Trade-offs in Coordination Strategies for Duet Jazz Performances Subject to Network Delay and Jitter

Huw Cheston, Ian Cross, &
Peter M. C. Harrison
University of Cambridge, Cambridge, United Kingdom

Coordination between participants is a necessary foundation for successful human interaction. This is especially true in group musical performances, where action must often be temporally coordinated between the members of an ensemble for their performance to be effective. Networked mediation can disrupt this coordination process by introducing a delay between when a musical sound is produced and when it is received. This can result in significant deteriorations in synchrony and stability between performers. Here we show that five duos of professional jazz musicians adopt diverse strategies when confronted by the difficulties of coordinating performances over a network—difficulties that are not exclusive to networked performance but are also present in other situations (such as when coordinating performances over large physical spaces). What appear to be two alternatives involve: 1) one musician being led by the other, tracking the timings of the leader's performance; or 2) both musicians accommodating to each other, mutually adapting their timing. During networked performance, these two strategies favor different sides of the trade-off between, respectively, tempo synchrony and stability; in the absence of delay, both achieve similar outcomes. Our research highlights how remoteness presents new complexities and challenges to successful interaction.

Received: October 6, 2023, accepted February 16, 2024.

Key words: joint action, synchronization, rhythm and timing, improvisation, coordination

Jazz ensemble performances involve intricate processes of coordination among the musicians involved. Through collective improvisation, performers exchange musical ideas, negotiate their roles in the ensemble, and strive to understand each other's intentions in the performance. In particular, jazz musicians display unique (and genre-specific) approaches to embodied musical timing, variously referred to as their "feel," "swing," "pocket," "time" (Berliner, 1994), and they must align their timing with their partners' in order to maintain a coherent performance (Doffman, 2014). Group jazz improvisation thereby exemplifies a particular kind of joint action in which coordination is achieved through multi-level interactive alignment, the process by which the participants in a joint activity come to understand aspects of their shared world in the same way as each other (Garrod & Pickering, 2004). In this respect, jazz improvisation has often been compared to spontaneous spoken conversation (Kello et al., 2017; Monson, 1996), where alignment towards shared conceptualizations of space and time can also occur between participants.

With the rapid advancements in high-quality audio streaming technology, interactive musical performances of genres including jazz can now take place over vast distances via the internet. Networked performance brings with it certain advantages; for example, improving access to live performances for those with disabilities and reducing travel-related costs, time, and environmental emissions. More recently, it became necessary for many musicians because of restrictions on face-to-face performances during the COVID-19 pandemic. However, using a network to mediate a musical performance has its limitations.

Whenever a sound is transferred over a network, a temporal delay—known as *latency*—is introduced between when it was first produced and when it is received, resulting from the time required to convert sound waves to digital packets and transmit them over a network. While some form of delay is present even when music is performed face-to-face, owing to the time taken for sound to transmit in air, the latency present in networked performances is often several orders of magnitude greater (Chafe et al., 2010). In addition, transmission errors and network congestion can cause fluctuation in packet arrival time, and this variability—known as *jitter*—can cause further instability in (or even the momentary loss of) the output signal during networked performances.

This variable delay poses significant challenges for successfully coordinating joint action in various forms

Music Perception, volume 42, issue 1, pp. 48–72, issn 0730-7829, electronic issn 1533-8312. © 2024 by the regents of the university of california all rights reserved. please direct all requests for permission to photocopy or reproduce article content through the university of california press's Reprints and Permissions web page, https://online.ucpress.edu/journals/pages/reprintspermissions. DOI: https://doi.org/10.1525/mp.2024.42.1.48

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

of time-dependent communication such as spoken conversation (Aagaard, 2022; Boland et al., 2022) and interactive music-making (Chafe et al., 2010; Chew et al., 2005; Rottondi et al., 2015). Network latency impedes the alignment of temporal models and precludes the extreme rhythmic synchrony that is typical of ensemble playing; networked performances have correspondingly been associated with decreased ratings of performance quality and reduced connectedness with co-performers compared to face-to-face music-making (Bartlette et al., 2006; Monache et al., 2019; Olmos et al., 2009). Moreover, unlike speech interaction, in musical performance action and interaction are generally organized around a continuous temporally periodic pulse (London, 2012), which is highly susceptible to disruption by latency and jitter. Network latency starts to cause problems for interactive musical performances at 28 milliseconds of one-way delay (Chafe et al., 2010), while the threshold for negotiating spoken conversations online without difficulty seems to be much higher, at 500 ms (Holub et al., 2007).

Prior research has focused on optimizing networking infrastructure to improve the fluidity and coherence of remote musical performance. This has included the development of low-latency, low-jitter platforms such as JackTrip (Cáceres & Chafe, 2010) and LOLA (Drioli et al., 2013) that intend to offer the networked ensemble an experience as close as possible to that of playing in the same room as each other. Strategies used by these systems include data buffering, where incoming data packets are stored by the application before being released at regular intervals to ensure consistency in the output signal. But while these technological advances can minimize the presence of a time-lag to performers, latency can never be eliminated completely. A networked ensemble must, therefore, find some way of coordinating their joint action that accommodates this delay in the process.

In this study, we aim to determine the relative optimality of strategies for coordinating group musical improvisation over a network, with all the reliance on periodicity and the tight temporal coordination of action that this process entails. In previous work, coordination in face-to-face performances has typically been modeled as a function of one performer's adaptation to any deviations from expected isochrony in another's playing (phase correction or "coupling": see Jacoby et al., 2021, Timmers et al., 2014, Wing et al., 2014; see also Demos & Palmer, 2023, for a recent review). In networked performances, one hypothesis is that all players in an ensemble should try hard to listen to each other and couple together, which could result in mutual and symmetrical adaptation to timing variability (Nowicki et al., 2013). An alternative hypothesis is that successful networked interaction requires a degree of asymmetry in the distribution of roles within a musical ensemble: one performer ignores the delayed feedback, while their partner(s) attempt to match and adapt to them (Carôt & Werner, 2009). Participants in prior studies have mentioned adopting either these or similar strategies to accommodate latency during networked performances (e.g., Bartlette et al., 2006), but the question of which might be optimal has not been systematically explored.

To address this issue, we conducted a series of experiments where five duos of professional jazz drummers and pianists improvised together over a simulated network. This network introduced delay and jitter of a magnitude up to and including that present on Zoom, a telecommunications platform commonly used in remote interaction and teleconferencing. To evaluate the relative optimality of their coordination strategies, we use a battery of objective and subjective metrics as indicators of the overall success of these performances. Data are gathered from MIDI recordings and evaluations are provided by both the musicians and a secondary sample of naive listeners recruited via an online perceptual study. We begin our analysis by outlining the results obtained from co-present (i.e., non-delayed) conditions, before then considering how the presence of latency and jitter affects these baselines. Next, we use a combination of linear causal modeling and participant self-reports to characterize the individual coordination strategies of each duo. Finally, we evaluate the relative optimality of these strategies by conducting computer simulations.

Method

PARTICIPANTS

Ten adult men with a median age of 24 (*SD* = 5, range = 21 to 36) participated in the study. All participants had at least three years of professional experience on either piano or drum kit and held an undergraduate (or higher) degree in music performance and were required to be fluent English language speakers. This combination of instruments was selected as they constitute part of the "rhythm section" in jazz: these musicians play continuously throughout a performance, and their interaction is considered vital to its overall success (see Supplementary Materials section §2.1. for further detail about the rhythm section, at online.ucpress.edu/mp). Participants were recruited from within the first author's network of performance contacts and were all

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

professional or semi-professional musicians based in London, UK. All individuals who were approached for inclusion participated in and completed the study, with the experiments being conducted during April–July 2022.

Participants were grouped into five duos, each consisting of one pianist and one drummer, with no participant performing in more than one duo. The two musicians in duo 3 had never played together before, while the remaining pairs reported performing with each other during the past year. However, none of the participants had any prior experience of networked performance with their duo partner before the present experiment.

This sample size was deemed appropriate because we treated each participant (and participant-duo) as an independent entity, extensively characterizing their individual coordination strategy over a large number of performances rather than focusing on group averages across the entire sample. This approach is typical for psychological studies of advanced musical performance where ensemble roles and interpretative strategies are both highly specialized and individualized, and may depend on pre-existing relationships within a particular participant-group (e.g., Jacoby et al., 2021; Pras et al., 2017; Timmers et al., 2014; Wing et al., 2014).

The experiment was approved by the Ethics Review Subcommittee at the Faculty of Music, University of Cambridge, UK, and all participants provided written informed consent. Participants were paid for their time and travel expenses.

