Listening Behaviors and Musical Coordination in Collective Free Improvisation

Arthur Faraco, Armand Schwarz, Coralie Vincent, Patrick Susini, Emmanuel Ponsot, Clément Canonne

sempre: Society for Education, Music and Psychology Research
MUSIC & SCIENCE

Research Article

Music & Science
Volume 7: 1–18

DOI: 10.1177/20592043241257023
journals.sagepub.com/home/mns
Sage

Listening Behaviors and Musical Coordination in Collective Free Improvisation

Arthur Faraco[1], Armand Schwarz[2], Coralie Vincent[2], Patrick Susini[2], Emmanuel Ponsot[2] and Clément Canonne[2]

Abstract

While empirical studies on joint music-making have shed light on many aspects of ensemble performance in the past few decades, the role of auditory attention in such a context has remained strikingly understudied. We draw here on a self-annotation methodology to investigate musicians' listening behaviors in freely improvised performances. Six trios of professional musicians were asked to freely improvise in a recording studio, hearing each other only through headphones. While they were playing, the loudness of each musician as sent to the other two musicians' headphones was covertly increased/decreased during random periods of time in order to enhance its relative saliency within the musical scene. Immediately after each improvisation, musicians were asked to listen to the improvisation that they had just performed and to continuously indicate, using a specific application, where their listening focus was as they were performing. The results demonstrated that during periods of loudness manipulation, musicians' attention was significantly drawn to the musician who had been made more salient. Two follow-up studies then investigated the extent to which joint auditory attention correlated with perceived togetherness, as well as whether improvisers' auditory attention during the performance aligned with that of external listeners attending to the recording of the performance. Taken together, our results suggest that, beyond the effects of saliency, musicians also tend to strategically adapt their listening behavior to the specificities of the interactional context and that musicians' collective listening behaviors have an impact on the performance, both at an acoustic level and at a perceptual level. By relying on attentional patterns that dynamically emerged from complex, ecological musical interactions, our studies provide a first attempt at assessing the effects of auditory attention on coordination and contribute to establishing sonic interactions as a promising setting for the study of the effects of joint attention.

Keywords

Improvisation, joint action, listening, saliency, togetherness

Submission date: 6 February 2024; Acceptance date: 3 May 2024

Introduction

The importance of listening for collective improvisation—whatever the medium—can't be understated. Students attending workshops of theatrical improvisation quickly learn that "when building a Long Form improv scene with someone else, there is nothing more important than listening […]. Listening needs to happen so that you and your scene partner are on the same page […]. A great improviser is a good listener" (Besser et al., 2013, p. 36); in jazz, "to say that a player 'doesn't listen' is a grave insult": on the contrary, jazz players aim at "attending to what everyone else is doing in the band" in order to engage in "active listening – being able to respond to musical opportunities or to correct mistakes" (Monson, 1996, p. 84); and even in dance improvisation, listening plays a role, not only in

[1] Universidade de Sao Paulo, Sao Paulo, Brazil
[2] STMS UMR 9912 (CNRS/IRCAM/SU), Paris, France

Corresponding Author:
Clément Canonne, STMS UMR 9912 (CNRS/IRCAM/SU), 1 Place Igor Stravinsky, Paris 75004, France.
Email: clementcanonne@hotmail.com

Data Availability Statement included at the end of the article

a metaphoric sense, but also in the very real sense of paying auditory attention to the sounds made by the other dancers' bodies in order to better connect with them at the somatic level (Johnson, 2012).

Collective Free Improvisation (CFI)—a musical genre in which the very shape and content of the performance emerge from the unfolding interaction between the musicians (Borgo, 2005), without relying on pre-defined plans or pre-existing musical structures (Canonne, 2018; Pressing, 1984; Saint-Germier & Canonne, 2020)—is no exception, quite the contrary. For John Corbett, part of the aesthetic pleasure we have in listening to CFI precisely lies in tracking how musicians listen to each other, to who they are listening to, and, sometimes, how they deliberately choose to *not* listen to one another (Corbett, 2016). Alain Savouret, who taught free improvisation at Paris' Conservatory for more than twenty years, even went on stating that, in collective free improvisation, "L'entendre génère le faire [hearing commands playing]"—thus making of listening the *primum movens* of every collective improvisation (Savouret, 2010). In short, in CFI, musicians listen to each other not only to assess whether their own individual contribution suits the others' but also, at least in many cases, to ensure communication (Goupil et al., 2021; Pelz-Shermann, 1998) or even to simply decide what to play and how to interact with the others (Golvet et al., 2024).

But while empirical studies on joint music-making have shed light on many aspects of ensemble performance in the past few decades (see Wöllner & Keller, 2017, for an overview), the role of auditory attention in such a context has remained strikingly understudied. For example, Peter Keller, who has greatly contributed to the field of ensemble performance studies, generally lists three crucial ensemble skills—anticipation, adaptation, and attention—but mainly relies on the former two in his ADAM model of ensemble performance (Van Der Steem & Keller, 2013), attention being reduced to issues related to prioritized integrative attending (Keller, 2008), a form of unequally divided attention between the self and the others. However, understanding to what exactly (or to who, in cases where there are more than two musicians involved) musicians are paying attention, and the temporal dynamics that might underlie such shifts in attentional focus, seems of crucial importance in a kind of interaction in which much of the relevant information and affordances (Clarke, 2005) that guide the unfolding of the joint action precisely depends on the exchange of sonic information—over and beyond the much more studied visual information (Moran et al., 2015) that can certainly act as a coordination smoother (Saint-Germier & Canonne, 2020), but generally cannot be regarded as a necessary ingredient for musicians' coordination, as suggested by various empirical evidence (see e.g., Bishop et al., 2022).

Part of the reason for such a lack of empirical studies might have to do with the absence of tools, as it is not clear how musicians' listening—and the sonic source(s) that they are selecting as their main focus of auditory attention—could be dynamically tracked in the course of the performance, in the same way that visual attention can be tracked through eye-tracking devices. Eye-tracking devices could still be used in the context of collective music-making, as visual and visuospatial information is certainly a relevant part of ensemble play. However, while it is likely that auditory attention and visual attention overlap to some extent, they might also dissociate at some point or in some contexts. It is for example not rare to observe concerts of freely improvised music in which musicians play with their eyes closed during the whole performance, or without making much eye contact with their co-performers, but it would make little sense to conclude from that that the musicians are not listening to each other. Recent advances in auditory attention decoding, based on the modeling of neural information collected through mobiles EEG (Straetmans et al., 2022), offer promising perspectives, but we are still far from being able to use such methods in a setting as sonically and interactionally complex as collective free improvisation—even though EEG has sometimes be used in the context of collective improvisation to demonstrate patterns of intra- and inter-brain synchronization, as well as an extended hyper-brain network involving the coupling between instrumental sounds produced and brain signals, that could point to musical roles during improvisation (Müller & Lindenberger, 2019; Müller et al., 2013). Relying on post-hoc verbalizations through interviews with musicians (Seddon, 2005) is of course an option but such verbalizations raise numerous challenges, from the difficulty to collect reports from multiple performers as soon as possible after the performance (given how fleeting performers' memories of their attentional foci are bound to be) to the fact that verbal self-reports are often discontinuous, as they tend to focus on "remarkable" episodes, leaving the researchers with large gaps within the performance during which the musicians' auditory attention is not accounted for and making it impossible to systematically cross-compare performers' listening behaviors with the fine-grained evolution of the music.

In the present article, we propose to draw on a self-annotation methodology to explore musicians' directional listening—to whom they are listening—in freely improvised group performances. In a nutshell, the methods consist in having the musicians perform without any specific constraints (besides duration constraints) and then asking them immediately after to continuously report on a digital interface where their attentional focus was (e.g., was their attentional focus more on themselves or on one of the other musicians?) during the performance while listening back to a recording of said performance.

A similar methodology has already been used in a case study dedicated to the listening strategies of a string quartet performing a work of indeterminate music by composer Éliane Radigue (Majeau-Bettez et al., 2023). However, this previous work suffered from two important limitations. First, there was no experimental validation of the annotation method used in the study. In particular, it remained a possibility that, in the annotation task, the musicians were reporting whom they were paying attention to *when listening to the recording*, rather than whom they

were paying attention to *during the performance*. Second, the study was based on a single performance, making it difficult to rely on inferential statistics and systematic corpus analyses to draw more general conclusions. The study introduced in this article precisely aims at addressing those two issues, by investigating the directional listening of six trios of professional musicians—each trio performing two long-form improvisations—and by independently and covertly manipulating in real time the acoustic signal produced by the musicians with the aim of inducing shifts in their listening behaviors while presenting them with a recording in the annotation step that did not contain any traces of these manipulations—the hypothesis being that if the musicians were impacted by our acoustic manipulations in their annotations, despite not hearing them anymore in the recording presented to them, then this should mean that the musicians are indeed reporting their past listening behavior, rather than their present listening behavior.

Using this methodology, we had three goals in mind. First, we wanted to gather some general observations about the distribution of attentional foci in a CFI setting, both at the individual and at the group level: Are musicians more likely to focus on one particular musician (themselves included) or are they more likely to try to pay similar attention to everyone in the group? To what extent do the various musicians in a group tend to focus on the same source (joint listening) or to one another (mutual listening)? Second, we wanted to investigate how musicians' auditory attention is modulated by the musical context; in particular, are there specific attentional patterns at the boundaries between the various parts or sequences that comprise an improvised performance? Third, and finally, we wanted to assess the extent to which auditory attention and coordination are correlated: Are musicians who listen more strongly to one another also more coordinated with one another on an acoustic level? A similar issue was recently addressed by Bishop (2023), who aimed to assess the effects of joint and mutual attention between musicians on their feeling of togetherness. However, she did so by scripting highly general goals to the musicians beforehand which unfortunately mixed, on the one hand, directional attention (attention to a given source) with aspectual attention (attention to a given aspect of the musical output), and, on the other hand, attentional goals with performance goals (e.g., "focus on your partner, and try to synchronize well"); and while the study did reveal interesting results (with joint and mutual attention found to somehow strengthen feelings of togetherness), it did not address how auditory attention could shape musicians' performance on a more fine-grained temporal scale. By relying on attentional patterns that dynamically emerged from complex, ecological musical interactions, our study provides a first attempt at assessing the effects of auditory attention—as reported by the musicians themselves—on coordination.

