and Risk-taking than the *Strict*, while there is little difference in these two metrics in the Mozart piece, echoing the performers' reports on the success of the performance.

Table 3. Combined statistical effects on the audience ratings identified using the multilevel model.

Effects of	statsistics	Improvisatory	Innovative	RiskTaking	Engaging	Convincing	Familiar	Sleepy	PC1
Main effects									
Sight	F(1,40)	0.01	0.00	0.06	0.06	0.42	0.01	0.32	0.01
	p-value	0.940	0.974	0.801	0.800	0.522	0.934	0.575	0.910
Composition	F(1,40)	**8.39**	**5.18**	3.12	0.18	1.70	1.72	**5.88**	3.83
	p-value	**0.006**	**0.028**	0.085	0.670	0.200	0.197	**0.020**	0.057
Mode	F(1,40)	**11.92**	**4.37**	**9.14**	3.29	0.00	2.77	0.79	**6.05**
	p-value	**0.001**	**0.043**	**0.004**	0.077	0.952	0.104	0.378	**0.018**
2-way interactions									
Sight × Composition	F(1,40)	2.07	0.95	0.45	0.34	0.14	2.78	**6.68**	0.00
	p-value	0.158	0.336	0.505	0.566	0.713	0.103	**0.014**	0.967
Sight × Mode	F(1,40)	0.01	1.07	1.34	1.65	1.15	0.55	0.06	1.26
	p-value	0.914	0.307	0.254	0.207	0.291	0.463	0.802	0.269
Composition × Mode	F(1,40)	**18.02**	**14.51**	**13.07**	0.82	0.15	**6.98**	2.06	**14.90**
	p-value	**0.000**	**0.000**	**0.001**	0.371	0.705	**0.012**	0.159	**0.000**
3-way interaction									
Sight × Composition × Mode	F(1,40)	**4.99**	1.60	**5.25**	1.28	2.30	0.06	0.30	**5.63**
	p-value	**0.031**	0.213	**0.027**	0.264	0.138	0.815	0.588	**0.023**

B.2 Effect of absorption on ratings and sync

Fig. 7 shows the distribution of absorption metrics in the audience and the relationships between absorption and PC1 of ratings. We observe a strong correlation between high absorption and high ratings regardless of other factors, further enforcing the idea that high absorption is linked to more positive musical experience in general[45].

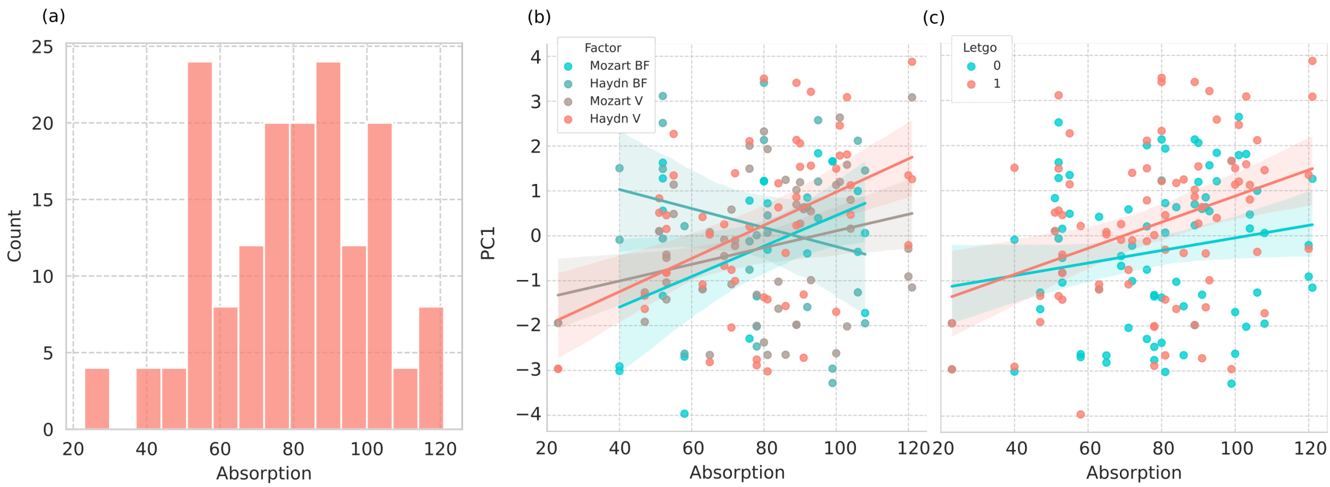

Figure 7. (a) Histogram of absorption metrics in the audience. (b) Relationship between absorption and PC1 of ratings, grouped by mode of performance. A consistent difference in slope is seen between the two groups across the two compositions. (c) Relationship between absorption and PC1 of ratings, grouped by blindfolding and composition. A consistent difference in slope is seen between the two groups across the two compositions.

Fig. 8 (a)-(g) shows scatter plots of the relationship between absorption and each audience rating. Results were obtained by performing linear regression with Python function `scipy.stats.linregress` between the absorption psychometric for each subject and their responses for each rating.

We also apply a non-linear measure of correlation: Székely's distance correlation[77], computed with the Python package `dcor`. A null distance correlation implies independence, a distance correlation of 1 implies full dependence between the variables.

The distance correlation with the absorption metric is highest for Emotionally Engaging (0.3), Innovative (0.25), and Improvisatory (0.22) and lowest for Sleepiness (0.14). Therefore, the most significant linear relationships seem to also show highest distance correlations, yet in itself the correlation is weak.

When comparing absorption with average synchrony for each subject directly, no strong linear correlations emerge. The Pearson ρ correlation coefficient, (computed with `scipy.stats.pearsonr`) is positive for Mozart but negative for Haydn, but the results are non-significant. Linear regression only yields non-significant results ($p>0.5$), and distance correlation between absorption and mean synchrony is between 0.21 and 0.27 for all pieces, which at best suggests only a weak correlation.

The interaction of the absorption, blindfolding and composition factors reveals an interesting difference between the two compositions, further reinforcing the significant difference between the way the mode of performance was executed. In the first piece by Mozart, both slopes are positive, suggesting blindfolding does not make much difference in how the subjects rate the piece. But in the second piece by Haydn, blindfolded subjects tend to rate the piece higher the higher their absorption, while sighted subjects show the expected positive correlation between absorption and ratings.