TESTBED CONFIGURATION

We designed a novel testbed for the experiment, shown in Figure 1. First, we generated a representative measurement of network latency and jitter by connecting two computers on the same network to a virtual call on a popular telecommunications platform (Zoom). Then, we positioned a metronome (playing at 80 quarter note beats per minute) next to one computer, so that an echo

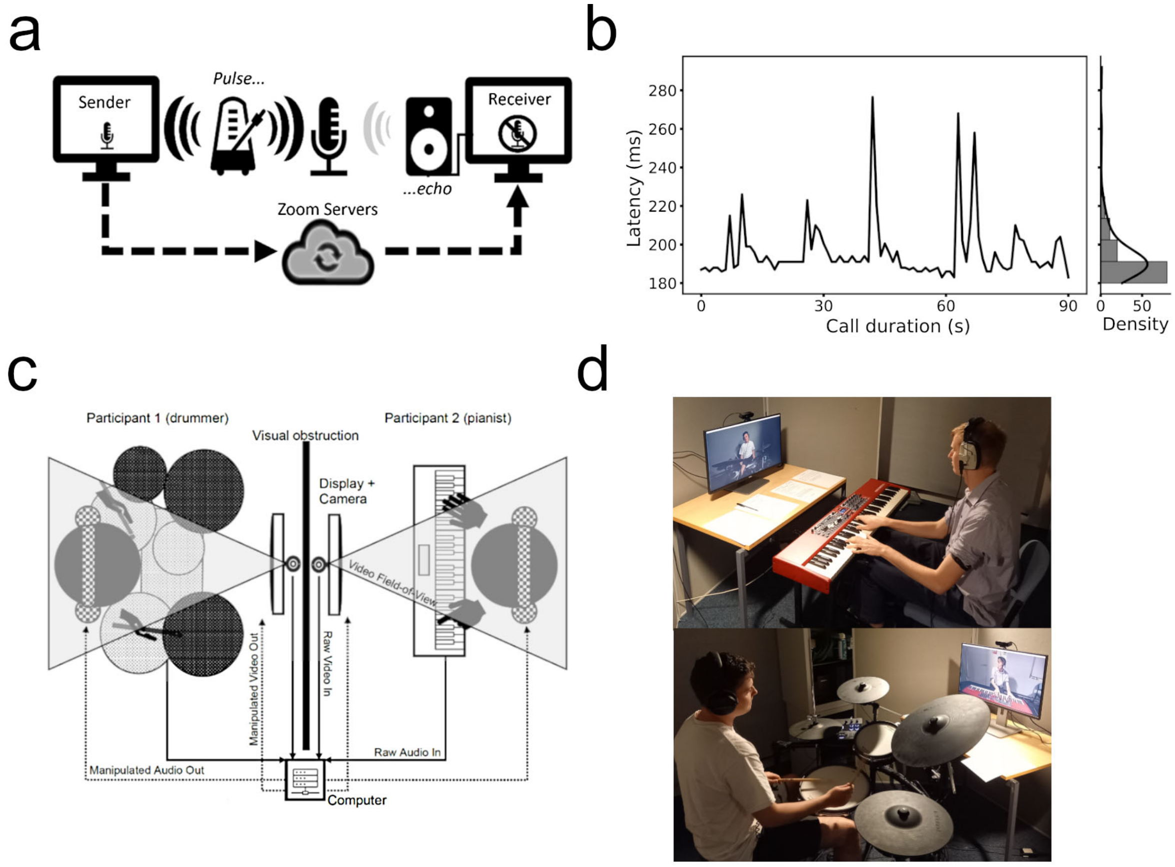

FIGURE 1. Testbed configuration. The top row of panels shows (a) the procedure used to measure network latency over Zoom and (b) the resulting 90-second latency and jitter time-series profile. The bottom row shows (c) a diagram of the testbed layout and (d) how this looked within the experiment room, for pianist and drummer.

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

could be heard from the speakers of the other computer as each pulse was transmitted (Figure 1a). We recorded audio of this process for 90 seconds and derived the variable roundtrip latency between pulse and echo by applying the onset detection algorithm contained in the librosa library, version 0.8.1 (McFee et al., 2015), to the recording. Network latency was generally stable around a median peak-to-peak delay of 192 ms (*SD* = 17.7, range = 181 to 293 ms), with occasional spikes caused by jitter (Figure 1b). We confirmed that identical tests conducted in three different locations produced similar results (see Supplementary Materials Figure S1), and that our measurements resembled the typical experience of using this platform for telecommunication.

The testbed was designed to apply these measurements to a performance in a controlled fashion. In the experiment room, an acoustically isolated recording studio, both participants sat apart from each other, with barriers placed so as to prevent direct visual contact (Figure 1c). Separate video and MIDI streams were captured of their performances, using an electronic keyboard (Nord Electro 6D-73), drum kit (Roland TD-27KV), and computer webcams (Logitech Brio 4K Pro). Video was captured at a resolution of 1920 x 1080 px and a rate of 30 frames per second. These signals were then transmitted to a computer server via USB connection and a 32-channel MIDI interface (M-Audio MIDISPORT 2 x 2).

Variable latency was applied to each musical track using the digital audio workstation REAPER. Each time a note was played, the corresponding latency was determined by looking up the current value in the latency time series, and the playback of that note was then delayed by that amount. Delay times were resampled from the latency time series at periodic intervals of 750 ms to replicate the metronome speed of the original test. Latency was applied to each video track in an identical manner using the computer vision library OpenCV.

The server then stored both the incoming (live) and outgoing (delayed) signals for later analysis before presenting them to participants. MIDI signals were first transcoded to audio using a high-quality virtual instrument library and then routed over closed-back headphones at a rate of 44.1 kHz through a 16-bit digital-to-analog converter. Video was shown on individual 27-inch computer monitors in front of each participant. Participants heard and saw delayed audio and video from their partner's performance, alongside unmanipulated audio of their own playing. They did not see video of themselves, due to the additional computational demand this would introduce and the likelihood that this would obstruct the view of their partner and their instrument (Figure 1d).

During testing, we found that the inherent delay added by our testbed signal path to the incoming MIDI signal was 4 ms for the keyboard and 3 ms for the drums. This is significantly lower than the inherent audio latency reported in previous studies (e.g., Olmos et al., 2009; Rottondi et al., 2015). This value is not included in the reported results as it is likely perceptually sub-threshold (Grant et al., 2004).

The inherent delay for the incoming video signals was substantially greater at 33 ms, which was not unexpected due to the greater computational demands involved in the real-time processing of video compared to audio. Rather than applying additional latency to the MIDI to compensate for this, we instead allowed both audio and video to be unsynchronized, which is common during networked performance and teleconferencing (Rottondi et al., 2016).

EXPERIMENTAL DESIGN

We selected a "twelve-bar blues" structure in the key of B♭ as the musical stimulus for the experiment. This simple, repetitive form is pervasive in jazz and popular music and was immediately familiar to all participants, all of whom chose to perform it from memory and without the aid of a musical score that was offered to them (Supplementary Materials Figure S2a). Performances lasted 90 seconds, the same duration as our measured latency time series, meaning that the total number of repeats of the blues form was dictated by the tempo established by the musicians. Before the experiment began, participants completed three warm-up performances without any latency and at a variety of tempi, allowing them to practice together in the testbed environment. The remaining performances comprised the experimental session and were characterized by differing amounts of latency and jitter.

We operationalized latency by transposing the original latency time series (Figure 1b) so that its minimum value became either 23, 45, 90, or 180 ms, not including the inherent delay introduced by the testbed. This latter value (180 ms) was within the upper limit of latency times tested in prior research and was essentially equivalent to the minimum delay we had originally measured on Zoom (181 ms); the other values tested were the integer quotients resulting from the successive division of 180. We manipulated jitter by keeping the minimum latency value as set above but scaling the deviations from this minimum value by either 1.0x (no change from original variation measured on Zoom), 0.5x, or 0.0x (no jitter, consistent/"flat" delay). The procedure