Two additional follow-up studies, based on our corpus of annotated performances, allowed us to complement our initial set of analyses by investigating, first, the extent to which improvisers' listening behaviors would impact the perception and appreciation of their musical performances by external listeners; and second, whether improvisers' auditory attention during the performance would simply align with that of external listeners attending to the recording of the performance, or whether it would tend to follow its own logic, different from that of external listeners. Taken together, our three studies thus aim to provide a comprehensive picture of the strategic dimension of listening in ensuring coordination in freely improvised musical interactions.

Study 1

Methods

Participants. Eighteen improvisers participated in the experiment (mean age = 39.6, $SD = 9$; 12 male, 2 female, 1 undefined, 1 non-binary and 2 that did not provide this information), divided into six trios. They were highly trained musicians (with a mean of 27.6 years of musical practice, $SD = 8.1$ years) and had significant experience with collective free improvisation (with a mean of 18 years of practice, $SD = 8.9$ years). The overall instrumentation was saxophone ($N = 5$), guitar ($N = 4$), trumpet ($N = 2$), drums ($N = 2$), piano, clarinet, bass clarinet, double bass and electronics. One participant also used voice during improvisations. Trios were intentionally composed to minimize potential effects of familiarity. Consequently, the majority of musicians within each trio were unacquainted or had not previously collaborated together. Participants assessed their prior familiarity with the other members of their trio using a seven-point Likert scale. As expected, mean familiarity was low ($M = 1.89$, $SD = 1.84$). All participants gave their informed written consent for the collection, use, and publication of their data (audio files and annotation files) and were compensated at the standard rate for the employment of professional musicians.

Procedure. The experiment was held in a professional recording studio. Each musician of the trio was allocated to an individual booth, so that the musicians could not see each other and were only able to listen to the overall sound scene produced—i.e., the amplified sound of their own instrument as well as the sound from the other two instruments—through professional headphones (Beyerdymamic DT 770 pro, 80 ohms). Importantly, the musicians' headphones were panned in such a way that they heard one improviser completely on the right side and one completely on the left side, while hearing themselves in the middle. The panning (i.e., which musician is heard on the right side, which is heard on the left side) was made randomly. Each musician could adjust the overall sound level of the total mix of the headphone so that they felt comfortable with it, and this setting was then kept unchanged for the whole experiment. The musicians were asked to freely improvise together twice, for roughly 7–10 min each time. This resulted in 12 improvisations with a mean duration of 460.8 s ($SD = 86.9$).

For each trio, one of the improvisations was subjected to a real-time acoustic manipulation: Two trios began with the manipulated improvisation, while the other three went through the manipulation during the second improvisation.[1] The manipulation consisted of introducing, at various points of the performance, variations of the perceived loudness (root mean square [RMS] levels) of a given musician during a 10 s window. There were two distinct patterns of variation: a *crescendo*, perceived as an increase in sound level, and a *decrescendo*, perceived as a decrease in sound level. Taking the musician's actual sound level as the baseline, the crescendo pattern featured a + 5 dB increase in the signal's amplitude over 2.5 s, sustained the heightened level for 5 s, and then decreased over 2.5 s back to the baseline level. Conversely, the decrescendo pattern started with a −5 dB decrease from the baseline over 2.5 s, maintained the reduced level for 5 s, and finally increased over 2.5 s back to the baseline. In order to implement these manipulations, six different automation tracks were created for each trio, each of them defining the moments in which these variations would occur for each musician. Importantly, the tracks were created in such a way as to independently manipulate the way the sound produced by a given musician would be presented to the mix received by the two others. For example, a given musician A could be made louder for musician B but softer for musician C (see Figure 1, e.g., at 1′30″). Finally, these automation tracks were applied to each musician's signals in real time during the improvisation through a Max-Msp patch integrated with the *ProTools* recording session.

In our two experimental conditions (i.e., with or without the real-time manipulations), the musicians were recorded separately, with final individual mixes (different for each musician, due to the individual-specific panning and the overall level of the mix freely adjusted prior to the experiment) being prepared at the end of each improvisation by the audio engineer in charge of the recording. Immediately after each improvisation, the musicians were asked to listen to the improvisation that they had just performed and to continuously indicate in a specific application where their listening focus was as they were performing. This application stored data points (annotations) in real time and could also provide information for the experimenters via a monitoring panel (Golvet et al., 2024; Matuszewski, 2019). Each participant was provided with an individual laptop (*MacBook Pro*) that ran the application. The musicians used the same DT 770 pro headphones to listen back to the individual mix they had while playing. Crucially, when they listened back to the improvisations in which their loudness was manipulated, the mixes presented to them were the original mixes, as recorded through ProTools before applying any of the acoustic manipulations. In other words, they heard the others' signals as they were actually played, and not as they heard them through our acoustic manipulations while performing.

The graphical interface for the annotation was an inverted triangle, with a white dot that participants could move freely within the triangle using the computer's trackpad (see Figure 2). Every 50 ms, the position of the dot was

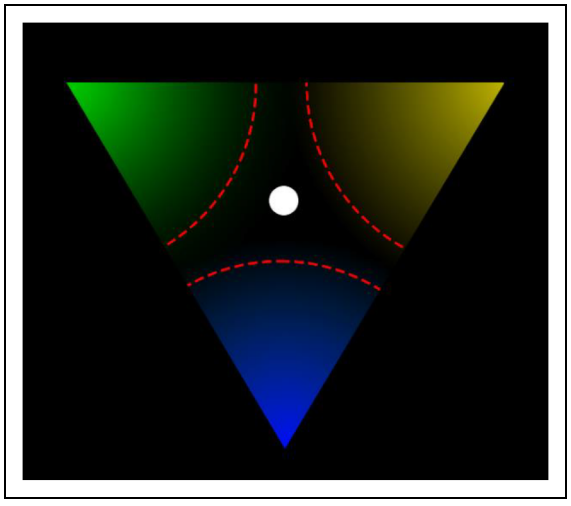

Figure 2. Graphical interface used by the participants to record their listening focus during improvisations. It featured a dynamic, draggable white dot that participants could move to continuously indicate their listening focus as they were listening back to their performance. Dashed lines delineate different zones, providing a visual guide to help musicians to accurately convey the intensity of their attentional focus.

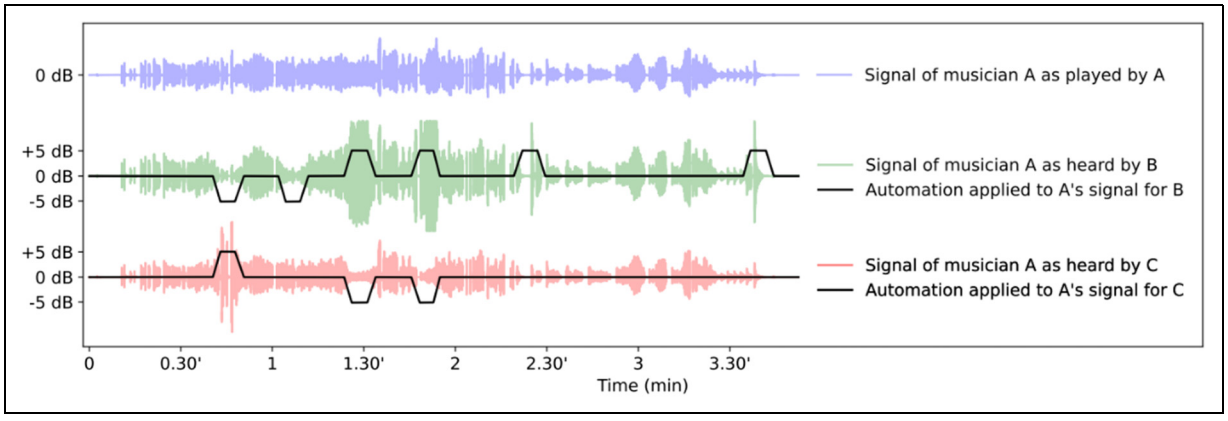

Figure 1. An example of the real-time loudness manipulations applied during study 1.

recorded, together with the current time of the audio file of the recorded improvisation.

The following instructions were given to the musicians for the annotation task:

> - The more you were listening to yourself during the performance, the more you move the dot toward the bottom vertex.
> - The more you were listening to the musician you heard on the right channel of your headphone during the performance, the more you move the dot toward the right vertex.
> - The more you were listening to the musician you heard on the left channel of your headphone during the performance, the more you move the dot toward the left vertex.

While the interface was indeed continuous (with proximity toward a specific vertex representing the intensity to which auditory attention was directed toward the musician associated with that vertex), dashed lines were inserted in order to delimit "zones" and aid participants in defining their listening behavior. As such, the triangle was divided into four main zones, with three zones meant to indicate local listening (one for each member of the trio) and the center of the triangle meant to represent global listening (auditory attention roughly equally divided between the three members of the trio).

Data Processing. Data obtained from the annotation interface were linearly interpolated between time points with a 4 Hz resolution to reduce the sheer size of these data. As mentioned above, the configuration of vertices assigned to each musician varied, with the bottom vertex consistently representing "self," while the top vertices corresponded to the two other musicians, as determined by the panning of their headphones. Therefore, to establish a consistent frame of analysis across different musicians, we applied a rotation matrix to the triangles, aligning the vertices in a standardized order with an arbitrary frame of reference. This ensured uniformity in the positioning of vertices for comparative purposes, with each vertex representing a musician (this also allowed us to create a video example of the three musicians' annotations synchronized with the music, which can be accessed here: https://figshare.com/s/13bd052138d57aa7967e).

Variables. As shown in Figure 3, for any given musician, we defined *the degree of focus toward another musician* as the Euclidean distance between the normalized position of the dot and the vertex associated with that musician (a lower value thus means a higher degree of focus toward that musician). In particular, for any given musician, we defined the *degree of self-listening* as the Euclidean distance between the normalized position of the dot and the vertex associated with the annotating musician (a lower value thus means a higher degree of self-listening).

We also defined the *degree of reciprocal listening* between any two musicians A and B by the mean between, on the one hand, the Euclidean distance between the position of A's dot and the vertex associated with B, and, on the other hand, the Euclidean distance between the position of B's dot and the vertex associated with A (a lower value thus means a higher degree of reciprocal listening).

Finally, we defined the *degree of similarity of listening behaviors* within the trio by the mean Euclidean distance between all three dots within the triangle (a lower value thus means a higher degree of similarity of listening behaviors).