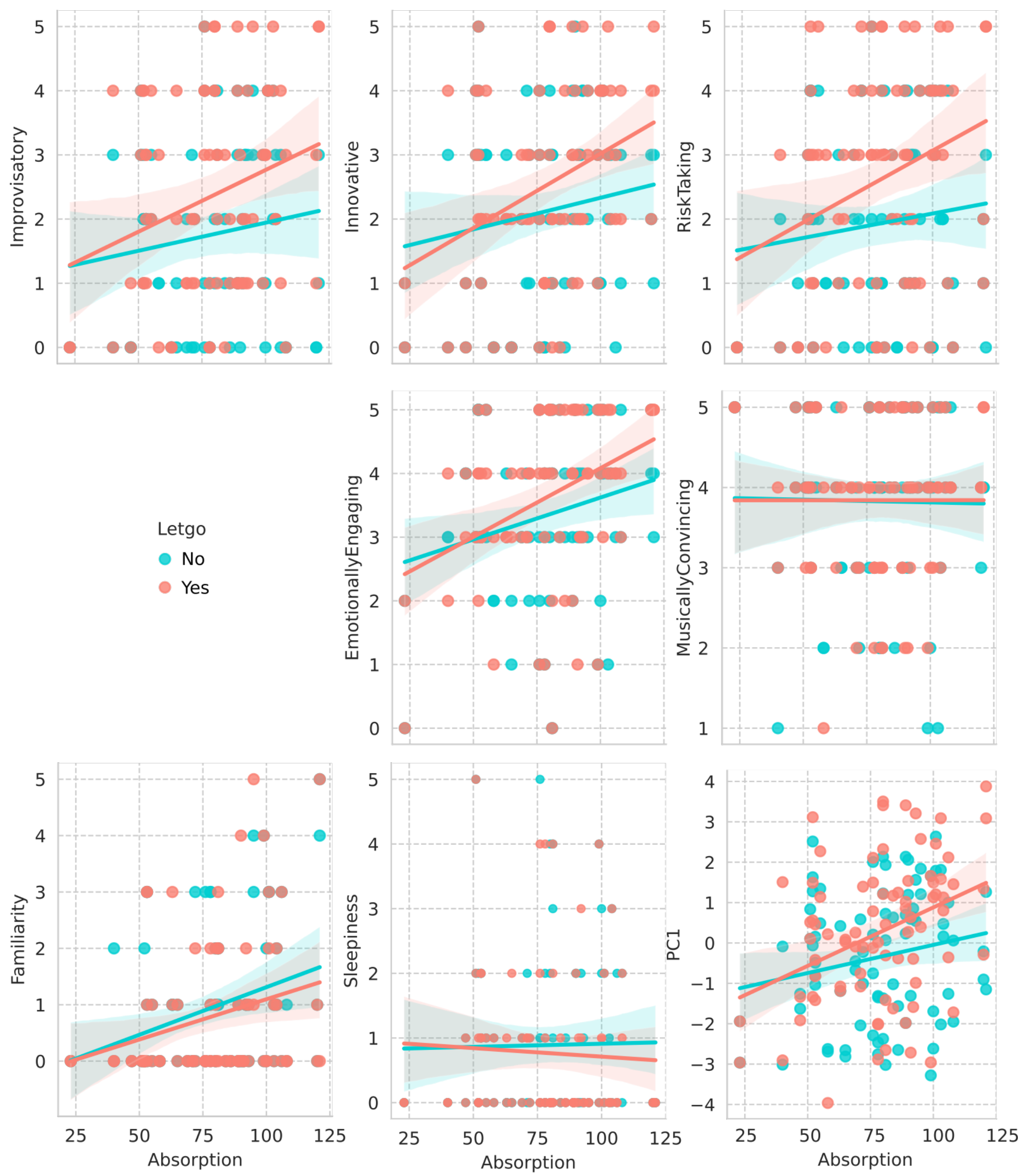

Figure 8. Relationship between absorption and each audience rating, grouped by mode of performance. (a) Improvisatory (b) Innovative (c) Musically Convincing (d) Emotionally Engaging (e) Risk Taking (f) Familiar. The lines indicate simple regression slopes for each mode of performance. A non-significant yet consistent difference in slope is seen between modes of performance in the Improvisatory, Innovative, Risk-Taking and Emotionally Engaging ratings. There is no relationship between absorption and the Musically Convincing rating. Mode of performance does not have an effect on the relationship between Familiarity or Sleepiness and Absorption.

C Physiology

C.1 Effect of composition

The performers reported that they failed to achieve the ideal *Let-go* performance in the first performance of Mozart (Piece 1). Audiences' perception was in accordance with this judgment by the performers. Therefore, to incorporate this post-hoc difference in the performance quality of the *Let-go* mode, we conducted the synchrony analyses separately for the two performances of each repertoire piece.

C.1.1 Audience's physical activity

To evaluate power of physical activity in different periods, a wavelet power spectrum (WPS), given by $\|W^X(t,s)\|^2$, was applied to the acceleration norms computed from the triaxial physical motion data. Log-scaled WPS was averaged over the time duration and subjected to group-level ANOVAs at each period. Bias in the wavelet power spectrum (WPS) was rectified using the method of[78].

Fig. 9 shows mean power spectra of the audience's physical activity during the two performances of the two repertoire pieces.

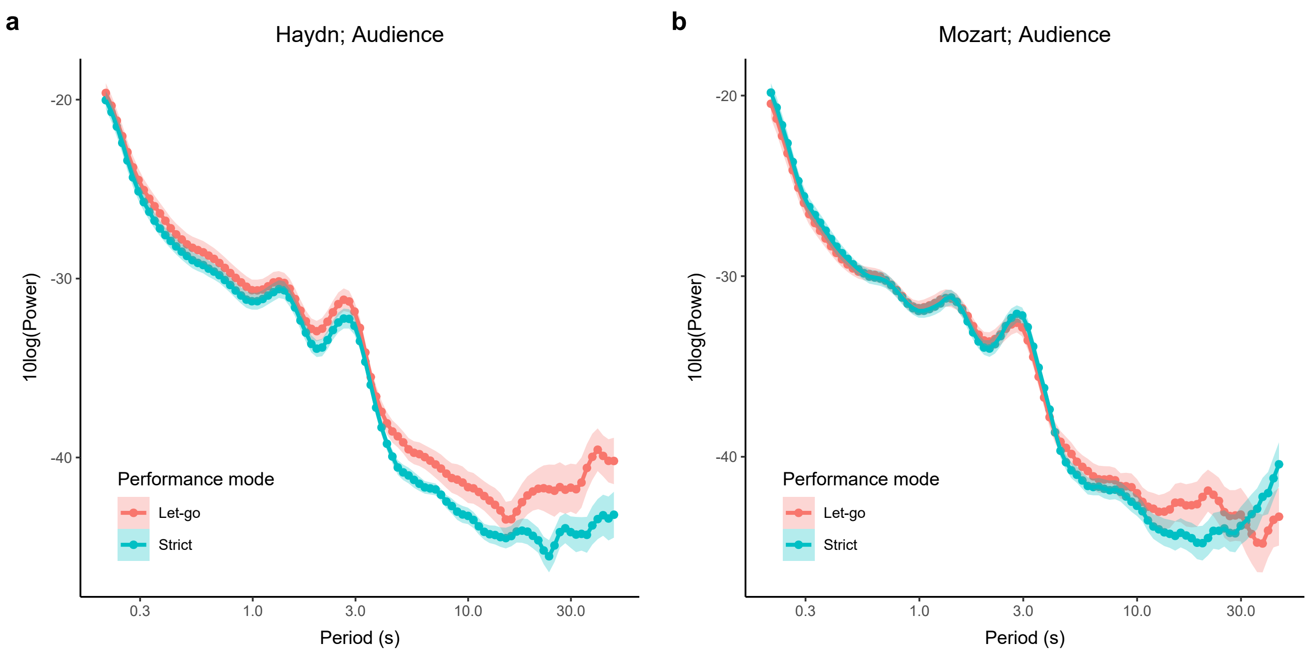

Figure 9. Mean power spectra of the audience's physical activities during the two performances of (a) Haydn's piece and (b) Mozart's piece, both in the *Let-go* and *Strict* performance modes. Spectra are calculated by time-averaging the log-transformed wavelet power in each performance. For the Haydn pieces, the effect is more pronounced, showing higher power during the let-go performance, yet the effects are not significant.