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

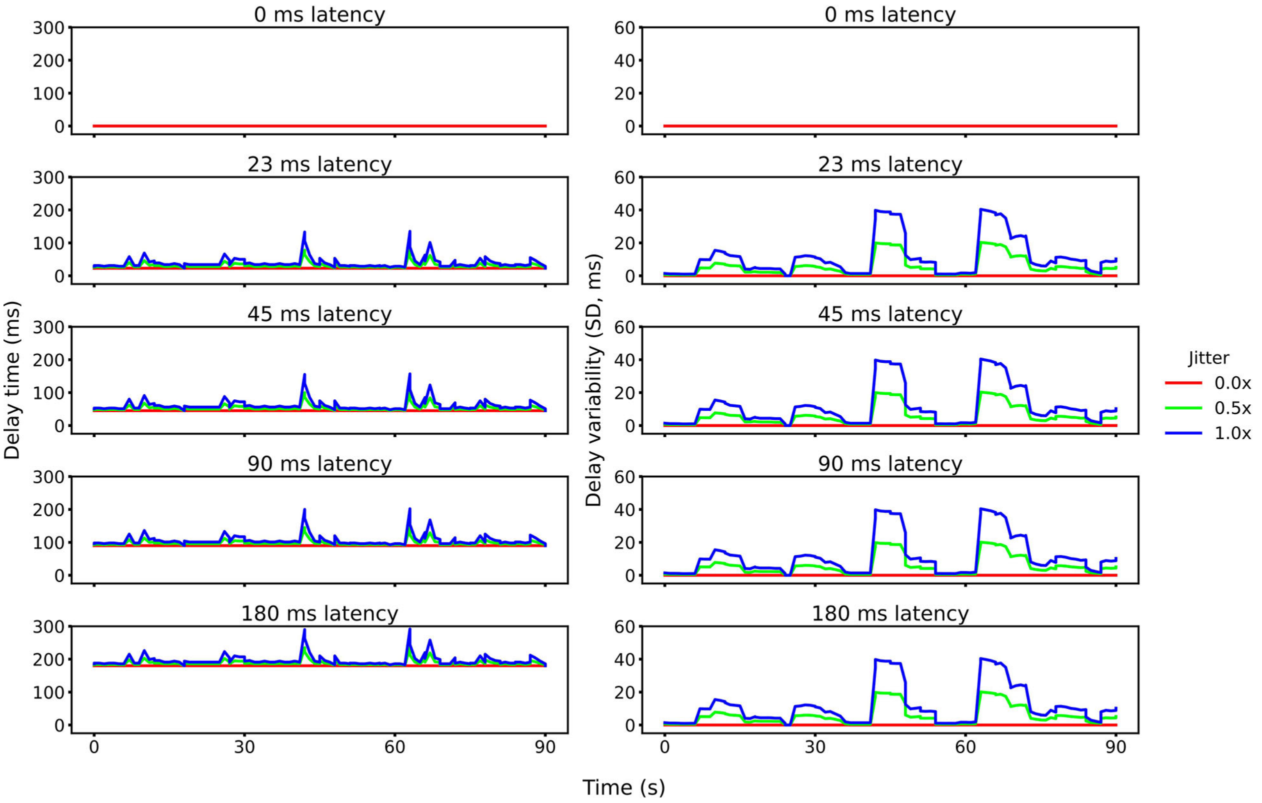

FIGURE 2. Experimental conditions. The thirteen conditions tested in the experiment were derived from the transformation of the measurements in Figure 1b. The rows indicate minimum latency values between 0 and 180 ms, with jitter scalings shown by the color of each line. The left column shows the raw latency timings over the 90-second duration of each performance. The right column displays the standard deviation of latency values obtained over a sliding window of four seconds, which we later use in the partial correlation analysis of timing variance shown in Supplementary Materials Figure S9.

used to transform the original latency time series d to the latency time series d', with minimum latency time L and jitter scaling J, can be written as:

$$d'^{N}_{1} = \left(J \times \left(d_i - min(d^{N}_{1})\right) + L\right)^{N}_{1} \qquad (1)$$

A final control condition was also tested, in which no latency or jitter was applied to the signal beyond that introduced by the testbed (i.e., $L = 0$, $J = 0$). These thirteen conditions were presented to duos in a randomized order across two successive sessions with a break of one hour in between, with each condition appearing once in every session. All experimental manipulations were delivered by the lead author. Figure 2 shows the transformations of the latency time series that were used in each condition.

EXPERIMENTAL PROCEDURE

Before each performance began, participants heard sixteen quarter-note pulses from a synthesized metronome over their headphones at a moderate tempo of 120 quarter-note beats per minute, which is a typical speed for a medium tempo jazz blues. The first of each group of four pulses was played at a higher pitch and greater dynamic level than the others to clarify the meter as four quarter-note beats per measure. The total duration of this count-in was eight seconds.

Participants were then instructed to improvise together over the blues form while maintaining both the tempo and meter established by the metronome pulses. To do so, they were asked to play continuous quarter notes, either in their left hand as part of a "walking bass" or "stride" accompaniment (pianist), or by "playing time" on their hi-hat and ride cymbals and bass kick drum in a swing style (drummer). In Supplementary Materials Figure S2b we provide notated musical examples of the typical patterns improvised by our participants, and in Figure S2c we analyze the frequency of MIDI note distribution across the total range of each instrument to demonstrate that their performances conformed to the given brief.

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

Participants were not made explicitly aware of the presence of latency and jitter. Instead, they were told that the feedback they would receive from their partner would change during the experiment, and that they should interact with them as they otherwise would during a "real" performance. To that end, we encouraged participants to improvise light musical embellishments while following their assigned brief, provided that this did not disrupt their ability to maintain the underlying quarter-note pulse. All instructions given to participants were in English.

Each performance lasted for 90 seconds before participants were instructed to stop. Recording was then maintained for several seconds to prevent notes from being cut-off if they occurred shortly before the 90-second point; the following analyses only consider the first 90 seconds of a performance, however.

After each performance, participants responded to questionnaires indicating their subjective experience of that performance in terms of: 1) the quality of the interaction with their duo partner; 2) the ease of coordination with their partner and; 3) the overall success of the performance (Supplementary Materials Figure S3a). These questions were administered using an online survey platform and are similar to those used in Setzler and Goldstone's (2020) earlier study of improvised duo interaction. Participants provided responses to each question using integers from 1 to 9 inclusive, with lower scores corresponding to negative and higher scores to positive evaluations. See Figure S3b for measurements of test-retest reliability for these questions. Participants were also able to comment in free text on a performance, with anonymized transcripts of these responses provided in Supplementary Materials section §3.

LISTENER EVALUATIONS

As we could not assume that our participants would provide an unbiased assessment of the quality of their own performance (Pras et al., 2017; Schober & Spiro, 2014), we also obtained equivalent evaluations from listeners recruited online via the Prolific platform. This experiment was implemented in PsyNet, version 10.2.0 (Harrison et al., 2020), a software package that enables large-scale perceptual studies to be conducted online through a browser-based interface. Participants were required to: 1) be at least 18 years old, 2) use headphones, 3) be in a quiet environment where they can clearly see their computer screen, and 4) use an up-to-date Google Chrome browser.

A pre-screening listening task that asked participants to discern differences in the volume of three synthesized sounds (Woods et al., 2017) was presented at the start of the experiment to exclude participants who were not listening attentively over headphones. Successful participants were then shown recordings (audio and video, with latency and jitter applied to both participants as in the original performance) of the first 45 seconds of 15 performances randomly sampled from the corpus, and were asked to rate the overall success of the performance. To ensure consistency between multiple evaluations of one performance, listener ratings were made using the same 9-point scale that was initially given to the musicians in the experiment. (Supplementary Materials Figure S4a). Of the three initially answered by our performers, the performance success question was selected as it was most likely to be comprehensible to listeners, who were not required to have any prior experience of listening to jazz. Participants were told only that the performances were taking place over the internet, and were not informed about latency or jitter.

Eighty-eight adults (44 women, 42 men, 2 nonbinary) with a median age of 38 (*SD* = 14, range = 18 to 75) participated in the study, excluding those who failed pre-screening tasks. For full demographic details, including information on participants' musical background, see Supplementary Materials Figure S4b. An appropriate sample size was determined by calculating the number of participants required to obtain 10 ratings of each of the 130 performances in the corpus, assuming 15 performances rated per participant (equivalent to a study duration of 11.25 minutes, excluding pre-screening). In reality, not all performances received the full number of ratings, due to attrition caused by network time-outs or participants otherwise ending the study early: 31 (23.8%) performances were rated by 9 participants and 6 performances (4.6%) were rated by 8.

The experiment was approved by the Ethics Review Subcommittee at the Faculty of Music, University of Cambridge, UK, and all participants provided written informed consent. Participants were compensated at a GBP £10/hour rate, according to the amount of the experiment they completed; thus, if a participant failed a pre-screening task or left the study early, they were still paid for the proportion of the task that they had completed.

BEAT EXTRACTION

For each of the 130 performances in the corpus, we extracted the position of the MIDI onsets corresponding to each quarter-note beat by manually removing any improvised embellishments from the unmanipulated (live) recordings. The precision of detected beats was ±0.5 ms, corresponding to an internal MIDI resolution of 960 pulses-per-quarter-note at the reference tempo of 120 quarter-note beats per minute.

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

In a small number of performances, the regular quarter-note pulse was occasionally disrupted, either due to mistakes made by participants or due to syncopated anticipation (“pushing”) of the quarter-note beat ahead of its expected metrical position (Berliner, 1994). Missing (or anticipated) quarter notes were detected through inspection of the MIDI data, using the video recordings of a performance for reference. We then realigned these notes into their expected position through linear interpolation between those beats occurring immediately before and after (Figure 3a).