To provide some more descriptive results, and to be able to cluster our recorded corpus into well-defined sequences characterized by various listening modes, we also used the dashed lined in our interface as a basis to define, for each musician and at each time, different listening modes: global listening, self-listening, and others-listening (see Figure 3a). By examining the positioning of any given pair of musicians, we could then determine their collective listening mode at that time. The following collective listening modes were defined:

- *Joint Global Listening:* Both participants are in the center of the triangle.
- *Joint Local Listening:* Both participants are together within the zone associated with the same musician.
- *Mutual Listening:* Each participant is in the zone associated with the other musician (for example, musician A listens to B and musician B listens to A).
- *Divergent Listening:* The two participants are in zones associated with distinct musicians, excluding the case described as mutual listening (for example, A listens to B but B listens to C; A listens to A and B listens to C; or A listens to A and B listens to B).
- *Composite Listening:* One musician is in global listening mode (i.e., in the center zone), while the other musician is in local listening mode (i.e., in one of the zones associated with a given musician).

Importantly, these definitions provide an *exhaustive* set (i.e., one that fully covers the space of possible behaviors) of *independent* (i.e., that cannot overlap at any time) listening modes to characterize the listening behavior of any defined pair of musicians within a given trio. Examples of the positionings corresponding to each of these collective listening modes can be seen in Figure 3b.

Statistical Analysis. In our statistical analysis, we primarily relied on linear mixed-effects models (LMMs), using categorical variables at different levels (e.g., moments with and without audio manipulation) to examine their effects on listening behaviors. LMMs offer greater analytical power compared to traditional comparison tests like repeated measures ANOVA or dependent *t*-tests, particularly in addressing the non-independence of our data (Brauer & Curtin, 2018). LMMs were also preferred over other potential analysis, such as multivariate methods, due to the nested structure of our experiment (participants in different trios). Additionally, our data collection method (self-annotation) yields highly auto-correlated data, which needs to be

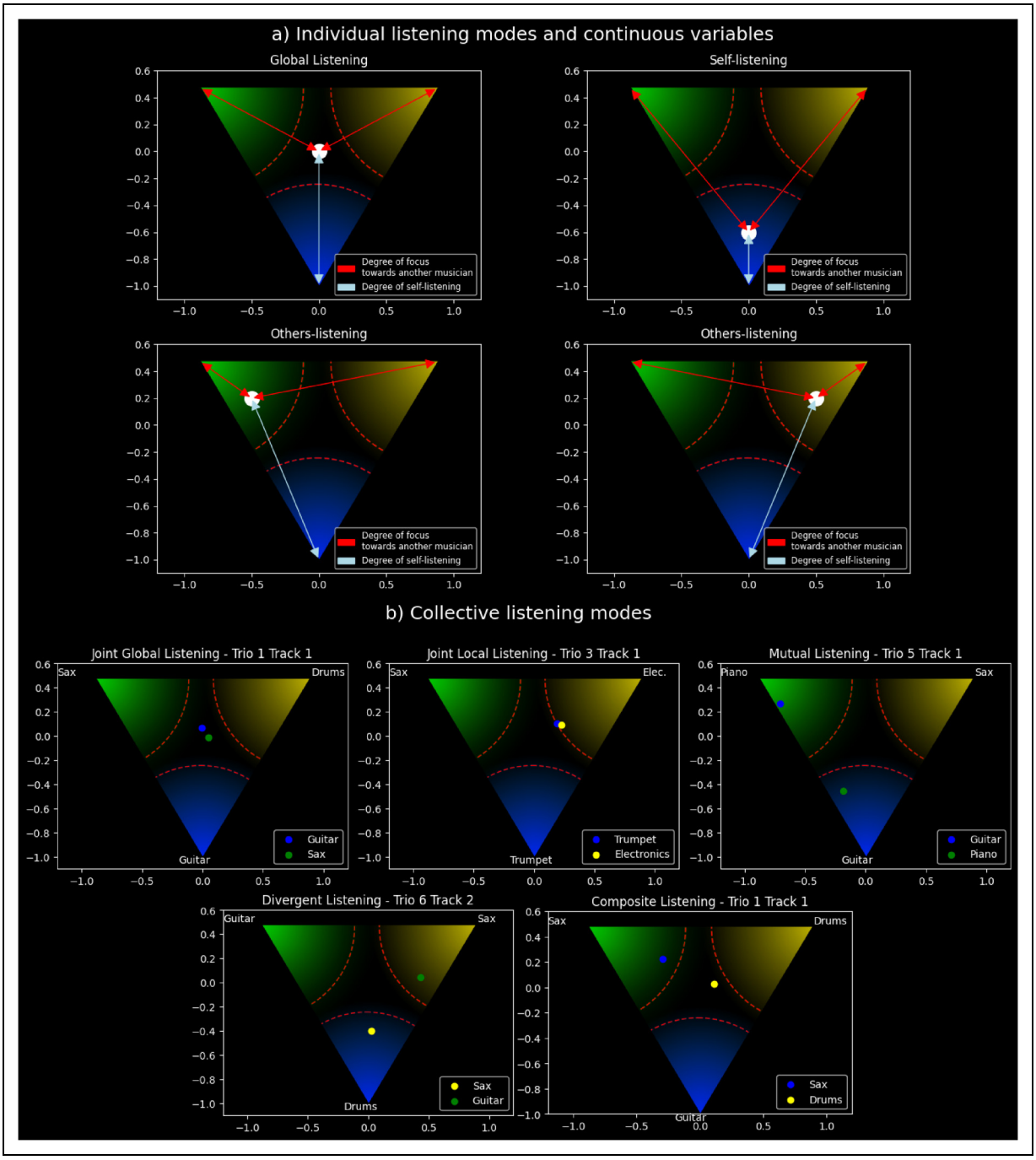

Figure 3. Examples of musicians' positioning considered either individually (a) or in pairs (b).

modeled in order to have unbiased estimates. Thus, our analyses were conducted by incorporating track number and participant ID as random intercepts and by applying an AR(1) correlation structure to manage the high autocorrelation in our dataset. All models were compared against a null model using a likelihood-ratio test (Gelman & Hill, 2006) and were fitted using the *nlme* (Pinheiro & Bates, 2000) and *lme4* (Bates et al., 2015) libraries in R.

Acoustic Analysis. As the musicians were recorded individually in separate booths, we had access to each musician's individual tracks. This allowed us to calculate audio descriptors of interest and explore potential relationships between these descriptors and the listening behaviors reported by the musicians in their self-annotations. Specifically, we computed two main audio descriptors often used to account for coordination in CFI: RMS and spectral centroid (Golvet et al., 2024; Goupil et al., 2021). RMS is indicative of the loudness in each musician's signal, while the spectral centroid provides important information on the signal's timbre (more specifically on its brightness). These two audio descriptors were computed

using the python library *Librosa* (McFee et al., 2015) from each musician's individual WAV files, with a 46 ms window size (the default settings of the functions). They were linearly interpolated between time points with a 4 Hz resolution, in order to match our dataset of annotations.

Results

Performers' Offline Annotations Were Impacted by the Online Acoustic Manipulations. A potential issue with our post-hoc annotation methodology is that participants could tend to annotate their present listening behavior (what they are paying attention to as they are listening back to the recording) rather than their past listening behavior (what they were paying attention to while playing).

Our experimental manipulation was precisely designed to address this issue. Previous research on acoustic saliency indeed showed that variations in loudness have the potential to catch listeners' auditory attention (Dalton & Lavie, 2004; Huang & Elhilali, 2017; Kaya et al., 2020). Introducing loudness variations during the performance while removing such variations from the recording listened to by the performers was thus a way to ensure that potential effects on musicians' annotations associated with our loudness manipulations would be driven by what occurred during the past performance rather than during the present listening session. Our hypothesis was that if musicians were indeed accurately recalling their listening focus during moments of variation, there would be a significant decrease in the distance between the annotating musician's position in the triangle and the vertex associated with the musician whose signal was manipulated.

We thus analyzed the effects of real-time manipulations on the degree of focus toward another musician. To do so, we compared for each musician the distance to the vertex associated with the manipulated musician during the 10 s window in which that musician underwent the saliency manipulation and during the 10 s window immediately preceding the manipulation. For each pattern (both crescendo and decrescendo), a linear mixed-effects model was fitted, using track number and participant ID as random intercepts and using an AR(1) correlation structure. This model included a categorical predictor variable with three levels (baseline: no variation; condition1: with crescendo; condition2: with decrescendo) and the distance to the vertex associated with the manipulated musician as the dependent variable. The results show that, in both crescendo and decrescendo scenarios, the degree of focus toward another musician was higher (i.e., the distance toward that musician's vertex was lower) when that musician underwent a saliency manipulation than in the 10 s window preceding the manipulation ($\beta = -0.037$, $SE = 0.011$, $t = -3.16$, $p = .0015$ for crescendo cases; $\beta = -0.042$, $SE = 0.015$, $t = -2.83$, $p = .004$ for decrescendo cases) (see Table 1 and Figure 4). A post-hoc pairwise test (estimated marginal means with the coefficients of the model, using the *emmeans* package in R) was also run to test the difference between the distances in crescendo and decrescendo manipulations in comparison to the distances before manipulation, and no significant differences between the crescendo and decrescendo patterns were found (see Table 2 and Figure 4).

Table 1. Coefficients from a linear mixed-effects model including two single-level categorical predictors corresponding to the moments where a loudness variation was induced (factor1: crescendo; factor2: decrescendo) as well as the 10 s preceding these moments (Intercepts).

Dependent variable	Factor	Estimate	*SE*	*T*	*p*
Degree of focus toward another musician	Intercept	0.987	0.029	32.9	0
	Crescendo	−0.036	0.011	−3.11	.002**
	Decrescendo	−0.044	0.015	−2.97	.003**

First, the presence of the observed effects validates our approach by demonstrating that our participants were indeed reporting on their past listening behaviors rather than on their present ones (since, again, the acoustic manipulations were not present anymore in the recordings used for the annotation task). Second, these results show that variations in loudness, whether a crescendo or a decrescendo, both attracted musicians' auditory attention significantly, in line with findings from previous studies, and that, strikingly, sudden drops in RMS level seemed to attract musicians' attention equally to sudden rises.