The power spectra indicate the existence of oscillatory components at the periods around 1.5s and 3s, possibly reflecting the audience's physiological markers (heartbeat and respiration) or their implicit bodily reaction to musical beats. We explore this relationship further in the sections that follow 2.2.Comparing the two performance modes, whilst no significant differences were found, we can observe trends by analysing the two pairs of performances separately. The audience showed a tendency towards larger amplitude movement during the *Let-go* performance compared to the *Strict* performance of Haydn's piece. The tendency was less clear between the *Let-go* and *Strict* performances of Mozart's piece.

C.1.2 Physical synchrony

When studying the synchrony regimes individually for each piece (Figs. 10 and 11), we observe the peaks and troughs in the time-averaged synchrony differ according to the piece being performed, thus explaining the interactions between performance mode and composition at certain periods.

C.1.3 Temporal variability of physical synchrony

Temporal variability of synchrony was higher in the *Let-go* mode in longer timescales for both compositions (Figs. 12 and 13). Note that the temporal variability is commonly lower in longer timescales because of the higher auto-correlation of synchrony

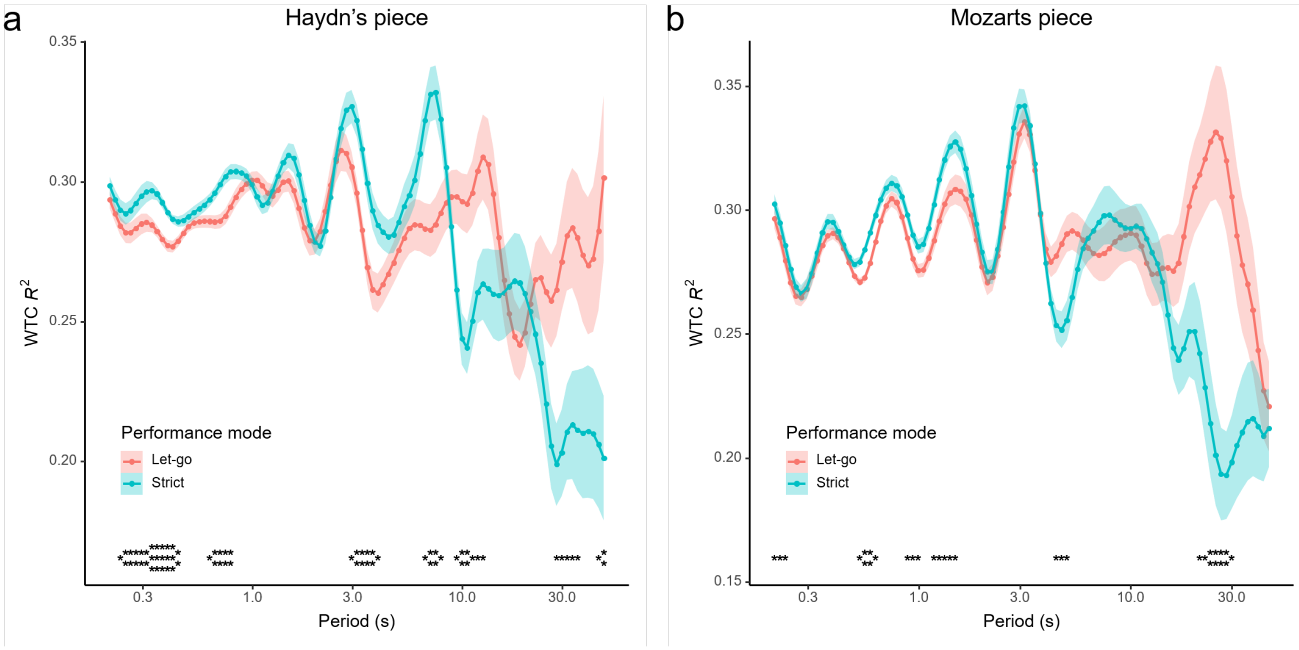

Figure 10. (a) Mean P-A sync over different timescales (periods) for the *Let-go* and *Strict* performances of Haydn's piece. (b) Mean P-A sync over different timescales for the two performances of Mozart's piece. Shaded areas represent SEM over 42 subjects. Periods with significant differences are marked by asterisks. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; FDR-corrected.

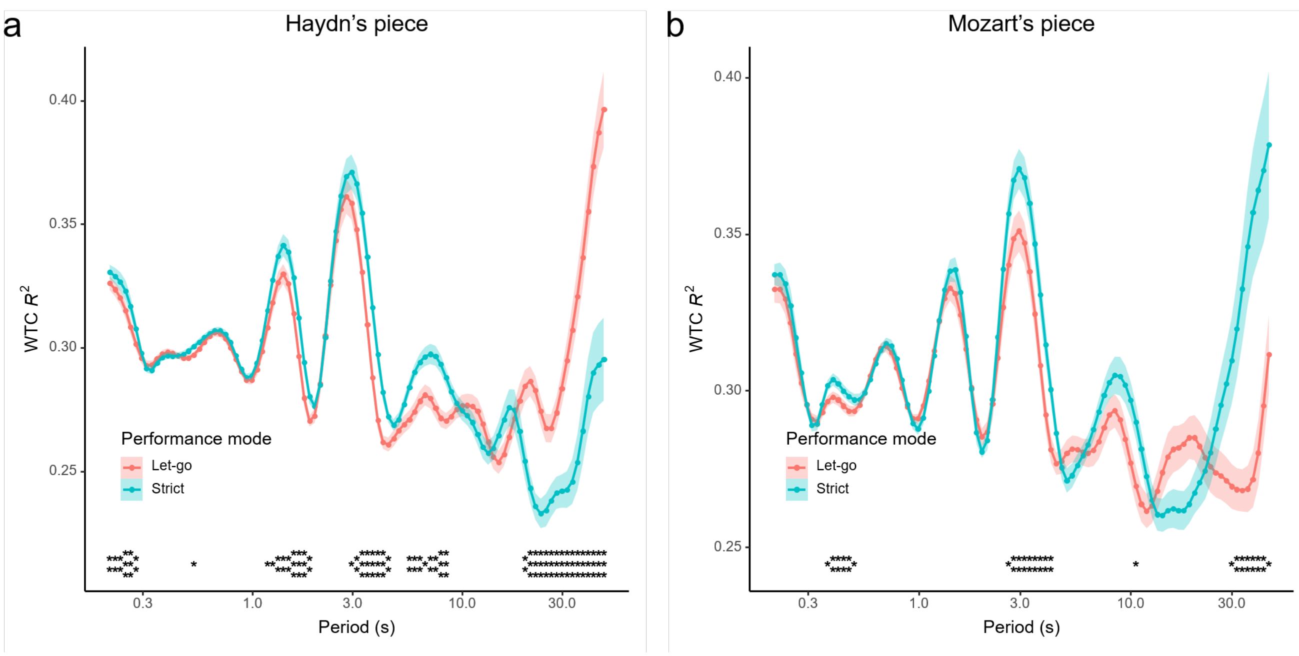

Figure 11. (a) Mean A-A sync over different timescales (periods) for the *Let-go* and *Strict* performances of Haydn's piece. (b) Mean A-A sync over different timescales (periods) for the two performances of Mozart's piece. Shaded areas represent SEM over 42 subjects. Periods with significant differences are marked by asterisks. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; FDR-corrected.

(i.e. the longer the timescale is, the slower the synchrony changes, limiting variability) and the more limited available time range due to the exclusion of the cone of influence to avoid edge effects in the WTC analysis.