Repeat notes (where one musical event was incorrectly registered as two or more MIDI notes) were filtered from performances in the corpus by discarding any quarter-note beats where a preceding beat had occurred less than 250 ms before. Repeat notes may inadvertently occur when a performer presses a piano key or hits a drum pad several times in rapid succession. We chose this threshold as it was half the duration of quarter-note interbeat intervals at the reference tempo provided to participants (500 ms), and we deemed it unlikely that participants would have accelerated to more than twice this initial tempo (see Supplementary Materials Figure S5).

A nearest-neighbor algorithm was then used to match each quarter-note beat from one performer with the closest equivalent beat played by their partner, with the addition of the latency applied by the testbed at that moment in the performance. In cases where two consecutive beats by one participant could conceivably be matched with the same beat by their partner, the pair with the maximum temporal distance between matched beats was excluded. This process accounted for instances where a performer may have inserted an additional quarter note into their performance to realign with their partner (see Berliner, 1994, pp. 381–382); several of their comments attested to the use of this procedure, e.g., “popped an extra beat in a middle fill” (drummer, duo 1), “I added an extra beat towards the end of the recording to get back on [beats] 2 and 4” (pianist, duo 3).

The raw dataset consisted of 46,640 quarter-note beats, 2.7% of which required linear interpolation. Filtering repeated MIDI notes from the performances amounted to a loss of 0.4% of the raw data, with the process of nearest-neighbor matching incurring a further loss of 4.0%. The final corpus comprised 44,605 matched quarter-note beats, extracted from four hours of performances (Figure 3b).

ANALYSIS

We extracted three objective measures of coordination success from the matched quarter-note beats dataset: 1) tempo slope, capturing systematic digression from the reference tempo; 2) asynchrony, the extent to which both performers remained “in time” with each other; and 3) timing irregularity, the local variability of quarter-note interbeat intervals. Tempo slope and asynchrony were calculated within groups, with a single performance yielding one value for both musicians. Timing irregularity was calculated individually for both participants, with a single performance yielding a separate value for each musician. Alongside these objective measurements, we also considered two subjective indices of performance success as provided by 1) the musicians themselves (two values per performance, one per

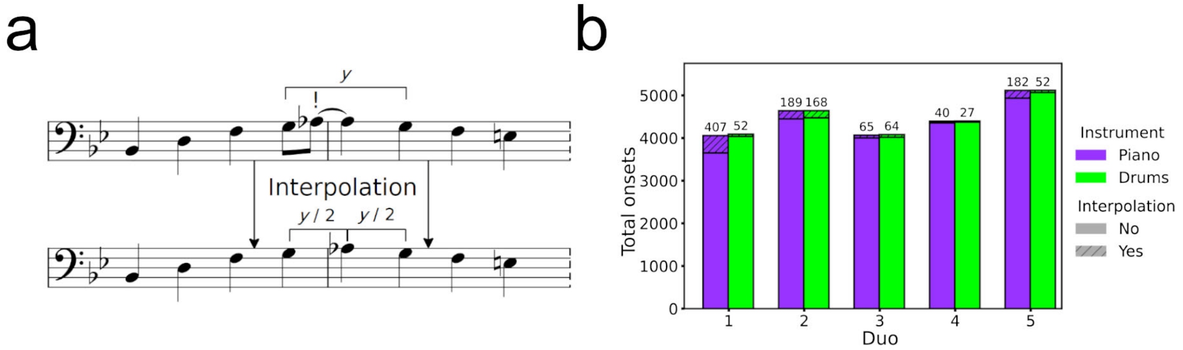

FIGURE 3. Beat extraction procedure. (a) Notation from a hypothetical performance where linear interpolation would have been applied. The note annotated with an exclamation point in the upper stave has been “pushed” before its expected position as the first beat of the next bar. The lower stave shows the performance after interpolation, with this note aligned midway between the quarter notes immediately before and after it (the interbeat interval between which is given by *y*). (b) The total number of beats contributed by each participant to the final corpus, after filtering and nearest-neighbor matching. The hatched segment of each bar shows the proportion of beats by that performer that required interpolation, with the exact number of beats given above the bar.

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

musician), and 2) the participants in our online perceptual study (one value per performance, averaged across all raters).

We define tempo slope as the signed overall tempo change per second within a performance, equivalent to the slope of a linear regression of instantaneous tempo against beat onset time such that a negative slope implies deceleration over time and a positive slope acceleration (Rottondi et al., 2016). Individual coefficients obtained for both musicians in a duo were averaged to reflect the fact that one performer's tempo would not feasibly be independent of their partner's when attempting to play together.

We define asynchrony as the root mean square of the temporal distance between all matched beat pairs articulating the same metrical position played by both musicians (Clayton et al., 2020), including the latency applied by the testbed to both beats. The single asynchrony value obtained from each performance represents the perspective of a hypothetical listener joined to the same virtual "call" as our duo, who would have experienced latency and jitter applied to both musicians' performance equally. This was the perspective adopted by the participants in our online rating study.

We define timing irregularity as the relative temporal instability of a performance, as characterized by the moment-to-moment variability in quarter-note beat durations. This metric was computed by sliding an overlapping window of four seconds duration (equivalent to two measures at the reference tempo and meter) over both performers' quarter-note interbeat intervals (such that the first window spanned the opening four seconds of a performance, and the second window seconds one to five), taking the standard deviation of interbeat intervals within each window, and then finally obtaining the median of all standard deviation values. We opted to use windowed statistics due to their increased robustness to trends and patterns in time series analysis, as the mean interbeat interval duration of a performance would be expected to change over time if the tempo slowed down or sped up.

We obtained separate values for the regularity of each performer's timing in a duo. This is because we were interested in modeling the variability inherent within their individual performance, which we assumed would be the result of both between-participant variance (i.e., differences in note/phrase choices during improvisation across all pianists and drummers), alongside the inherent differences between the roles occupied by the pianist and drummer in a jazz rhythm section (Kilchenmann & Senn, 2015); see Supplementary Materials section §2.1.

We define performer-reported success as the response given by participants to the question "how successful was the performance?" with a rating of 1 indicating an "extremely unsuccessful" performance and 9 an "extremely successful" one. This question was selected for analysis as it demonstrated the best test-retest and inter-rater reliability of the three that we initially asked performers (Supplementary Materials Figure S3b). However, we note here that levels of inter-performer agreement were still not especially high, as indicated by values of Pearson's *r* obtained from the correlation of all pianist-drummer scores in one duo: mean *r* = .40 (*SD* = .24, range = .17 to .73), suggesting only moderate agreement on average. Disagreements of this kind are common in group musical improvisation, however, where performers rarely share the same understanding of what unfolded (Pras et al., 2017; Schober & Spiro, 2014), and this was not taken to indicate any inherent lack of reliability in how this question had been presented.

We define listener-reported success as the ratings of overall performance success given by listeners in our online perceptual experiment, in response to the same question initially presented to our performers. Individual ratings of the same condition were generally consistent across participants (mean *SD* = 1.64; see also Supplementary Materials Figure S4b), which led us to average ratings obtained for each performance. Average levels of listener-pianist and listener-drummer agreement were broadly equivalent with the levels of pianist-drummer agreement given above: when values of the correlation coefficient *r* were averaged across all duos, listener-pianist mean *r* = .33 (*SD* = .39, range = -.32 to .72), listener-drummer mean *r* = .40 (*SD* = .15, range = .19 to .56). Our musicians did not hold a privileged understanding of the success of their own performances (Schober & Spiro, 2014): they agreed with listeners to an equivalent degree that they agreed with each other.

TRANSPARENCY AND OPENNESS

We report how we determined our sample size, all data exclusions, all manipulations, and all conditions used in both studies described in this article, and we follow Journal Article Reporting Standards (Kazak, 2018). All data and research materials (with no exceptions) are posted under a permissive license to a trusted third-party repository (accessible using the stable link https://doi.org/10.5281/zenodo.7773824). Our complete analysis scripts, software, and code book have also been made publicly accessible to enable readers to replicate our analyses (code: https://github.com/HuwCheston/

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

Jazz-Jitter-Analysis, testbed software: https://github.com/HuwCheston/AV-Manip, online perceptual study software: https://github.com/HuwCheston/2023-duo-success-analysis). Statistical analyses were conducted using the SciPy (Virtanen et al., 2020) and Statsmodels (Seabold & Perktold, 2010) packages in the Python programming language (version 3.10.2). The design and analysis of the studies reported in this paper were not preregistered.