Distribution of Listening Modes. As previously mentioned, we defined three main zones in our annotation interface: global listening, self-listening, and others-listening (Figure 3a). The proportion of time spent by each participant in each of these three zones is shown in Figure 5, while the overall proportion of time spent in each zone by all musicians across our entire corpus is shown in Figure 6a. The musicians in fact spent the majority of their time engaged in global listening. This tendency is probably due to the very nature of CFI, in which musicians must constantly assess the contextual relevance of their own sonic actions against the overall group sound. But a significant amount of time was also devoted to more local forms of listening, whether to one specific other musician or to one's own sound. Global listening was also more stable compared to self-listening and others-listening, that is, the musicians tended to stay in the global position for a greater amount of time, as shown by an ANOVA test that revealed significant differences between the mean duration of individual listening modes ($F = 10.7$, $p < .001$, see Table 3 for Tukey HSD post-hoc pairwise analysis; see also Figure 7a).

We also looked at the distribution of collective listening modes across our entire corpus. This is shown in Figure 6b. Given that the musicians individually spent most of their time engaged in global listening, it is not surprising that

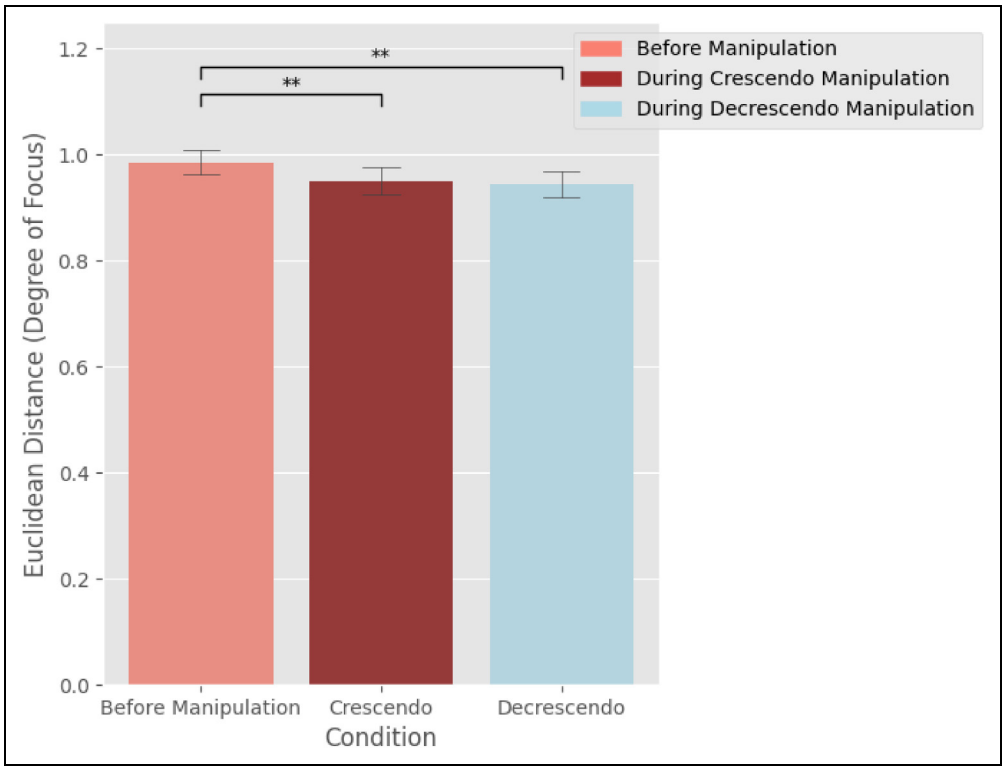

Figure 4. Impact of loudness manipulations on the degree of focus toward another musician (lower values mean greater degree of focus). Error bars show standard error (95% interval), and the black asterisks show significant differences (** for $p < 0.01$; *** for $p < 0.001$).

Table 2. Post-hoc pairwise comparison with estimated marginal means of the coefficients of the model.

Pairs	Estimate	*SE*	*t*	*p*
Before manipulation × crescendo	0.036	0.011	3.11	.005**
Before manipulation × decrescendo	0.044	0.015	2.97	.008**
Crescendo × decrescendo	0.007	0.18	0.43	.9

collective listening modes in which at least one musician is engaged in global listening (i.e., Composite Listening and Joint Global Listening) made up most of the performance time. However, there was one important difference between Composite Listening and Joint Global Listening: Joint Global Listening indeed appeared to be much more stable (i.e., pairs of musicians tended to remain engaged in joint global listening for longer stretches of time) than all other listening modes (including Composite Listening), as confirmed by an ANOVA test, which revealed significant differences in the mean duration of the various collective listening modes ($F = 11.92$, $p < .001$; see Table 3 for the Tukey HSD post-hoc pairwise analysis; see also Figure 7b). This establishes Joint Global Listening as the "default" listening mode, with other listening modes appearing in a more sporadic way, either because of their transient nature or because they were only triggered by specific contexts.

A remaining question is thus whether those various collective listening modes tend to be associated with specific interactional patterns. This will be investigated below, focusing on the listening modes that are arguably most meaningful: Joint Global Listening (a joint monitoring of the overall sound), Joint Local Listening (a shared attraction to a same sonic source), Mutual Listening (reciprocal attention), and Divergent listening (paying attention to different sonic sources).

Listening Behaviors Change at Segmentation Points. As suggested by Majeau-Bettez et al. (2023), listening might play a strategic role in negotiating the unfolding of a collective musical performance, with a particular listening behavior (both at the individual and at the collective levels) being privileged depending on the interactional and musical context. We thus investigated whether the musicians would tend to rely on specific listening behaviors in moments of articulation between the various parts or sequences that comprised their joint performance (Canonne & Garnier, 2015). Even though there are no explicit idiomatic rules or conventions that prescribe how CFI performances are formally organized, such performances are typically characterized by a segmental form, i.e., consisting in a succession of *sequences*, each having a stable musical identity of its own. This feature of CFI has been independently described by several analysts (Bertolani, 2019; Borgo, 2005; Burrows & Reed, 2016; Canonne & Garnier, 2012, 2015). The passage between one sequence to another can then be

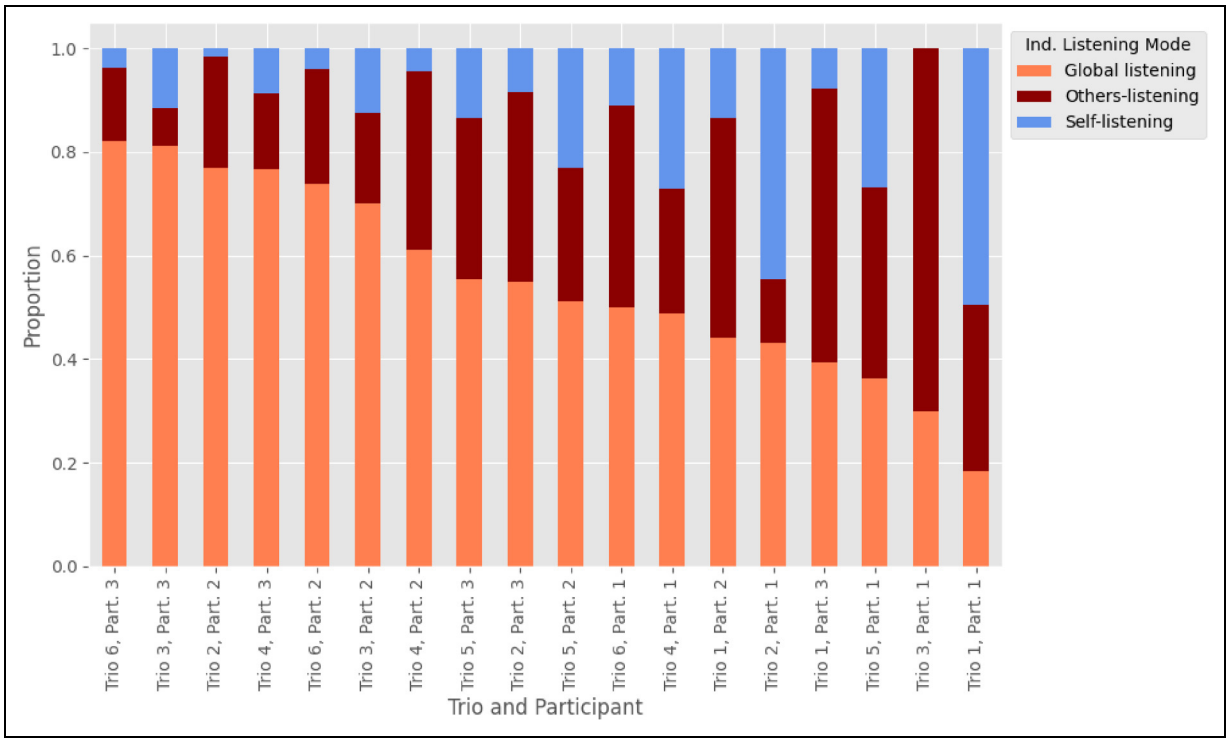

Figure 5. Proportion of time spent by each participant in each individual listening mode.

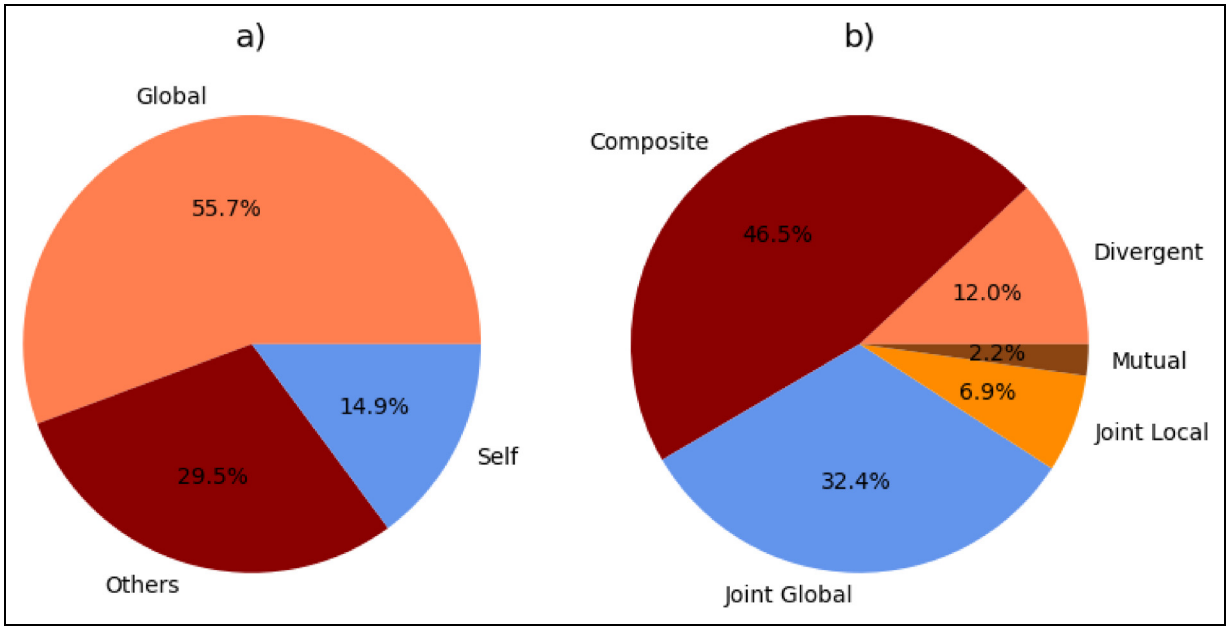

Figure 6. Proportion of time spent in each individual listening mode (a) and in each collective listening mode (b) across all participants.

seen as a moment of *articulation,* a coordination problem faced by musicians in CFI, in which "the consolidation of a sequence loses momentum, or starts to become unstable" (Saint-Germier & Canonne, 2020, p. 459), and musicians need to collectively move to another sequence.