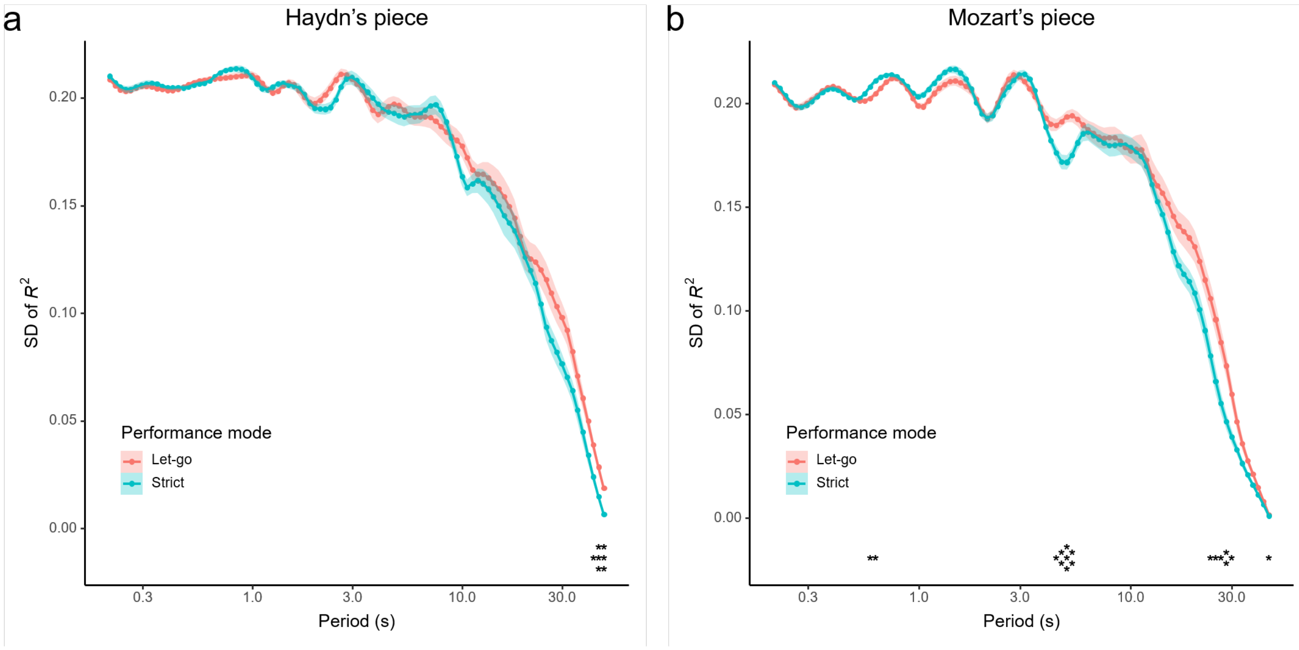

Figure 12. (a) Temporal variability (standard deviation over time) of the P-A physical sync at different timescales (periods) for the *Let-go* and *Strict* performances of Haydn's piece. (b) Mean temporal variability of the P-A sync at different timescales for the two performances of Mozart's piece. Shaded areas represent SEM over 42 subjects. Periods with significant differences are marked by asterisks. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; FDR-corrected.

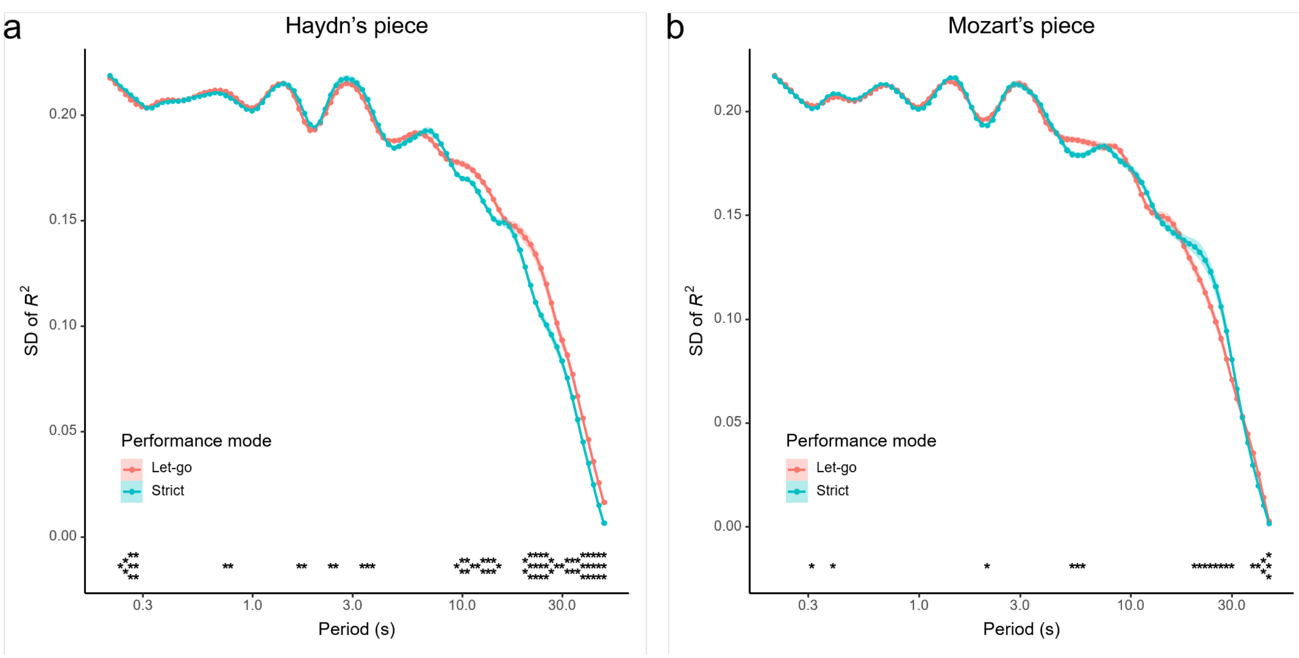

Figure 13. (a) Temporal variability (standard deviation over time) of A-A sync at different timescales (periods) for the *Let-go* and *Strict* performances of Haydn's piece. (b) Mean temporal variability (standard deviation over time) of A-A sync at different timescales for the two performances of Mozart's piece. Shaded areas represent SEM over 42 subjects. Periods with significant differences are marked by asterisks. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; FDR-corrected.

C.2 Higher order correlations in Breathing Synchrony

Along with individual entropy rate and pairwise synchrony measured using PLV, we explored higher order effects among audience members (as triplets). We used the framework of multivariate auto-regressive (MVAR) model to fit the oscillatory breathing signals for a given performance. The noise covariance matrix obtained from the model fit was then used to infer Ω - *information* and Σ - *information* for triplets of participants[79]. Average, Ω and Σ information was estimated for each participant was estimated by averaging over all triplets involving the participant.