Results

BASELINE MEASUREMENTS FOR NON-DELAYED PERFORMANCES

In the following discussion, we outline the baseline results obtained from the control condition, when no latency or jitter was applied to the performance (Figure 4). The baseline mean tempo slope for all our duos during the control condition was 0.04 beats-per-minute-per-second (BPM/s) (*SD* = 0.03, range = −0.02 to 0.10 BPM/s), indicating that the tempo of performances remained stable when no latency was present, though with a slight tendency towards acceleration. In Figure 4 we compare these results with tempo slope coefficients obtained from previous networked performance studies, noting that the behavior of our duos did not differ from expected standards. See Supplementary Materials Figure S5 for individual "tempo map" plots for each performance.

The baseline mean asynchrony for our duos in the absence of latency and jitter was 33.1 ms (*SD* = 7.2, range = 22.5 to 48.8 ms), which we compare in Figure 4 to prior studies of real-time ensemble performance across various musical genres. We note that this value is greater than the asynchrony observed in a prior study of jazz bass and drums synchronization (Kilchenmann & Senn, 2015), closer instead to the synchronization of Cuban salsa or North Indian raga musicians (Clayton et al., 2020). This may be because of the random placement of the control within each experimental session and participants' overall lack of awareness of our manipulations, such that they could have been primed to adopt particular strategies in non-delayed performances as a result of earlier conditions where latency had been present. We refer to the comment of one pianist here, that it "felt slightly more difficult to coordinate, and to be creative" (pianist, duo 4) during the control in comparison to the previous condition they encountered.

Another explanation is that the digital environment created by the testbed simply made it harder to play music together effectively versus the face-to-face, copresent conditions used in these reference studies, even without latency and jitter (see Doherty-Sneddon et al., 1997). The inherent latency of the video footage, for instance, could have been sufficiently distracting to the performers to result in a higher-than-expected baseline asynchrony.

The baseline timing irregularity was 16.1 ms (*SD* = 5.8, range = 7.9 to 22.1 ms) for drummers and 26.4 ms (*SD* = 4.5, range = 21.9 to 36.7 ms) for pianists, which suggests significantly lower variability in the timing of drummers compared to pianists. Note that no participant demonstrated precise temporal isochrony: indeed, anisochronous timing has been theorized as aesthetically preferable over quantized isochrony in "groove-based" music such as jazz, due to the increased rhythmic interest it imparts ("participatory discrepancies"; see Keil, 1987).

The baseline mean performer-reported success score was 7.9 (*SD* = 0.7, range = 7 to 9) for drummers and 6.9 (*SD* = 1.9, range = 2 to 8) for pianists, suggesting that real-time performances were regarded as more successful than unsuccessful. Inter-participant agreement was typically higher in their evaluation of the control condition than in the remainder of the experiment, with an absolute difference between pianist and drummer scores of no more than 1 obtained for all control performances bar one.

The baseline mean listener-reported success score was 7.1 (*SD* = 0.6, range = 6 to 8) for the control condition; in Supplementary Materials Figure S4c, we demonstrate that there were no significant differences in mean listener score for performances in the control condition across any duo. These results again indicate that the non-delayed, real-time performances of all duos were consistently regarded as successful.

Linear Associations Between Variables

In Figure 5, we show the pairwise associations between the five performance success variables discussed above; to minimize overplotting, each datapoint on this graph corresponds with the response of one duo to a single condition, averaged across both sessions of the experiment. Note that this also means that the timing irregularity and performer-reported success variables are averaged over both members of a duo to ensure the number of values plotted remains consistent. Tempo slope is given in its absolute (unsigned) form to show the relationship between individual metrics and deviations from the reference tempo, regardless of whether this change had occurred via acceleration or deceleration.

We observe here that the three objective metrics derived from the extracted quarter-note beats (absolute

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

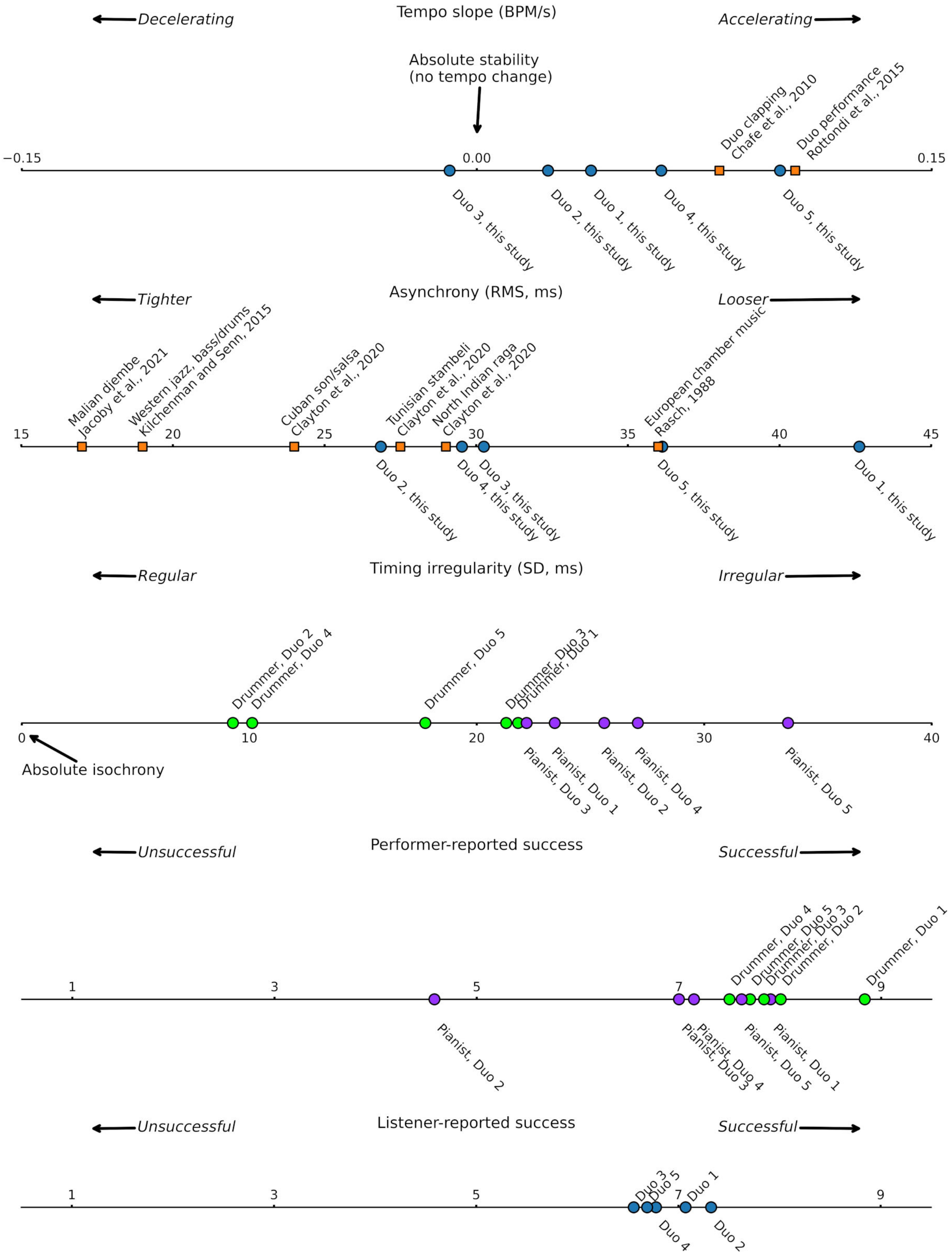

FIGURE 4. Baseline results. Number lines showing baseline values obtained for tempo slope, asynchrony, timing irregularity, performer- and listener-reported success, averaged for both repeats of the control condition by a duo. Note that performer-reported success values are randomly displaced horizontally for increased visual clarity and to prevent over-plotting.