In previous studies on CFI, such as Canonne and Garnier (2015) or Goupil et al. (2020), segmentation analysis was made by asking external listeners to segment the improvisations while listening to them. In the present study, given the sheer volume of music that would need to be segmented and the difficulty of finding enough CFI experts to do that in a reliable way, segmentations were generated automatically by relying on *librosa's agglomerative segment* function, which uses agglomerative clustering based on the spectral content in order to generate segmentations in the music.[2] In order to obtain a meaningful number of

Table 3. Tukey's HSD results with pairwise comparison between, first, individual listening modes, and second, collective listening modes. Only significant results are reported.

Listening modes	Pair	Difference	*SE*	Critical mean	*p*
Individual	Global × Self	25.03	3.834	12.761	<.001***
	Global × Others	16.377	3.857	12.835	.008**
Global	JG × Com.	5.489	0.976	3.775	<.001***
	JG × Div.	9.73	1.543	5.966	<.001***
	JG × Mut.	11.541	2.363	9.135	.005**
	JG × JL	12.068	1.456	5.63	<.001***
	Com. × JL	6.578	1.362	5.268	.006 **

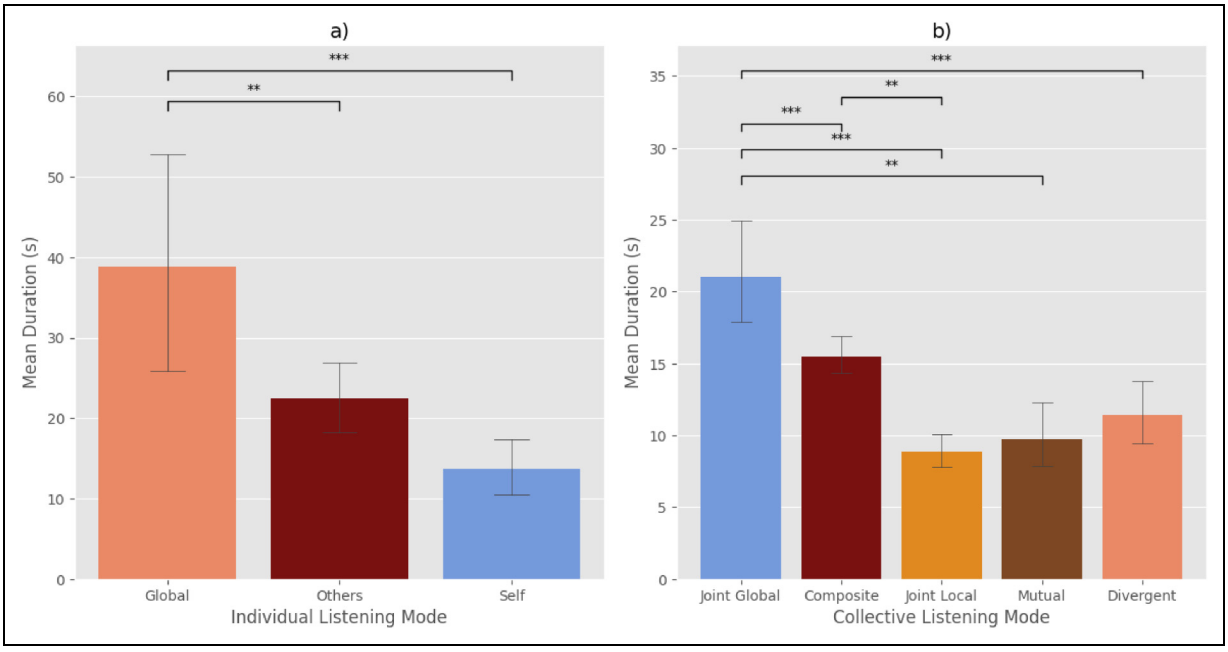

Figure 7. Mean duration of individual (a) and collective listening modes (b). Error bars show standard error (95% interval), and the black asterisks show significant differences (** for $p < 0.01$; *** for $p < 0.001$).

segmentations per improvisation, we took the mean sequence duration from the 32 segmented improvisations found in Canonne and Garnier (2012, 2015) and Faraco (2024). Overall, this mean duration was 58.3 s, with a standard deviation of 29.6 s. Then, in order to establish the number of segments per improvisation (needed as part of the agglomerative segment function), we divided the duration of each improvisation by this average sequence duration. Consequently, each improvisation in our corpus had a varying number of segments, appropriately scaled to its duration. Overall, the segments resulting from our automatic clustering had a mean of 56.91 s, with a large standard deviation 50.43 s.[3]

We then focused on investigating whether the degree of self-listening (i.e., the extent to which a musician tended to listen to themselves or, on the contrary, to others) varied when going through an articulation point—more specifically by comparing moments within an articulation with moments immediately before the articulation and after the articulation. We defined our periods as follows: 1) Articulation periods: These are defined by taking the timestamp of each segmentation point and extending it by 5 s both before and after. 2) Pre-Articulation periods: The 10 s windows before the beginning of an Articulation period. 3) Post-Articulation Period: The 10 s windows after the end of an Articulation period (see Figure 8 for an example).

Our working hypothesis was that the musicians would exhibit an increased attention toward the others during articulation periods, as the success of such articulations seems to depend on a heightened coordination with the other musicians. To test for this hypothesis, we fitted a linear mixed-effects model with "Condition" as a categorical variable, comprising two levels: "Articulation period" and "Pre- and Post-articulation period." This model included track number and participant ID as random intercepts and incorporated an AR(1) correlation structure to address the autocorrelation in the data. As shown in Table 4, the analysis revealed that the degree of self-listening was indeed lower during articulation periods as compared to pre- and post-articulation periods ($p < .001$). This suggests that, in these moments of instability, the musicians needed to pay more attention to other improvisers, either to ensure that their gestures aligned with that of

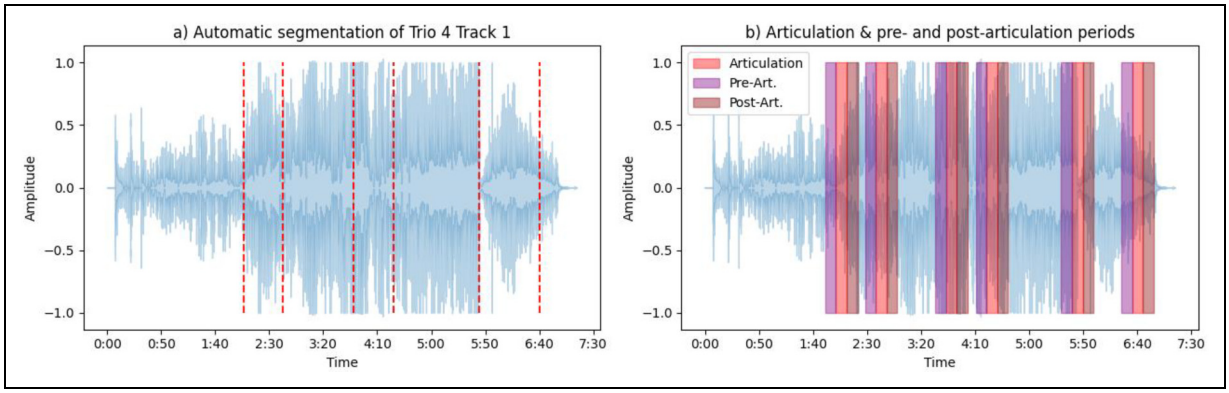

Figure 8. An example of an automatic segmentation (a) and of the delineation of articulation periods, pre-articulation periods, and post-articulation periods (b).

Table 4. Results of the LMM analyses comparing, first, mean distances to the self vertex during articulation periods compared with pre- and post- articulation periods and, second, mean distances between the improvisers in articulation periods compared with pre- and post-articulation periods.

Dependent variable	Factors	Estimate	*SE*	*t*	*P*
Degree of self-listening	Intercept	0.988	0.018	53.01	0
	Articulation	0.068	0.01	6.65	<.001***
Degree of similarity of listening behaviors	Intercept	0.319	0.027	11.5	0
	Articulation	0.063	0.015	4.16	<.001***

the other musicians or to somehow draw inspiration from what the other musicians were doing.

We also fitted a similar linear mixed-effects model to test whether the degree of similarity of listening behaviors within the group was modified during articulations. As shown in Table 4, the musicians' listening behaviors tended to be less similar during articulation periods as compared to pre- and post-articulation periods ($p < .001$). This suggests that, during articulations, musicians might have had different attentional foci. Such dissimilarity could be explained by divergent individual strategies—with some musicians drawing inspiration from, for example, the most salient sonic gesture, while another would draw inspiration from, for example, the most stable sonic proposal. An alternative interpretation would be that such divergence in listening orientations is precisely what triggers the articulation period, because of the momentary discoordination thus created. This would be compatible with previous findings from Goupil et al. (2020), which show that articulations between sequences are not so much the result of a shared intention to change the music as the product of split intentions within the group, between those who want to change the music and those who want to continue with the same idea. This idea—that divergent listening is more likely to elicit a feeling of discoordination among external listeners than other listening modes—will be investigated below in Study 2.