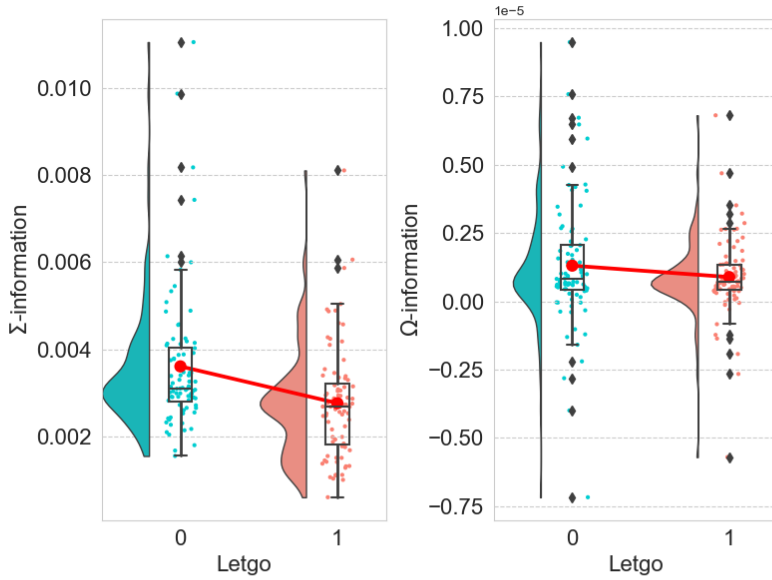

Figure 14. Higher order differences between strict and Letgo modes using triplet level (a) Σ-information and (b) Ω-information

Figure 14 shows that Σ-information, which is known to correlate with TSE complexity[79] decreased during letgo performances. Whereas, no significant change was observed for Ω-information, which measures the balance between synergy and redundancy among the parts. For the case of triplets Ω-information is equivalent to co-information, which is negative for synergistic interactions[80].

C.3 Effect of audience's vision

Figure 15(a) shows a comparison of physical activity power spectra between the blindfolded and non-blindfolded audiences. Blindfolded audiences tended to show less physical activity than those who could see the performance, but the differences were not significant. Figures 15(b) and (c) show comparison of P-A sync and A-A sync between the audiences' sight type, respectively. For A-A sync, blindfolded audiences showed higher level of synchrony in both shorter and longer time scales. Similar tendency was also observed in P-A sync, but the difference was not significant. Figures 15(d) and (e) show comparison of temporal variability in P-A sync and A-A sync between the audiences' sight type, respectively. No significant effect of sight types was observed.

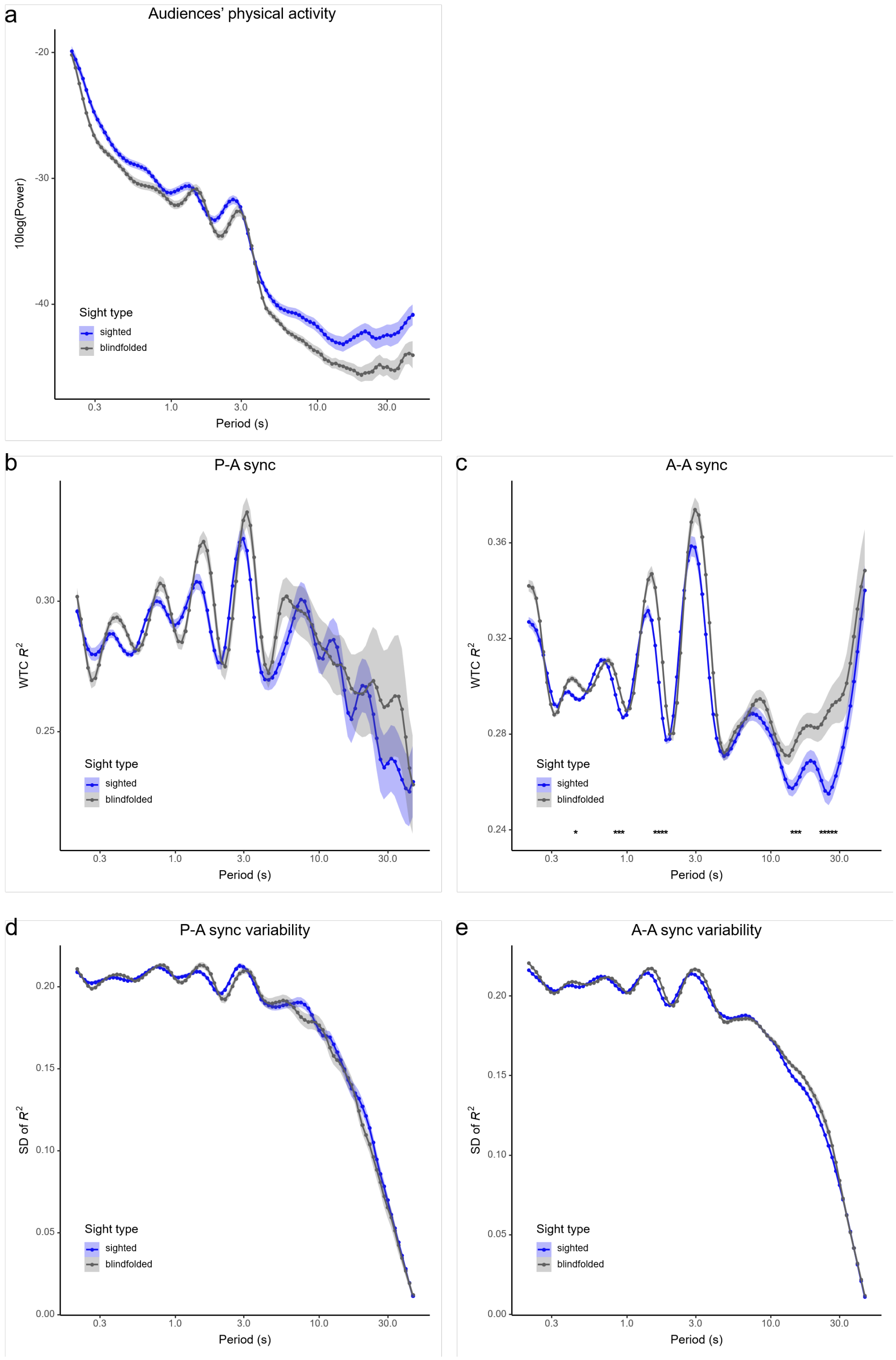

Figure 15. (a) Mean power spectra of the audience's physical activity of the two groups, comparing the effect of wearing a blindfold. Blindfolded audience show slightly less physical activity. (b) Mean P-A sync over different timescales (periods) for the two groups. (c) Mean A-A sync over different timescales (periods) for the two groups. Shaded areas indicate SEM over the four performers. Periods with significant difference between the sighted and blindfolded groups are marked by asterisks. *: p <0.05; FDR-corrected.

D Relationship between psychology and physiology

D.1 Absorption and physical synchrony

When studying the effect of Absorption and mean Synchrony on PC1, Absorption was the only significant term, and no interaction with mean synchrony in either ‘beat-sync’ or ‘music-sync’ bands was seen. When adding interactions with Blindfolding, we observe a significant interaction with the P-A ‘music-sync’ band ($p = 0.02$). The linear model using ‘music-sync variability’ shows both Absorption ($p = 0.0003$), P-A ‘music-sync variability’ ($p = 0.013$) and A-A ‘music-sync variability’ ($p = 0.018$) are significant. We also observe a significant three-way interaction between Blindfolded, Absorption, and P-A ‘music-sync variability’ ($p = 0.043$).

D.2 Analysis by composition

When separating the performances by the compositions, the analysis of the correlation between subjective ratings and physical synchrony at different timescales revealed stronger correlations and anti-correlations for the performances of Haydn’s composition (Table 4) that Mozart’s. On the other hand, for the performances of Mozart’s composition, the correlations were negligible (Table 5). These further supports the musicians’ assessment of the performance itself, with the modes of performance being more strongly differentiated in the piece by Haydn than the piece by Mozart.

Table 4. Correlation between audience ratings and mean physical synchrony and its temporal variability, with only the two pieces composed by Haydn.