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

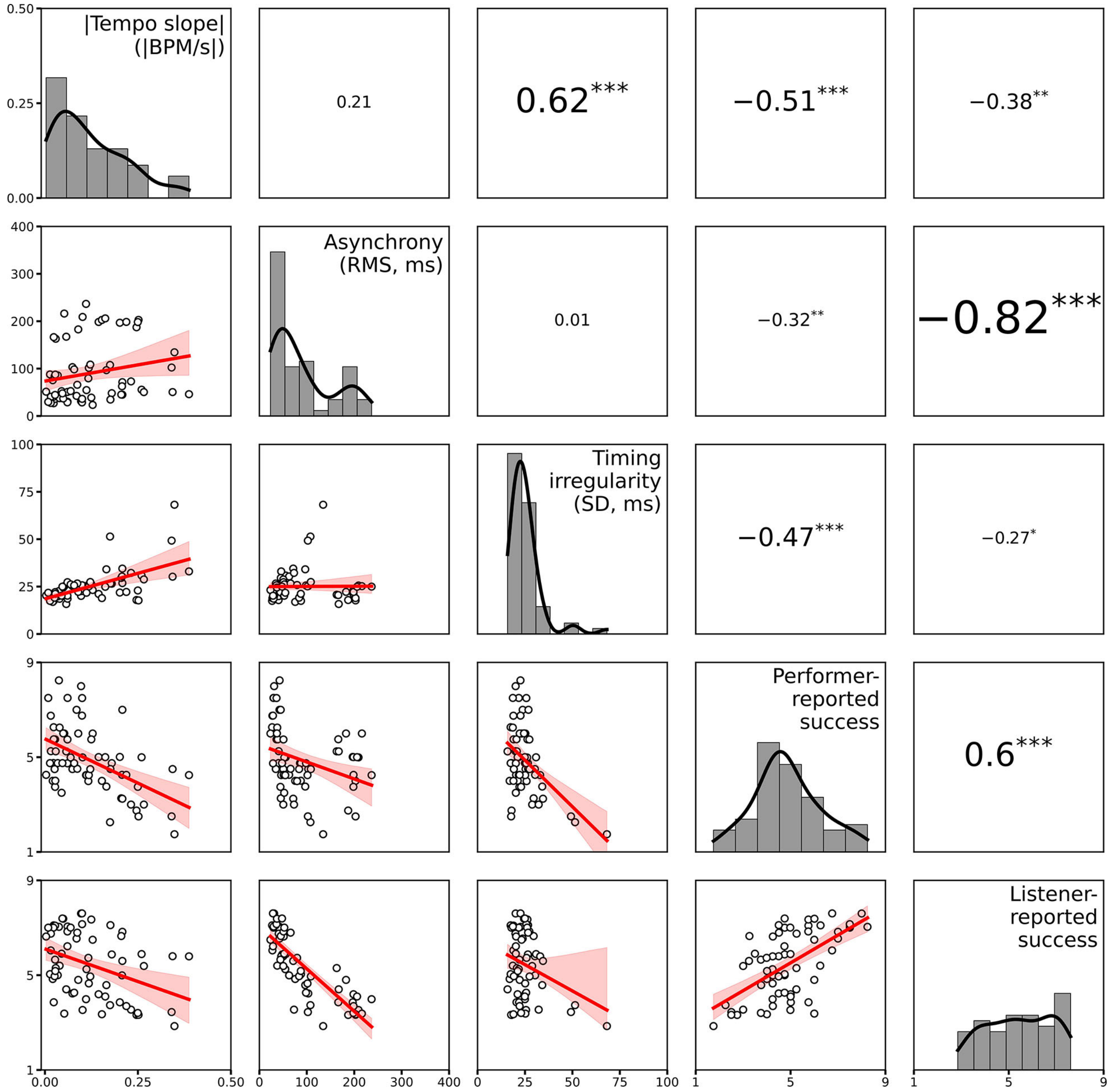

FIGURE 5. Univariate and bivariate distributions. The histograms on each diagonal show the distribution of the variable plotted in that column/row. The scatter plots below the diagonal show the pair of variables obtained at the intersection of every column and row, with markers representing the score obtained by each duo for all thirteen conditions, averaged across instruments and sessions of the experiment. The straight red lines show a linear regression model fit between both variables, with error bars denoting 95% confidence intervals generated via bootstrapping with 10,000 replicates. Likewise, values above the diagonal denote the coefficient of Pearson's r calculated between the corresponding variable pair, with the font size also indicating the strength of the correlation. Asterisks indicate the significance of the correlation coefficient, *p < .05. **p < .01. ***p < .001. Values for the tempo slope variable are given in their absolute (unsigned) form.

tempo slope, asynchrony, and timing irregularity) were all negatively correlated with both reported success variables, such that performances that diverged from the reference tempo and displayed lower synchronicity and isochronicity were evaluated less favorably by participants and listeners. Additionally, tempo slope coefficient and timing irregularity were also positively correlated with each other, with a larger magnitude of tempo change associated with more unstable performances.

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

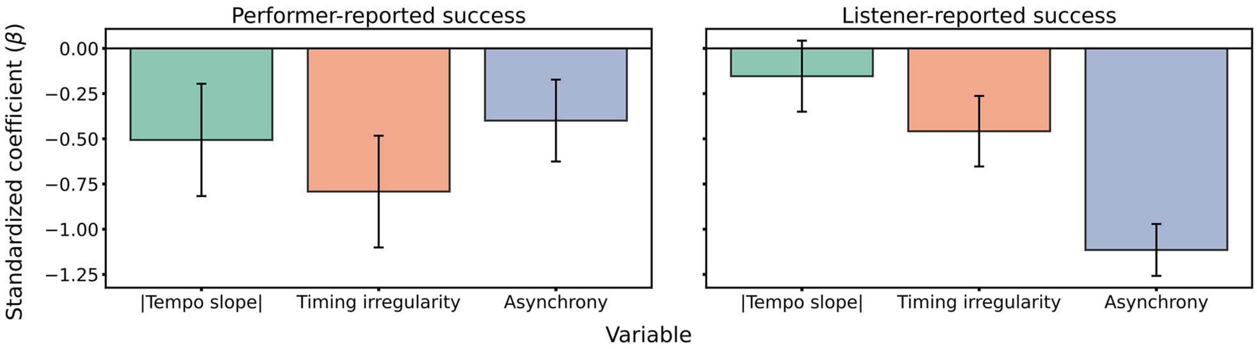

FIGURE 6. Objective features affecting success evaluations. The panels show standardized regression coefficients (β) for a mixed-effects model predicting success ratings from musicians and naive listeners respectively. Error bars represent 95% confidence intervals computed by bootstrapping with 10,000 replicates; where an error bar does not cross 0, the effect of that predictor is statistically significant.

To further establish the relative importance of the different factors considered when evaluating a networked performance, we fitted a mixed-effects model to the dataset, predicting performer- and listener-rated success using a combination of absolute tempo slope, timing irregularity, and ensemble asynchrony (Figure 6). A random effect (intercepts and slopes) of duo number was included in both models, as we wished to model group-specific variations in both the baseline rating and in the effects of the predictor variables. The statistical significance of the fixed effects was assessed by bootstrapping over duos with 10,000 replicates to produce 95% confidence intervals. For performers, an increase in timing irregularity was the strongest predictor of a decrease in subjective rating, although all three predictors were significant; for listeners, timing irregularity and asynchrony were both significant predictors of rating decreases (with asynchrony having the stronger effect), while tempo slope was not significant.

The amount of variance explained by both the fixed and random effects (conditional R^2) was .762 for the performer reports and .856 for the listener reports. The amount of variance explained by only the fixed effects (marginal R^2) was .483 for the performer reports and .735 for the listener reports. The standard deviation in scores estimated for the random effect of duo was .970 for the performer reports and .520 for the listener reports. These statistics indicated that the objective metrics effectively summarized the proximal causes of subjectively evaluated performance success; listeners cared most about whether performers remained stable and synchronized, but not whether they changed tempo, while musicians considered all three factors to be important.

NETWORK LATENCY IMPAIRS MUSICAL PERFORMANCE

To evaluate the effect of testbed configuration on the performance of each group, we fitted a separate linear model to the data obtained for each of our five duos, using each of our performance success variables as a response measure (twenty-five models total). Latency and jitter were included as predictors in every model and were treated categorically (with the control condition as the reference category), due to the possibility of non-monotonic effects. For models predicting timing irregularity and performer-reported success, where separate values had been obtained individually for both musicians in a duo, instrumental role was additionally included as a predictor (with the drummer's performance used as the reference category).

We accounted for our repeated-measures design by averaging the results obtained from a duo for a particular condition across both sessions of the experiment. Values of Pearson's r obtained from the correlation of scores from both sessions of the experiment indicated good to excellent test-retest reliability for each metric (Supplementary Materials Figure S6a): for tempo slope, $r(63) = .66$, $p < .001$, for timing irregularity, $r(63) = .86$, $p < .001$, for asynchrony, $r(63) = .98$, $p < .001$, for performer-reported success, $r(63) = .67$, $p < .001$, and for listener-reported success, $r(63) = .78$, $p < .001$. Figure S6b reports bootstrapped confidence intervals for the difference in mean scores (N = 10,000 replicates) for each metric across both sessions. Barring duo 2 demonstrating a significantly greater mean tempo slope coefficient in the second session compared to the first (mean difference = 0.03, 95% CI: [0.02, 0.04]), there were otherwise no significant differences in means. These analyses suggest that the behavior of each duo was consistent across both sessions of the experiment, validating our decision to average their scores.