Reciprocal Listening Correlates with Acoustic Coordination. Correlation between audio descriptors has recently been used as a measure of coordination in musical practices. For example, Papiotis et al. (2012) used, among other measurements, Pearson correlations of dynamics and intonation to measure the interdependence of musicians in a string quartet. In the context of CFI, Golvet et al. (2024) demonstrated that a higher correlation of RMS, spectral centroid, and fundamental frequency between two musicians was more likely to be found when at least one musician had the intent to play *with* the other musician (as opposed to play *against* or *without*).

Building on these previous studies, our aim was to investigate whether a higher degree of reciprocal listening between two musicians would be associated with a higher acoustic correlation between the two musicians, for both RMS (loudness) and spectral centroid (timbre). In order to do so, we computed for all possible pairs of musicians multiple Pearson correlation values on 5 s windows for each audio descriptor. We then performed a median split on our data, resulting into two categories for each series of correlation: high correlation vs. low correlation. We fitted an LMM with the degree of reciprocal listening as dependent variable, the RMS correlation and spectral centroid correlation as two categorical predictors (low/high) with interaction, and the duo of musicians as a random intercept. The results show that the degree of reciprocal listening was higher (i.e., lower distance between the two musicians) with RMS (marginally significant, $p = .077$) and that the interaction between the two factors was significant ($\beta = -0.01$, $SE = 0.002$, $t = -4.28$, $p < .001$): when both RMS and spectral centroid correlations were in the "high" category, the degree of reciprocal listening was higher than when both descriptor correlations were in the

Table 5. LMM results for values of degree of reciprocal listening as a function of the coordination level of the two audio descriptors RMS and Spectral Centroid as well as their interaction, with Intercept as the baseline (low RMS / low Spectral Centroid).

Dependent variable	Factors	Estimate	*SE*	*t*	*p*
Degree of reciprocal listening	Intercept	1.021	0.008	118.9	0.000
	RMS	0.003	0.001	1.76	.077
	Spectral Centroid	−0.0003	0.001	−0.19	.85
	RMS × Spectral Centroid	−0.01	0.002	−4.28	<.001***

"low" category (see Table 5). Overall, our analysis thus shows that higher levels of acoustic coordination between two musicians are associated with a higher degree of reciprocal listening between those musicians—suggesting that coordination might in fact be mediated by auditory attention.

Study 2

Study 1 revealed two interesting patterns tying various collective listening behaviors and the coordination dynamics between the musicians. In particular, we showed that a higher degree of reciprocal listening between two musicians was associated with a higher acoustic coordination and that the three musicians exhibited a higher degree of divergent listening during articulation periods. We also showed that Joint Global Listening appeared to be more stable than other listening modes, suggesting that it might be associated with phases during which musicians are well-coordinated with one another and do not feel the urge to change what they are doing. We thus designed a follow-up study based on the musical material collected in Study 1 to assess in a more systematic way whether our perception of musicians' coordination would vary as a function of the collective listening mode they are currently engaged in.

Methods

Participants. A total of 29 participants (15 male, 13 female, 1 other; mean age = 25.55 years, *SD* = 5.06) were recruited. Participants were screened based on their musical practice (a five-year minimum; mean musical practice = 11.75 years, *SD* = 6.45). Participants signed a written consent form for the collection, use, and publication of their data and were compensated at a standard rate for participating in the experiment.

Stimuli and Variables. To select our experimental stimuli, we relied on the four most meaningful collective listening modes described in Section 2.2.2 (i.e., Joint Global Listening, Joint Local Listening, Mutual Listening, and Divergent Listening). Since the listening modes were analyzed in duos of musicians, each trio improvisation (comprising three individual tracks) was artificially divided into three duo improvisations by removing a different individual track each time.

To prepare our stimuli, we first identified all moments from the duo improvisations that exhibited one of the aforementioned listening modes. Every moment lasting less than 10 s was discarded; every moment lasting more than 10 s was randomly cut to a 10 s excerpt. From there, we excluded the excerpts in which one musician was silent for at least 5 s.

Table 6. Results of the LMM with ratings as dependent variable and collective listening modes as fixed effect.

Dependent variable	Factors	Estimate	*SE*	*t*	*p*
Ratings	Intercept	5.63	0.178	31.61	0
	Divergent	−0.773	0.145	−5.3	<.001***
	Joint Local	−0.364	0.145	−2.5	.012 *
	Mutual	−0.481	0.145	−2.29	.001 **

We then looked for the listening mode that had the least number of excerpts: it was Mutual Listening, with 17 excerpts. We thus randomly selected 17 excerpts for each of the remaining listening modes (i.e., Joint Global, Joint Local, and Divergent) from our selection of excerpts. This resulted into a total of 68 excerpts—17 for each listening mode. Finally, a 1 s fade-in and fade-out was inserted in the final mix of each excerpt.

Procedure. Participants listened to each excerpt in random order. After each excerpt, participants had to rate, on a continuous scale, the extent to which they found that the two musicians they were hearing were connected with one another (from "Not at all"—0 to "Very much"—10).

Results

In order to assess the impact of our experimental factor (Collective Listening mode) on participants' ratings, the data were analyzed through a linear mixed-effects model with ratings by participants as the dependent variable, collective listening mode as a fixed effect, participant ID and excerpt number as random intercepts, and Joint Global Listening as the base level. Joint Global Listening was associated with the highest average rating (mean = 5.63, *SD* = 2.36), while Divergent Listening was associated with the lowest average rating (mean = 4.85, *SD* = 2.32). Joint Local Listening was the second-best-rated mode (mean = 5.26, *SD* = 2.47), and Mutual Listening was the third (mean = 5.14, *SD* = 2.51). As shown in Table 6, the results of our LMM revealed that Joint Global Listening's ratings were significantly higher than Divergent

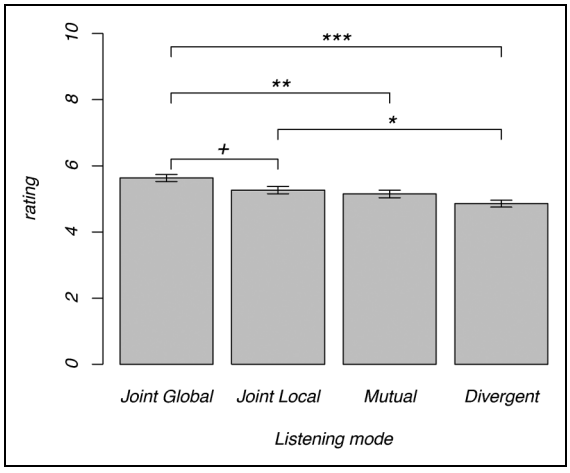

Figure 9. Mean coordination ratings (0: not at all – 10: very much) by independent musician listeners of 10 s musical excerpts as a function of our four main collective listening modes. Error bars show standard error (95% interval), and the black asterisks show significant differences (+ for 0.06; * for $p < 0.05$; ** for $p < 0.01$; *** for $p < 0.001$).

Listening's ratings ($\beta = -0.773$, $SE = 0.145$, $t = -5.30$, $p < .001$), Joint Local Listening's ratings ($\beta = -0.364$, $SE = 0.145$, $t = -2.5$, $p = .012$), and Mutual Listening's ratings ($\beta = -0.481$, $SE = 0.145$, $t = -2.29$, $p = .001$). We also performed a post-hoc pairwise comparison by estimated marginal means, which showed that Joint Global Listening's ratings were higher, although only marginally, when compared to Joint Local Listening ($p = .06$) (see Figure 9 and Table 7). In other words, when both musicians had their auditory attention focused on the overall group sound, they were perceived as more coordinated than when the musicians focused their attention toward a specific sonic source. Conversely, the post-hoc comparison tests showed that Divergent Listening's ratings were significantly lower than Joint Local Listening's ratings ($p = .026$), suggesting that unshared auditory attention came at a cost for the perceived coordination of the musicians.

It should be noted that the effect size remains rather small; however, this limited effect size might also be due to the very short duration of the excerpts—which is likely to have made it harder for the participants to perceive strong differences between the various excerpts. Overall, this study thus confirms that musicians engaged in Joint Global Listening are more likely to be found well-coordinated with one another, and, conversely, that musicians engaged in Divergent Listening are more likely to be found uncoordinated. Interestingly, it also shows that there is no special advantage associated with both Joint Local Listening and Mutual Listening in terms of the perceived coordination between musicians, suggesting that musicians might precisely rely on such listening behaviors as a way to strategically repair some coordination problem. We investigated this idea in a second follow-up study, dedicated to the specific case of Joint Local Listening.

Table 7. Post-hoc pairwise comparison (estimated marginal means) of ratings in collective listening modes.

Pairs	Estimate	*SE*	*t*	*p*
Global × Divergent	0.774	0.146	5.3	<.001***
Global × Local	0.365	0.146	2.49	.06
Global × Mutual	0.482	0.146	3.3	.005**
Divergent × Local	−0.409	0.146	−2.8	.026*
Divergent × Mutual	−0.292	0.146	−2.004	.186
Local × Mutual	0.117	0.146	0.8	.854

Study 3

During Joint Local Listening periods, the musicians attend to a same sonic source (i.e., one of the three musicians of the trio). In these cases, one might wonder whether such attention is due to bottom-up processes (e.g., musicians being attracted by a salient sonic gesture) or to top-down processes (e.g., musicians independently deciding to focus on a same source because they both feel that they need to do so). We thus designed a study to investigate whether external listeners would be similarly drawn to the musician who was the focal point of attention during Joint Local Listening periods, thus suggesting that such a focal point was mainly the result of acoustic saliency (available to an external listener) rather than strategic decisions (unavailable to an external listener).

Methods

Participants and Stimuli. The participants in this study were the same as in Study 2. The order in which the participants underwent the two studies was randomized. Our stimuli also consisted of the same 17 Joint Local Listening excerpts used in Study 2. However, in this case, the excerpts were presented dichotically, with one musician's track entirely on the right side and the other's completely on the left side of the headphones. The pannings were randomized in such a way that the target track (the one the musicians focused on during the Joint Local Listening period) was positioned on the right channel for half of the trials and on the left channel for the other half. This procedure follows that used by Huang and Elhilali (2017) to identify salient events in two sound sequences presented respectively in the right and left ear.

Procedure. The software utilized in this experiment was a standalone Max-MSP patch. Participants were asked to listen to each excerpt and to indicate with the keyboard arrows (left and right) whether, during the 10 s of the excerpt, their auditory attention was overall more focused on the musician heard on the left or on the musician heard on the right. The order of the excerpts was randomized for each participant.