	Beat-sync			**Music-sync**			**Music-sync variability**		
Rating	t_{41}	p		t_{41}	p		t_{41}	p	
				P-A sync					
PC1	−2.82	0.007	**	3.27	0.002	**	2.10	0.042	*
Improvisatory	−3.44	0.001	**	3.06	0.004	**	3.30	0.002	**
Innovative	−2.18	0.035	*	3.75	0.001	***	3.00	0.005	**
RiskTaking	−3.65	0.001	***	3.40	0.001	**	3.03	0.004	**
Engaging	−2.08	0.044	*	1.50	0.141		1.26	0.216	
Convincing	0.66	0.511		0.61	0.548		0.42	0.675	
Familiar	−0.19	0.848		1.62	0.112		1.36	0.182	
Sleepy	−0.57	0.569		−1.48	0.147		0.49	0.629	
				A-A sync					
PC1	−1.73	0.090		3.23	0.002	**	3.57	0.001	***
Improvisatory	−2.22	0.032	*	3.79	0.000	***	4.43	0.000	***
Innovative	−1.57	0.125		3.18	0.003	**	4.43	0.000	***
RiskTaking	−2.25	0.030	*	3.38	0.002	**	4.01	0.000	***
Engaging	−1.91	0.063		1.76	0.085		0.73	0.472	
Convincing	1.77	0.085		−0.35	0.727		−0.52	0.609	
Familiar	−1.73	0.091		1.29	0.203		1.43	0.160	
Sleepy	0.08	0.938		−0.19	0.853		0.56	0.576	

As an example, Figure 16 illustrates the correlations between the first principal component of the audience ratings (PC1) and P-A sync in shorter and longer timescales as well as the temporal variability of the P-A sync in the longer timescales for the pieces of Haydn’s composition.

Table 5. Correlation between audience ratings and mean physical synchrony and its temporal variability, with only the two pieces composed by Mozart.

Rating	Beat-sync t_{41}	Beat-sync p		Music-sync t_{41}	Music-sync p		Music-sync variability t_{41}	Music-sync variability p	
				P-A sync					
PC1	−0.69	0.496		0.25	0.808		0.71	0.481	
Improvisatory	−0.70	0.489		0.31	0.759		0.09	0.927	
Innovative	−0.28	0.780		0.29	0.775		1.03	0.307	
RiskTaking	−0.66	0.513		0.34	0.734		1.29	0.205	
Engaging	−0.77	0.444		0.38	0.703		1.21	0.233	
Convincing	−0.04	0.970		−0.56	0.576		1.08	0.285	
Familiar	2.42	0.020	*	−2.16	0.037	*	−2.81	0.008	**
Sleepy	0.92	0.362		−0.68	0.500		−1.16	0.253	
				A-A sync					
PC1	0.96	0.344		—	—		−0.03	0.977	
Improvisatory	0.61	0.545		—	—		−0.73	0.468	
Innovative	1.16	0.251		—	—		−0.98	0.335	
RiskTaking	1.01	0.320		—	—		−0.38	0.704	
Engaging	0.57	0.571		—	—		0.318	0.752	
Convincing	0.84	0.405		—	—		0.19	0.847	
Familiar	1.68	0.101		—	—		−1.62	0.112	
Sleepy	1.01	0.316		—	—		−1.41	0.165	

Remark: For the physical synchrony between audience (A-A sync), there were no periods of interest (timescales) where the performance modes showed significant effect of *Let-go* > *Strict* on the average synchrony (music sync).

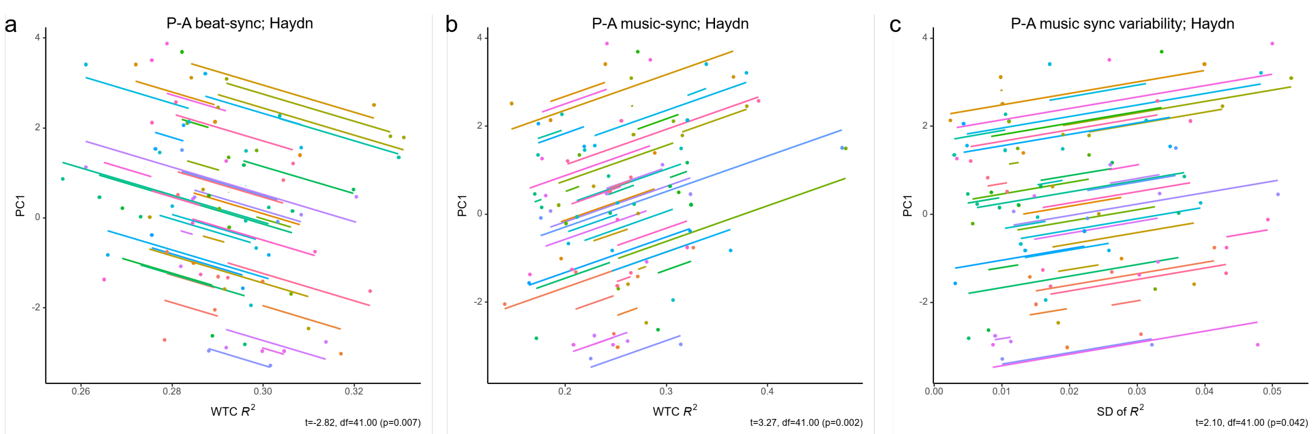

Figure 16. Relationship between the mean P-A sync in the shorter timescales (beat-sync; a), in the longer timescales (music-sync; b) and P-A sync temporal variability in the longer timescales (music-sync variability) and the audiences' perception (PC1) for Haydn's pieces. Coloured points represent the two performances with the different modes for each subject. Coloured lines indicate the best linear fit of the relationship between the synchrony and the ratings for each subject, estimated using the multilevel models with the same slope (fixed effect) and varying intercepts (random effect).