The average R^2_{adj} value for our models was .529 when predicting tempo slope (SD = .529, range = −.390 to .884), .985 when predicting asynchrony (SD = .009, range = .973 to .993), .908 when predicting timing

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

irregularity (*SD* = .051, range = .842 to .961), .717 when predicting performer-reported success (*SD* = .175, range = .407 to .820), and .885 when predicting listener-reported success (*SD* = .062, range = .821 to .971). We took this as an indication that testbed configuration and instrumental role alone were generally strong enough predictors to explain a large degree of the variance in each performance success metric, albeit with greater spread of R^2_{adj} values obtained for some variables than others.

Figure 7 plots the coefficients and confidence intervals obtained for every predictor and categorical level in our models. The following discussion is organized to address in turn the effect of testbed configuration (latency and jitter) on indicators of overall performance success for each duo. Discussion of instrumental role as a predictor is contained in Supplementary Materials section §2.3, and inferential statistics for each success metric across every latency and jitter value (averaged over all duos) are shown in Supplementary Materials Figure S7.

Effects of Latency

Increases in latency were strongly correlated with increases in ensemble asynchrony, $r(63) = .97$, $p < .001$, as would be expected. The effect of latency on the remaining metrics was more complex. The lowest amount of latency we tested, 23 ms, predicted significantly reduced ratings of performer-reported success for three duos, but no equivalent changes in any other metric. Above this value, latency had a detrimental effect on many performance features. Both 45 and 90 ms of latency produced significant decreases in tempo slope and increases in timing irregularity for two duos, alongside reductions in performer-reported success for all duos. For listener-reported success values, 45 ms of latency predicted significant decreases in ratings for two duos and 90 ms for all duos. These differences between performer- and listener-reported success at 45 ms latency could suggest a lower tolerance for latency exists when performing versus listening to music. In total, our results replicate many of the "classic" findings of prior networked performance studies, where latency typically contributes to a recursive slowing in the tempo of a performance and reductions in timing regularity and ensemble synchrony, alongside reductions in subjective assessments of performance quality (Bartlette et al., 2006; Chafe et al., 2010; Monache et al., 2019; Rottondi et al., 2015).

Surprisingly, the maximum amount of latency we tested, 180 ms, was not associated with significant decreases in tempo slope coefficient for any duo, which contradicts the argument in Chafe et al. (2010) of a linear relationship between increases in delay time and decreases in tempo. Instead, this amount of latency was associated with positive (albeit nonsignificant) tempo slope coefficients for four out of five duos, suggesting that their tempo had either accelerated or remained stable. We suggest two possible explanations for this phenomenon; extreme latency values could have resulted in participants" 1) operating either on "autopilot," ignoring their partner and focussing solely on their own performance; or 2) perceiving their partner as playing ahead of (rather than behind) them and speeding up in an attempt to catch them, rather than slowing down to meet them. This phenomenon is explored in further detail in Supplementary Materials section §2.2.

The free-text responses given by participants after each performance attested to these effects of latency. Comments ranged both from acknowledging the changes in performance tempo ("to me it felt like a constant rallentando that kept failing to land": drummer, duo 1, "there was a tendency to rush": drummer, duo 3) and timing regularity ("when the time isn't settled I have a tendency to tense up": drummer, duo 2) caused by latency, to annoyance at their inability to interact with their partner successfully ("frustrating to not be able to use body language effectively": drummer, duo 1), and finally to a dislike of their performance in the testbed environment ("absolute carnage and I think it sounded utterly awful": pianist, duo 1). Additionally worth noting here is the typically more positive tone used to describe performances made at 180 ms latency ("everything seemed to align this time": pianist, duo 1, "we were on the same page with this one": pianist, duo 3, "lots of creative energy": pianist, duo 4) than lower values.

Effects of Jitter

While a spike in network latency time caused by jitter inevitably causes large asynchronies in performances at a local scale, at a global scale the jitter conditions we tested had a much smaller impact on our performance success metrics than latency time alone. The presence of 1.0x jitter predicted significant increases in asynchrony for two duos, increases in timing irregularity for one duo, decreases in performer-reported success for one duo, and decreases in listener-reported success for two duos. The magnitude of these effects, however, was typically small in comparison to the other predictors included in each model. Removing jitter as a predictor resulted in little change to how well each model fit the data obtained for any variable or duo (average

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

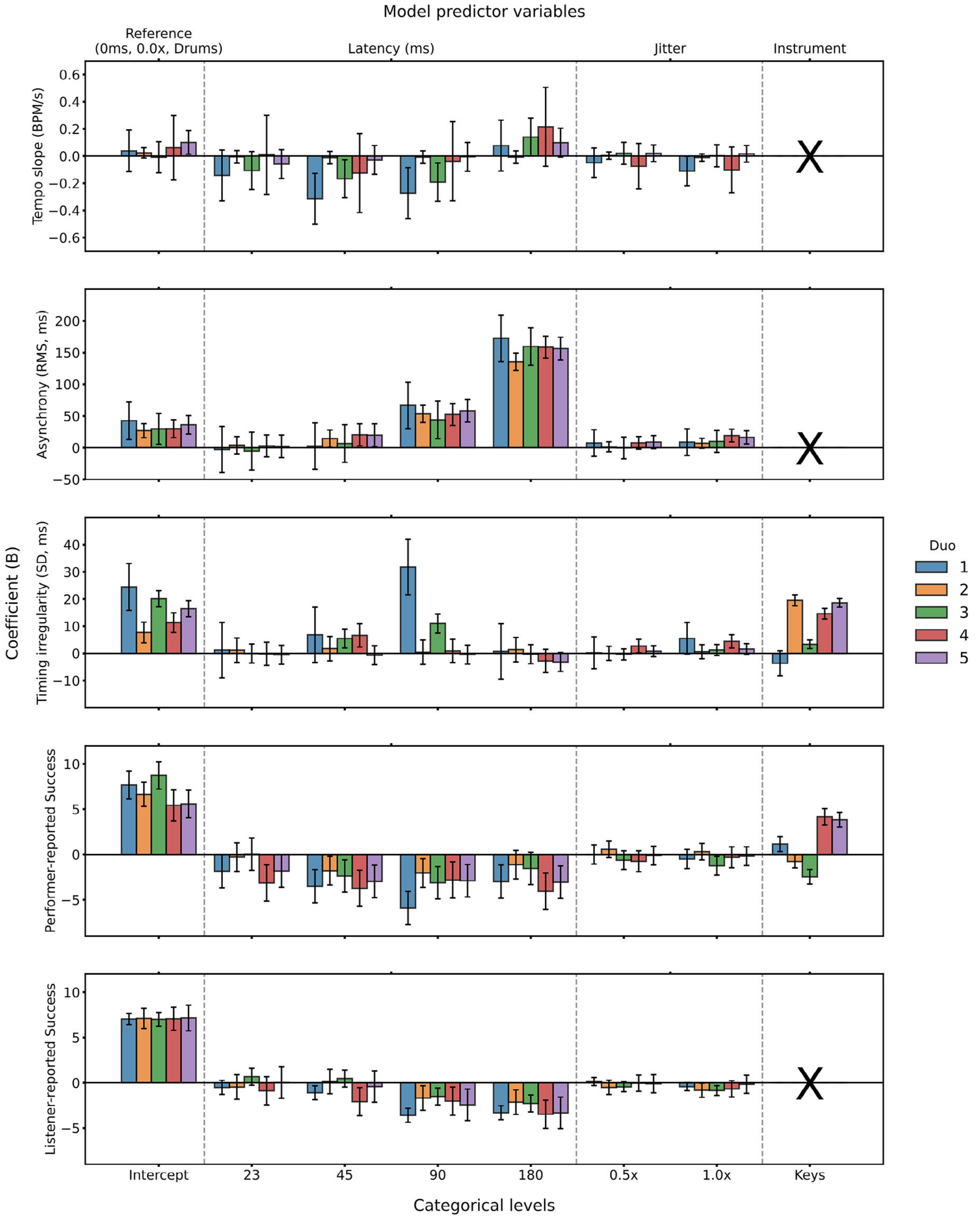

FIGURE 7. Effects of testbed configuration on performance success variables. The bars show regression coefficients from models predicting one of the five performance success variables, with results split by duo. The reference category for each model corresponds to the performance during the control condition (i.e., with no latency or jitter); if values were obtained separately for each instrument for a given variable, the reference category also corresponds to the drummer's performance. Crosses indicate where a particular predictor variable was omitted from that model. Error bars represent 95% confidence intervals computed by the model; where an error bar does not cross 0, the effect of that category is statistically significant.

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

ΔR^2_{adj} = 0.01), in comparison to removing either latency (average ΔR^2_{adj} = 0.69) or instrumental role (average ΔR^2_{adj} = 0.61) (Supplementary Materials Figure S8).