Results

Participants selected the musician who was the actual focus of attention of the improvisers only 45.4% of the time.

To determine if these results differed significantly from chance, we conducted a one-proportion Z-test with a hypothetical proportion of 0.5. The results were significant ($x^2 = 5.15$, $Z = -2.32$, $p = .02$), indicating that the choices made by participants were not merely due to chance. Therefore, it appears that participants were more likely to focus their attention on the musician who was *not* the primary focus of the improvisers during the selected moments of the improvisation. This suggests that Joint Local Listening modes were more likely to emerge as a result of strategic decisions (e.g., because both musicians feel that they need to invest more attention in a given source to repair a coordination problem) rather than mere acoustic saliency.

Discussion

Taken together, our three studies shed new light on listening dynamics in collectively improvised music. First, we showed that real-time saliency manipulations, both through crescendos or through decrescendos, had an effect on the improvisers' auditory attention, by making them more likely to focus more intensely on the musician who had been made more salient. Second, we found that the musicians tended to strategically adapt their listening behavior to the specificities of the interactional context: in particular, we found that the musicians focused more on others when entering a new part or sequence of their performance and that improvisers were more likely to jointly focus on a given sonic source for interactional reasons than for mere acoustic reasons. Third, and finally, we showed that the musicians' collective listening behaviors had an impact on the performance, both at an acoustic level (with a higher degree of reciprocal listening being associated with a higher degree of acoustic coordination) and at a perceptual level (with musicians attending to the overall group sound being more likely to be found coordinated with one another than musicians attending to divergent sonic sources).

Our results provide additional empirical support to the idea that acoustic saliency is not only a matter of specific acoustic features (e.g., rugosity, see Arnal et al., 2019; or shorter inter-onset intervals, see Suied et al., 2010), but also a matter of contextual information (see Kothini & Elhilali, 2023 for a recent discussion)—since local alterations in loudness, both through crescendos and decrescendos, were enough to attract musicians' attention. While this idea has already been tested through highly controlled psychoacoustic tasks, in which participants are presented with short artificial sequences of sounds (see, e.g., Bouvier et al., 2023), our study extends the validity of these results to an interactional context in which participants rely on a variety of instrumental sounds that are on a far greater level of acoustic complexity.

However, another important lesson that can be drawn from our studies is that top-down processes also play an important role in shaping musicians' listening during a performance. Free improvisation is often pictured as a purely emergent artform, in which music is constructed step by step as a result of the ongoing interactions between the performers. But that does not mean that musicians' auditory attention is only driven by bottom-up factors, such as acoustic saliency—with musicians merely passively responding to the sonic tapestry they jointly produce. While such factors obviously play a role—as demonstrated by our first study—there is also room for strategies, goals, and on-the-spot reasoning that can orient musicians' listening behaviors in a particular direction, for example to negotiate a transition from one sequence to another or to find a solution to what appears to be a coordination problem (e.g., finding a way to include a musician who seems unable to find their place in the current musical situation, see Canonne & Garnier, 2012; or finding a way to bring the performance to a satisfying end, see Goupil et al., 2021). Of course, our current methodology did not allow us to explore in details the goals that might underlie the improvisers' choices in the distribution of their auditory attention. Further works on the topic could benefit from integrating retrospective verbalizations (see e.g., Canonne & Garnier, 2012) with the more systematic approach explored here, as a way to investigate how listening strategies might precisely relate to the kind of local, short-term goals that typically emerge in the course of an improvised performance (Saint-Germier & Canonne, 2020).

It is also crucial to note that, despite the few results observed here at the sample level, there are likely to be considerable individual differences in how one listens during a performance (see also Figure 5 above). Numerous factors come in mind: the instrument played (e.g., whether you play an instrument associated with the "rhythm section" or the "frontline"); the musical background (e.g., whether your initial encounter with music was through the so-called "New Complexity"—with its emphasis on complex individual lines—or through drone music—with its emphasis on an overall, always-evolving sonic texture); the level of expertise; etc. Relational factors should not be downplayed either: for example, whether or not two improvisers are well-acquainted could make a difference in how they attend to one another during a performance. The study of individual differences in listening behaviors during musical performances opens the way for exciting new investigations, which could fruitfully combine methods from the sociology of taste (Hennion, 1997) and from experimental psychology.

One might wonder whether our results could extend to non-improvised musical practices. In score-based, well-rehearsed performances, the distribution of musicians' auditory attention might be taken to be mostly guided by the structural information contained within the score (e.g., where the melodic part lies) or the "ideal sound" and other performance goals that are built through rehearsals. It is for example not uncommon for musicians practicing chamber music to indicate on their individual scores who they are supposed to listen to (or to look at) at such and such a point. However, on a more fine-grained scale, such musicians must also continuously adapt to each other,

opening the way for the kind of combination of bottom-up factors—a musician's attention being suddenly drawn by an accent a bit stronger than usual—and top-down processes—a musician deciding to focus on another group member who started the accelerando a bit earlier than anticipated, in order to preserve the fluidity of the overall agogic movement—observed in our studies.

That being said, our study suffers from four important limitations that prevent straightforward generalization. First, our experiment took place in a recording studio, with musicians hearing one another through a highly artificial mix (one musician completely on the right channel and the other on the left channel). Such conditions (in which the various sonic sources are well-segregated) are likely to favor a more analytical form of listening; conversely, the absence of a shared acoustic space might have made merging strategies and the emergence of a group sound more difficult, which in turn might have impacted the overall distribution of individual listening behaviors (for example, by overemphasizing a more encompassing listening behavior to compensate for the highly segregated mix). The present study having provided a first validation of the post-hoc annotation methodology as a valid strategy to investigate auditory attention during performance, further experiments could now rely on more ecological settings, by allowing performers to play in the same room and/or to see each other. Second, the absence of an audience might also have impacted how musicians listen to each other. In particular, this might have favored a more conversational approach (with musicians focusing alternatively on one another), with less attention to the overall result of their interaction (and thus less attention to the group sound as a whole). This could be controlled for by relying on decoy audiences in follow-up studies on the same topic. Third, it remains possible that our annotation interface biased participants toward indicating a global listening behavior (as it corresponded to the center of the interface); further experiments could explore alternative design choices (e.g., using a discrete interface rather than a continuous one) to mitigate such potential biases. Fourth, we should also account for the limitations of working only with improvisation trios. Group size can affect coordination and the unfolding of joint actions (Dyer et al., 2009), such as conversational behavior (Fay et al., 2000) and the way that musicians improvise (Goupil et al., 2020; Saint-Germier et al., 2021). As such, it is also likely that musicians' auditory attention would be affected by group size, for example because of the tendency of improvisers to create sub-groups in larger ensembles (Goupil et al., 2020). There would perhaps be a smaller tendency toward global listening in larger groups, with sub-groups having their own listening dynamics. Further studies should account for this possibility by contrasting groups of various sizes (for example trios and sextets).

Finally, and on a more general note, the relationship we observed between auditory attention and coordination also deserves further discussion. The correlational nature of the analyses conducted here do not allow us to conclude that musicians appear to be more coordinated *because* they are paying more attention to one another or *because* they are paying joint attention to the overall group sound; indeed, it might just be that musicians listen more to one another because they rely on similar acoustic material (and are spontaneously attracted to the musician that sounds more like themselves) or that they can endorse a more encompassing listening behavior precisely because everything is going along smoothly, with no obvious misunderstanding between the musicians. A possible way to gain further insight on this issue would be to introduce artificial points of uncoordination (for example by replacing in real time a given musician's actual signal by a recorded sample of the same musician) to investigate whether such uncoordination phases would cause an alteration in the distribution of the musicians' auditory attention. Another possible way to further investigate the relationship between coordination and auditory attention would be to consider the latter as a mediator variable in future studies, together with other elements that are believed to enhance coordination in CFI (such as shared local goals). But at the very least, our results show that coordination and auditory attention are closely associated, which is compatible with the idea that attentional processes underlie coordination in the context of freely improvised musical interactions.

From that same perspective, our results are also compatible with the idea that joint attention might facilitate coordination (Majeau-Bettez et al., 2023) and the emergence of a feeling of togetherness (Bishop, 2023), at least as perceived by external listeners. But an interesting aspect of our results is that joint attention to the overall group (or to the overall result created by the musicians' individual actions) might be more efficient than more local forms of joint attention (i.e., paying attention to a same sonic source). This paves the way for exciting new investigations: Is the group's emergent music treated as an abstract agent of its own, which can be an object of auditory attention in the same way as a given member of the group, as suggested by some of the musicians interviewed in Canonne (2018)? Should we treat global and local joint listenings as distinct attentional processes? While musical interactions have not often been used as an experimental paradigm for the study of the effects of joint attention on group behavior, they in fact appear as a highly promising setting for the exploration of new theoretical questions.

Acknowledgments

This work was supported by the INSEAD-Sorbonne Université Behavioural Lab and by the Collegium Musicae. This work was also supported by a FAPESP grant (2022/05792-1) to A.F.

Action Editor

Andrew Goldman, Indiana University, Jacobs School of Music, Department of Music Theory.

Peer Review

David Borgo, University of California San Diego, Department of Music.

Sarah Faber, Simon Fraser University, Faculty of Applied Sciences, Institute for Neuroscience and Neurotechnology.

Authors Contributions

All authors designed Study 1. C.C. and A.F. designed Studies 2 and 3. C.C., A.F., A.S., and C.V. collected the data for Study 1. C.C. and A.F. collected the data for Studies 2 and 3. C.C. prepared the analysis plan. A.F. analyzed the data. C.C. and A.F. interpreted the results. C.C. wrote the article with contributions from A.F. and comments from the other authors.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

The studies reported in this article were approved by the INSEAD's IRB (Protocol ID: 2023-16).

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Collegium Musicae.

ORCID iDs

Arthur Faraco https://orcid.org/0000-0002-0641-9625
Clément Canonne https://orcid.org/0000-0003-2617-6971

Data Availability Statement

The datasets generated and analyzed during the current study are available in Faraco et al. (2024).