References

1. Castano, E., Yzerbyt, V. & Bourguignon, D. We are one and i like it: The impact of ingroup entitativity on ingroup identification. *Eur. J. Soc. Psychol.* **33**, 735–754, DOI: https://doi.org/10.1002/ejsp.175 (2003).
2. Bak-Coleman, J. B. *et al.* Stewardship of global collective behavior. *Proc. Natl. Acad. Sci.* **118**, DOI: https://doi.org/10.1073/pnas.2025764118 (2021).
3. Rennung, M. & Göritz, A. S. Prosocial consequences of interpersonal synchrony: A meta-analysis. *Zeitschrift fur Psychol.* **224**, 168–189, DOI: 10.1027/2151-2604/a000252 (2016).
4. Vicaria, I. M. & Dickens, L. Meta-analyses of the intra-and interpersonal outcomes of interpersonal coordination. *J. Nonverbal Behav.* **40**, 335–361, DOI: 10.1007/s10919-016-0238-8 (2016).
5. Mogan, R., Fischer, R. & Bulbulia, J. A. To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *J Exp Soc Psychol* **72**, 13–20, DOI: 10.1016/j.jesp.2017.03.009 (2017).
6. Nozawa, T. *et al.* Prior physical synchrony enhances rapport and inter-brain synchronization during subsequent educational communication. *Sci. Reports* **9**, 12747, DOI: 10.1038/s41598-019-49257-z (2019).
7. Konvalinka, I. *et al.* Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proc. Natl. Acad. Sci.* **108**, 8514–8519, DOI: 10.1073/pnas.1016955108 (2011).
8. Kettner, H. *et al.* Psychedelic communitas: intersubjective experience during psychedelic group sessions predicts enduring changes in psychological wellbeing and social connectedness. *Front. Pharmacol.* 234, DOI: 10.3389/fphar.2021.623985 (2021).
9. Higo, N. *et al.* Interpersonal similarity between body movements in face-to-face communication in daily life. *PLoS One* **9**, e102019, DOI: 10.1371/journal.pone.0102019 (2014).
10. Varni, G., Volpe, G. & Camurri, A. A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. *IEEE Transactions on Multimed.* **12**, 576–590, DOI: 10.1109/tmm.2010.2052592 (2010).
11. Trehub, S. E., Becker, J. & Morley, I. Cross-cultural perspectives on music and musicality. *Philos. Transactions Royal Soc. B: Biol. Sci.* **370**, 20140096–20140096, DOI: https://doi.org/10.1098/rstb.2014.0096 (2015).
12. Mehr, S. A., Singh, M., York, H., Glowacki, L. & Krasnow, M. M. Form and function in human song. *Curr. Biol.* **28**, 356–368.e5, DOI: https://doi.org/10.1016/j.cub.2017.12.042 (2018).
13. Clarke, E., DeNora, T. & Vuoskoski, J. Music, empathy and cultural understanding. *Phys. Life Rev.* **15**, 61–88, DOI: https://doi.org/10.1016/j.plrev.2015.09.001 (2015).
14. Laird, L. Empathy in the classroom. *Music. Educ. J.* **101**, 56–61, DOI: https://doi.org/10.1177/0027432115572230 (2015).
15. Cho, E. The relationship between small music ensemble experience and empathy skill: A survey study. *Psychol. Music.* 030573561988722, DOI: https://doi.org/10.1177/0305735619887226 (2019).
16. Keller, P. E. Joint action in music performance. In Morganti, F., Carassa, A. & Riva, G. (eds.) *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions*, 17 (IOS Press, 2008).
17. Chang, A., Livingstone, S. R., Bosnyak, D. J. & Trainor, L. J. Body sway reflects leadership in joint music performance. *Proc Natl Acad Sci U S A* **114**, E4134–E4141, DOI: 10.1073/pnas.1617657114 (2017).
18. Walton, A. E., Richardson, M. J., Langland-Hassan, P. & Chemero, A. Improvisation and the self-organization of multiple musical bodies. *Front. Psychol.* **06**, DOI: 10.3389/fpsyg.2015.00313 (2015).
19. Noy, L., Levit-Binun, N. & Golland, Y. Being in the zone: physiological markers of togetherness in joint improvisation. *Front. Hum. Neurosci.* **9**, DOI: https://doi.org/10.3389/fnhum.2015.00187 (2015).
20. Csikszentmihalyi, M. *Flow: the Psychology of Optimal Experience* (Harper and Row, New York, 1990).
21. Shehata, M. *et al.* Team flow is a unique brain state associated with enhanced information integration and neural synchrony. *bioRxiv (Cold Spring Harb. Lab.* DOI: https://doi.org/10.1101/2020.06.17.157990 (2020).
22. Brand, G., Sloboda, J., Saul, B. & Hathaway, M. The reciprocal relationship between jazz musicians and audiences in live performances: A pilot qualitative study. *Psychol. Music.* **40**, 634–651, DOI: https://doi.org/10.1177/0305735612448509 (2012).
23. Toelle, J. & Sloboda, J. A. The audience as artist? the audience's experience of participatory music. *Music. Sci.* **25**, 102986491984480, DOI: https://doi.org/10.1177/1029864919844804 (2019).

24. Jansen, E. Complexity and musical improvisation. *Music. & Sci.* **1**, 205920431877980, DOI: 10.1177/2059204318779807 (2018).

25. Cook, N. *Beyond the score : music as performance* (Oxford University Press, New York, 2013).

26. Matare, J. Creativity or musical intelligence? a comparative study of improvisation performance by european and african musicians. *Think. Ski. Creat.* **4**, 194–203, DOI: https://doi.org/10.1016/j.tsc.2009.09.005 (2009).

27. Creech, A. *et al.* Investigating musical performance: Commonality and diversity among classical and non-classical musicians. *Music. Educ. Res.* **10**, 215–234, DOI: 10.1080/14613800802079080 (2008).

28. Nettl, B. Thoughts on improvisation: a comparative approach. *The Music. Q.* **LX**, 1–19, DOI: https://doi.org/10.1093/mq/lx.1.1 (1974).

29. Levin, R. Improvising mozart. *Bull. Am. Acad. Arts Sci.* **55**, 87–90.

30. Dolan, D., Sloboda, J., Jensen, H. J., Crüts, B. & Feygelson, E. The improvisatory approach to classical music performance: An empirical investigation into its characteristics and impact. *Music. Perform. Res.* **6** (2013).

31. Leech-Wilkinson, D. Listening and responding to the evidence of early twentieth-century performance. *J. Royal Music. Assoc.* **135**, 45–62, DOI: https://doi.org/10.1080/02690400903414822 (2010).

32. Dolan, D. *et al.* The improvisational state of mind: A multidisciplinary study of an improvisatory approach to classical music repertoire performance. *Front. Psychol.* **9**, 21, DOI: 10.3389/fpsyg.2018.01341.

33. Volpe, G., D'Ausilio, A., Badino, L., Camurri, A. & Fadiga, L. Measuring social interaction in music ensembles. *Philos. Transactions Royal Soc. B: Biol. Sci.* **371**, 20150377, DOI: 10.1098/rstb.2015.0377 (2016).

34. Chang, A., Kragness, H. E., Livingstone, S. R., Bosnyak, D. J. & Trainor, L. J. Body sway reflects joint emotional expression in music ensemble performance. *Sci. Reports* **9**, DOI: https://doi.org/10.1038/s41598-018-36358-4 (2019).

35. Bernardi, N. F. *et al.* Increase in synchronization of autonomic rhythms between individuals when listening to music. *Front Physiol* **8**, 785, DOI: 10.3389/fphys.2017.00785 (2017).

36. Ardizzi, M., Calbi, M., Tavaglione, S., Umiltà, M. A. & Gallese, V. Audience spontaneous entrainment during the collective enjoyment of live performances: physiological and behavioral measurements. *Sci Rep* **10**, 3813, DOI: 10.1038/s41598-020-60832-7 (2020).

37. Czepiel, A. *et al.* Synchrony in the periphery: inter-subject correlation of physiological responses during live music concerts. *Sci. Reports* **11**, DOI: https://doi.org/10.1038/s41598-021-00492-3 (2021).

38. Tschacher, W. *et al.* Physiological synchrony in audiences of live concerts. *Psychol. Aesthetics, Creat. Arts* DOI: 10.1037/aca0000431 (2021).

39. 1. Madsen, J. & Parra, L. C. Cognitive processing of a common stimulus synchronizes brains, hearts, and eyes. *PNAS Nexus 1* **1**, pgac020, DOI: 10.1093/pnasnexus/pgac020 (2022).

40. Burger, B., Thompson, M. R., Luck, G., Saarikallio, S. & Toiviainen, P. Influences of Rhythm- and Timbre-Related Musical Features on Characteristics of Music-Induced Movement. *Front Psychol* **4**, 183, DOI: 10.3389/fpsyg.2013.00183 (2013).

41. Ellamil, M., Berson, J., Wong, J., Buckley, L. & Margulies, D. S. One in the dance: Musical correlates of group synchrony in a real-world club environment. *PLoS One* **11**, e0164783, DOI: 10.1371/journal.pone.0164783 (2016).

42. Vieillard, S. *et al.* Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cogn. & Emot.* **22**, 720–752, DOI: https://doi.org/10.1080/02699930701503567 (2008).

43. Donnay, G. F., Rankin, S. K., Lopez-Gonzalez, M., Jiradejvong, P. & Limb, C. J. Neural substrates of interactive musical improvisation: An fmri study of "trading fours" in jazz. *PLoS ONE* **9**, e88665, DOI: https://doi.org/10.1371/journal.pone.0088665 (2014).

44. Mayo, O. & Gordon, I. In and out of synchrony—behavioral and physiological dynamics of dyadic interpersonal coordination. *Psychophysiology* **57**, e13574, DOI: 10.1111/psyp.13574 (2020).

45. Haijen, E. C. H. M. *et al.* Predicting responses to psychedelics: A prospective study. *Front. Pharmacol.* **9**, DOI: 10.3389/fphar.2018.00897 (2018).

46. Grinsted, A., Moore, J. & Jevrejeva, S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process. Geophys.* **5/6**, 561–566, DOI: 10.5194/npg-11-561-2004 (2004).

47. Schmidt, R., Nie, L., Franco, A. & Richardson, M. J. Bodily synchronization underlying joke telling. *Front Hum Neurosci* **8**, 633, DOI: 10.3389/fnhum.2014.00633 (2014).

48. Issartel, J., Bardainne, T., Gaillot, P. & Marin, L. The relevance of the cross-wavelet transform in the analysis of human interaction–a tutorial. *Front Psychol* **5**, 1566, DOI: 10.3389/fpsyg.2014.01566 (2015).

49. Fujiwara, K. & Daibo, I. Evaluating interpersonal synchrony: Wavelet transform toward an unstructured conversation. *Front Psychol* **7**, 516, DOI: 10.3389/fpsyg.2016.00516 (2016).

50. Wiltshire, T. J., Steffensen, S. V. & Fiore, S. M. Multiscale movement coordination dynamics in collaborative team problem solving. *Appl Ergon* **79**, 143–151, DOI: 10.1016/j.apergo.2018.07.007 (2019).

51. Schirmer, A., Lo, C. & Wijaya, M. When the Music's No Good: Rhythms Prompt Interactional Synchrony But Impair Affective Communication Outcomes. *Commun. Res.* **50**, 30–52, DOI: 10.1177/00936502211015900 (2023).

52. Godøy, R. I. & Leman, M. *Musical Gestures* (Routledge, 2010).

53. Bernardi, L., Porta, C. & Sleight, P. Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: the importance of silence. *Heart* **92**, 445–452, DOI: https://doi.org/10.1136/hrt.2005.064600 (2005).

54. Mediano, P. A. *et al.* Spectrally and temporally resolved estimation of neural signal diversity. *bioRxiv* 2023–03 (2023).

55. Krumhansl, C. L. An exploratory study of musical emotions and psychophysiology. *Can. J. Exp. Psychol. canadienne de psychologie expérimentale* **51**, 336 (1997).

56. Aydore, S., Pantazis, D. & Leahy, R. M. A note on the phase locking value and its properties. *Neuroimage* **74**, 231–244 (2013).

57. Høffding, S. *A phenomenology of musical absorption* (Cham Palgrave Macmillan, 2018).

58. Hancock, F. *et al.* Metastability demystified – The foundational past, the pragmatic present, and the potential future. *Preprints* (2023).

59. Pressing, J. The micro- and macrostructural design of improvised music. *Music. Percept.* **5**, 133–172, DOI: https://doi.org/10.2307/40285390 (1987).

60. Lerdahl, F. & Jackendoff, R. S. *A Generative Theory of Tonal Music, reissue, with a new preface* (MIT Press, 1996).

61. Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A. & Zatorre, R. J. Predictability and uncertainty in the pleasure of music: a reward for learning? *J Neurosci* **39**, 9397–9409, DOI: 10.1523/JNEUROSCI.0428-19.2019 (2019).

62. Stupacher, J., Matthews, T. E., Pando-Naude, V., Foster Vander Elst, O. & Vuust, P. The sweet spot between predictability and surprise: Musical groove in brain, body, and social interactions. *Front Psychol* **13**, 906190, DOI: 10.3389/fpsyg.2022.906190 (2022).

63. Vuust, P., Heggli, O. A., Friston, K. J. & Kringelbach, M. L. Music in the brain. *Nat. Rev. Neurosci.* **23**, 287–305, DOI: 10.1038/s41583-022-00578-5 (2022).

64. Dolan, D. Interpretation through improvisation.

65. Nozawa, T., Uchiyama, M., Honda, K., Nakano, T. & Miyake, Y. Speech discrimination in real-world group communication using audio-motion multimodal sensing. *Sensors* **20**, 2948, DOI: 10.3390/s20102948 (2020).

66. Tellegen, G., Auke; Atkinson. Openness to absorbing and self-altering experiences ("absorption"), a trait related to hypnotic susceptibility. *J. Abnorm. Psychol.* **83**, 268–277, DOI: 10.1037/h0036681 (1974).

67. Rhodes, L. A., David, D. C. & Combs, A. L. Absorption and enjoyment of music. *Percept. Mot. Ski.* **66**, 737–738, DOI: 10.2466/pms.1988.66.3.737 (1988).

68. Stekhoven, D. J. & Bühlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**, 112–118, DOI: 10.1093/bioinformatics/btr597 (2012).

69. Torrence, C. & Compo, G. P. A practical guide to wavelet analysis. *Bull. Am. Meteorol. society* **79**, 61–78, DOI: 10.1175/1520-0477(1998)079⟨0061:APGTWA⟩2.0.CO;2 (1998).

70. Grinsted, A. Cross wavelet and wavelet coherence toolbox.

71. Phan, D., Bonnet, S., Guillemaud, R., Castelli, E. & Thi, N. P. Estimation of respiratory waveform and heart rate using an accelerometer. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 4916–4919 (IEEE, 2008).

72. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J Stat Softw* **67**, 1–48, DOI: 10.18637/jss.v067.i01 (2015).

73. Yaremych, H. E., Preacher, K. J. & Hedeker, D. Centering categorical predictors in multilevel models: Best practices and interpretation. *Psychol Methods* **28**, 613–630, DOI: 10.1037/met0000434 (2023).

74. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmertest package: tests in linear mixed effects models. *J Stat Softw* **82**, DOI: 10.18637/jss.v082.i13 (2017).

75. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc. Ser. B (Methodological)* **57**, 289–300, DOI: 10.1111/j.2517-6161.1995.tb02031.x (1995).

76. Bakdash, J. Z. & Marusich, L. R. Repeated measures correlation. *Front Psychol* **8**, 456, DOI: 10.3389/fpsyg.2017.00456 (2017).

77. Székely, G., Rizzo, M. & Bakirov, N. Measuring and testing dependence by correlation of distances. *The Annals Stat.* **35**, 2769–2794, DOI: doi:10.1214/009053607000000505 (2007).

78. Liu, Y., San Liang, X. & Weisberg, R. H. Rectification of the bias in the wavelet power spectrum. *J Atmos Ocean. Technol* **24**, 2093–2102, DOI: 10.1175/2007JTECHO511.1 (2007).

79. Rosas, F. E., Mediano, P. A., Gastpar, M. & Jensen, H. J. Quantifying high-order interdependencies via multivariate extensions of the mutual information. *Phys. Rev. E* **100**, 032305, DOI: 10.1103/PhysRevE.100.032305 (2019).

80. Bell, A. J. The co-information lattice. In *Proceedings of the fifth international workshop on independent component analysis and blind signal separation: ICA*, vol. 2003 (2003).