Notes

1. Due to a technical issue, both of the first trio's improvisations had to be performed without any kind of audio manipulation.
2 Agglomerative (or hierarchical) clustering is a bottom-up approach to cluster analysis where each data point initially forms its own cluster, and these are progressively merged based on similarity until a certain predetermined number of clusters is reached (Ackermann et al., 2012). In the agglomerative segmentation function of *Librosa*, this clustering method is applied to a similarity matrix derived from feature extraction. In our study, segmentation was conducted based on variations in the Mel-spectrogram, which captures the power spectrum of sound over time. The decision to use the Mel-spectrogram was influenced by its ability to represent changes in timbre, rhythm, pitch, and texture. This provides a broader scope for analyzing the audio, as the Mel-spectrogram effectively enfolds significant acoustic variations. This function was applied to the final mix of each improvisation.
3 We conducted a Silhouette test for each improvisation to determine the optimal number of clusters, which identified the minimum number of clusters possible (two) for nearly all improvisations, except for four that aligned with our segmentation definition. This outcome, possibly influenced by the Mel-spectrogram data's complexity, results from the Silhouette test comparing each value within and across clusters, yielding a score from −1 to 1. The diverse Mel-spectrogram values in CFI likely lead to lower scores, suggesting fewer clusters. Our approach, we argue, captures more accurately the segmental structure observed in CFI, as supported by existing literature.

References

Ackermann, M. R., Blömer, J., Kuntze, D., & Sohler, C. (2012). Analysis of agglomerative clustering. *Algorithmica*, *69*(1), 184–215. https://doi.org/10.1007/s00453-012-9717-4

Arnal, L. H., Kleinschmidt, A., Spinelli, L., Giraud, A., & Mégevand, P. (2019). The rough sound of salience enhances aversion through neural synchronisation. *Nature Communications*, *10*(3671). https://doi.org/10.1038/s41467-019-11626-7

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

Bertolani, V. (2019). Improvisatory exercises as analytical tool: The group dynamics of the Gruppo di Improvvisazione Nuova Consonanza. *Music Theory Online*, *25*(1). https://doi.org/10.30535/mto.25.1.1

Besser, M., Roberts, I., & Walsh, M. (2013). *The upright citizens brigade comedy improvisation manual*. Comedy Council of Nicea.

Bishop, L. (2023). Focus of attention affects togetherness experiences and body interactivity in piano duos. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. https://doi.org/10.1037/aca0000555

Bishop, L., Cancino-Chacón, C., & Goebl, W. (2022). Beyond synchronization: Body gestures and gaze direction in duo performance. In R. Timmers, F. Bailes, & H. Daffern (Eds.), *Together in music: Coordination, expression, participation* (pp. 182–187). Oxford University Press.

Borgo, D. (2005). *Sync or swarm: Improvising music in a complex age*. Bloomsbury Academic.

Bouvier, B., Susini, P., Marquis-Favre, C., & Misdaris, N. (2023). Revealing the stimulus-driven component of attention through modulations of auditory salience by timbre attributes. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-33496-2

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3). https://doi.org/10.1037/met0000159

Burrows, J., & Reed, C. G. (2016). Free improvisation as a path-dependent process. In G. Lewis & B. Piekut (Eds.), *The Oxford handbook of critical improvisation studies* (Vol. 1, pp. 396–415). Oxford University Press.

Canonne, C. (2018). Rehearsing free improvisation? An ethnographic study of free improvisers at work. *Music Theory Online*, *24*(4). https://doi.org/10.30535/mto.24.4.1

Canonne, C., & Garnier, N. (2012). Cognition and segmentation in collective free improvisation: An exploratory study. In Proceedings of the 12th international conference on music perception and cognition 8th triennial conference of the European society for the cognitive sciences of music, Thessaloniki, Greece.

Canonne, C., & Garnier, N. (2015). Individual decisions and perceived form in collective free improvisation. *Journal of New Music Research*, *44*. https://doi.org/10.1080/09298215.2015.1061564

Clarke, E. (2005). *Ways of listening: An ecological approach to the perception of music meaning*. Oxford University Press.

Corbett, J. (2016). *A listener's guide to free improvisation*. University of Chicago Press.

Dalton, P., & Lavie, N. (2004). Auditory attentional capture: Effects of singleton distractor sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 180. https://doi.org/10.1037/0096-1523.30.1.180

Dyer, J. R., Johansson, A., Helbing, D., Couzin, I. D., & Krause, J. (2009). Leadership, consensus decision making and collective behaviour in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 781–789. https://doi.org/10.1098/rstb.2008.0233

Faraco, A. (2024). Perception of structure in collective free improvisations and its context dependency: An exploratory analysis. *Empirical Musicology Review*, *18*(1), 63–81. https://doi.org/10.18061/emr.v18i1.8875

Faraco, A., Schwarz, A., Vincent, C., Susini, P., Ponsot, E., & Canonne, C. (2024). *Listening behaviors and musical coordination in collective free improvisation [Dataset]*. figshare. https://doi.org/10.6084/m9.figshare.25146296.v1

Fay, N., Garrod, S., & Carletta, J. (2000). Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science*, *11*(6), 481–486. https://doi.org/10.1111/1467-9280.00292

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Golvet, A., Goupil, L., Saint-Germier, P., Matuszewski, B., Assayag, G., Nika, J., & Canonne, C. (2024). With, against, or without? Familiarity and copresence increase interactional dissensus and relational plasticity in freely improvising duos. *Psychology of Aesthetics, Creativity, and the Arts*, *18*(2), 182–195. https://doi.org/10.1037/aca0000422

Goupil, L., Saint-Germier, P., Rouvier, G., Schwarz, D., & Canonne, C. (2020). Musical coordination in a large group without plans nor leaders. *Scientific Reports*, *10*. https://doi.org/10.1038/s41598-020-77263-z

Goupil, L., Wolf, T., Saint-Germier, P., Aucouturier, J., & Canonne, C. (2021). Emergent shared intentions support coordination during collective musical improvisations. *Cognitive Science*, *45*(1). https://doi.org/10.1111/cogs.12932

Hennion, A. (1997). Baroque and rock: Music, mediators and musical taste. *Poetics*, *24*(6), 415–435. https://doi.org/10.1016/S0304-422X(97)00005-3

Huang, N., & Elhilali, M. (2017). Auditory salience using natural soundscapes. *The Journal of the Acoustical Society of America*, *141*(3), 2163–2176. https://doi.org/10.1121/1.4979055

Johnson, D. (2012). The art of listening: intuition & improvisation in choreography.

Kaya, E. M., Huang, N., & Elhilali, M. (2020). Pitch, timbre, and intensity interdependently modulate neural responses to salient sounds. *Neuroscience*, *440*, 1–14. https://doi.org/10.1016/j.neuroscience.2020.05.018

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions* (pp. 205–221). IOS Press.

Kothini, S. R., & Elhilali, M. (2023). Are acoustics enough? Semantic effects on auditory salience in natural scenes. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1276237

Majeau-Bettez, E., Golvet, A., & Canonne, C. (2023). Tracking auditory attention in group performances: A case study on Éliane Radigue's Occam Delta XV. *Musicae Scientiae*, 10298649231203641. https://doi.org/10.1177/10298649231203641

Matuszewski, B. (2019). Soundworks—A framework for networked music systems on the web – state of affairs and new developments. In Proceedings of the web audio conference (WAC) 2019. https://hal.archives-ouvertes.fr/hal-02387783/document

McFee, B., Raffel, D., Dawen, L., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference (pp. 18–25). https://doi.org/10.25080/Majora-7b98e3ed-003

Monson, I. (1996). *Saying something: Jazz improvisation and interaction*. University of Chicago Press.

Moran, N., Hadley, L. V., Bader, M., & Keller, P. E. (2015). Perception of 'back-channeling' nonverbal feedback in musical duo improvisation. *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0130070

Müller, V., & Lindenberger, U. (2019). Dynamic orchestration of brains and instruments during free guitar improvisation. *Frontiers in Integrative Neuroscience*, *13*(50). https://doi.org/10.3389/fnint.2019.00050

Müller, V., Sänger, J., & Lindenberger, U. (2013). Intra- and inter-brain synchronization during musical improvisations on the guitar. *PLoS ONE*, *8*(9). https://doi.org/10.1371/journal.pone.0073852

Papiotis, P., Marchini, M., & Maestre, E. (2012). Computational analysis of solo versus ensemble performance in string quartets: intonation and dynamics. In Proceedings of the 12th international conference on music perception and cognition and the 8th triennial conference of the European society for the cognitive sciences of music.

Pelz-Shermann, M. (1998). *A framework for the analysis of performer interactions in western improvised contemporary art music* [Unpublished Doctoral Dissertation]. University of California.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer. https://doi.org/10.1007/b98882.

Pressing, J. (1984). Cognitive processes in improvisation. In W. R. Crozier & A. J. Chapman (Eds.), *Cognitive processes in the perception of art* (pp. 345–363). North-Holland.

Saint-Germier, P., & Canonne, C. (2020). Coordinating free improvisation: An integrative framework for the study of collective improvisation. *Musicae Scientiae*, *26*(3), 455–475. https://doi.org/10.1177/1029864920976182

Saint-Germier, P., Goupil, L., Rouvier, G., Schwarz, D., & Canonne, C. (2021). What it is like to improvise together? Investigating the phenomenology of joint action through improvised musical performance. *Phenomenology and the Cognitive Sciences*. Advance online publication. https://doi.org/10.1007/s11097-021-09789-0

Savouret, A. (2010). *Introduction à un solfège de l'audible : l'improvisation comme outil pratique*. Symétrie.

Seddon, F. A. (2005). Modes of communication during jazz improvisation. *British Journal of Music Education*, *22*(1), 47–61. https://doi.org/10.1017/S0265051704005984

Straetmans, L., Holtze, B., Debener, S., Jaeger, M., & Mirkovic, B. (2022). Neural tracking to go: Auditory attention decoding and saliency detection with mobile EEG. *Journal of Neural Engineering*, *18*. https://doi.org/10.18112/1741-2552/ac42b5

Suied, C., Susini, P., McAdams, S., & Patterson, R.D. (2010). Why are natural sounds detected faster than pips?. *The Journal of the Acoustical Society of America*, *127*(3), EL105–EL110. https://doi.org/10.1121/1.3310196

Van Der Steem, M. C., & Keller, P. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Science*, *7*. https://doi.org/10.3389/fnhum.2013.00253

Wöllner, C., & Keller, P. E. (2017). Music with others: Ensembles, conductors and interpersonal coordination. In R. Ashley & R. Timmers (Eds.), *Routledge companion to music cognition* (pp. 313–324). Routledge.