A complementary way to quantify the effect of jitter is to analyze whether moment-to-moment fluctuations in latency were followed by increased variability in performance timing. To do so, we measured the standard deviation of the quarter-note interbeat intervals played by a performer and the latency time applied by the testbed across a sliding window of four seconds duration. We then computed partial correlations between these two time series, with latency variability lagged at increasing one-second intervals between 0 and 8 seconds (or two bars at the reference tempo and meter). Any prior variation in a performers' timing up to this lag was controlled for in the correlation, to account for the probability that subsequent quarter-note beat durations in jazz performance may display autocorrelation with previous values (Cheston, 2022). Put differently, when computing the partial correlation between timing variability *t* and latency variability *d* at lag *k* seconds, we controlled for prior timing variability at all lags up to and including lag *k*, except for when $k = 0$ (in which case the correlation coefficient used was Pearson's *r*, with no controls).

We calculated these partial correlations separately for all performances made using the three jitter values tested in the experiment, regardless of minimum latency time (Supplementary Materials Figure S9): note that minimum delay and delay variability were independent of each other (see Figure 2, right column). We observed that, when the 1.0x jitter scaling was used, previous variation in network latency was positively correlated with future variation in performance timing. The strongest association between prior increases (or spikes) in latency and future increases in timing variability occurred at a lag of two seconds: averaged across performances by all musicians made with 1.0x jitter, mean *r* = .11 (95% CI: [.08, .14], obtained via bootstrapping across all obtained values of *r* with 10,000 replicates).

However, the small magnitude of this correlation suggests that jitter only had a slight impact on performance stability, and the musicians' own comments validate this claim. While they were evidently aware of its presence ("moments of ride cymbal jolting": pianist, duo 4, "I noticed the [video] fluctuating": pianist, duo 5), participants were able to develop strategies to accommodate jitter. These included inserting or removing beats from the underlying meter ("there were several points where we were suddenly playing on different beats to each other but it was easy to add/drop a beat to come back in time": drummer, duo 3), cultivating a deeper sense of rhythmic intensity in the performance ("subtle disruptions in the feed were subsumed within the strength of our interaction/groove": pianist, duo 4), and looking at the video feed ("eye contact and watching the fingers on the [piano] keys helped": drummer, duo 3). Indeed, one participant even claimed that the disruption caused by jitter had had a positive effect ("disruptions through the feed, but these helped with the flow of the music," ". . . we were able to use [the disruptions] to interact consistently and vibrantly": pianist, duo 4).

MUSICAL ENSEMBLES ADOPT DIVERSE COORDINATION STRATEGIES DURING NETWORKED PERFORMANCE

The members of any musical ensemble coordinate via complex, distributed processes of mutual attending and adaptation (Clayton et al., 2020; Jacoby et al., 2021; Timmers et al., 2014; Wing et al., 2014), which may be referred to by jazz musicians as "hooking up," "grooving," and "swinging" (Doffman, 2014; Monson, 1996). For a group of musicians to remain synchronized, they must adapt to any small deviations from isochrony in each other's performances. When one musician adapts to match variation in another's performance, we can say that they are influenced by—or "coupled with"—that musician; vice-versa, the absence of coupling implies that one musician does not correct for variability in another's performance, and is thus not influenced by them (Jacoby et al., 2021; Konvalinka et al., 2010).

In the context of networked music-making, the perception of timing variability in a performance will be affected by any network instability or jitter present in the output signal: a musician does not, therefore, couple with the real-time performance of their partner, but with the delayed feedback they receive from the network. We begin this section by describing the linear phase correction model we employed to model this process, alongside a series of control analyses we conducted to validate it. We then describe the results from applying these models to the performances in our corpus, including evidence suggesting the presence of two distinct coordination strategies employed by the duos who participated in our experiment.

Phase Correction Modeling

We model the coordination in a networked ensemble using a process of linear phase correction (Vorberg & Wing, 1996), where a performer's upcoming quarter-note interbeat interval is predicted from both the duration of their prior interbeat interval and the asynchrony with their partner at the previous quarter note beat (Jacoby et al., 2021). We consider the particular coordination strategy adopted by a networked ensemble to be

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

equivalent to the complete system of coupling responses established between every pairwise combination of musicians in an ensemble, and hence we create a separate model for each performer in a duo.

As has been noted previously, the tempo of performances in our corpus often drifted as a result of the latency applied by the testbed. When a performance accelerates or decelerates over time in this manner, the mean duration of the quarter note beats by one performer will trend towards shorter or longer values, respectively. As such, rather than using the actual duration of these quarter-note intervals directly, we instead considered the difference between the duration of successive quarter notes played by a single performer when creating our model. We represent values that have been transformed in this manner using prime notation.

Formally, our model can be written as:

$$T'^{i,i}_{k+1} = \alpha_{i,i}T'^{i,i}_{k} + \alpha_{i,j}(T^{i,j}_{k} + d'^{j}_{k}) + \alpha_{i,0} + \varepsilon^{i}_{k} \quad (2)$$

where $T'^{i,i}_{k+1}$ is the difference between the durations of the quarter note interbeat intervals by musician i at beats $k+1$ and k, $T'^{i,i}_{k}$ is the difference between the durations of the two quarter note interbeat intervals directly preceding $T'^{i,i}_{k+1}$ by musician i (i.e., at beats k and $k-1$), $T^{i,j}_{k}$ is the asynchrony between musicians i and j at beat k, d'^{j}_{k} is the variable delay applied to musician j's performance by the testbed at beat k, $\alpha_{i,i}$ is the influence of the previous interbeat interval difference by musician i on the duration of i's future quarter notes, $\alpha_{i,j}$ is the coupling coefficient reflecting the influence of musician j on future quarter note durations by their partner i, $\alpha_{i,0}$ is an intercept term specific to musician i, and ε_i is residual noise (Figure 8a; see Supplementary Materials section §2.4. for a full description of this model).

Control Analyses

The average R^2_{adj} for our model was .487 (SD = .128, range = .094 to .742), meaning that it typically captured about half of the total variability in differenced interbeat interval durations during a performance. We further verified the robustness of our model by conducting a series of control analyses, described in detail in Supplementary Materials section §2.5. and Supplementary Materials Figure S10.

We began by comparing the partner-coupling coefficients $\alpha_{i,j}$ obtained from each participant across the first and second session of the experiment (Supplementary Materials Figure S10a), and across the first and second (45-second) half of each piece (Figure S10b). A strong positive correlation existed between coupling behavior in the first and second session of the experiment, $r(63)$ = .81, $p < .001$, and across both halves of one performance, $r(128)$ = .67, $p < .001$. We then explored whether a higher-order linear phase correction model (Vorberg & Wing, 1996) that considered a longer prior history—incorporating values of $T'^{i,i}$ and $T^{i,j}$ at further quarter-note lags of k, up to one measure—would perform better than our initial model. The results were similar to the main analysis (Figure S10c), suggesting that a lag of one quarter note was sufficient when predicting future interbeat intervals.

Finally, we generated a series of simulated performances for each condition tested in the experiment, wherein the durations of artificial interbeat intervals for two musicians were predicted from the models created for each duo in that condition, with the same amount of latency and jitter as was applied in the experiment. Motor variance was simulated by applying Gaussian noise to every artificial interbeat interval, with zero mean and standard deviation equal to residual noise term ε^{i}_{k}. Simulated and observed tempo slope coefficients were strongly correlated, $r(63)$ = .89, $p < .001$, as were simulated and observed values of ensemble asynchrony, $r(63)$ = .96, $p < .001$ (Supplementary Materials Figure S10d; see also Figure S12). In summary, our control analyses indicated that our model had adequately captured both the rhythmic adaptation present within each duo and the effect that this had had on their networked performance and, indeed, that coupling was a robust and internally consistent measurement of ensemble coordination.

Coupling Responses

Supplementary Materials Figure S11a plots distributions of the self-coupling coefficients $\alpha_{i,i}$, partner-coupling coefficients $\alpha_{i,j}$, and intercept terms $\alpha_{i,0}$ obtained from all models. Consistent with previous studies employing similar models (e.g., Jacoby et al., 2021), all observed values of the coupling coefficient $\alpha_{i,j}$ were positive, mean 0.474 (SD = 0.299, range = 0.015 to 1.215), suggesting that participants had coupled to their partner to some degree in every performance. Figure 8b shows the average coupling of each participant to their partner; Figure S11b depicts the average coupling response for each participant across both repeats of one condition, with tempo slope values given also as an indicator of overall performance success.

To compare the relative influence of both musicians, we obtained bootstrapped confidence intervals (N = 10,000 replicates) for the mean difference between pianist-drummer and drummer-pianist coupling

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

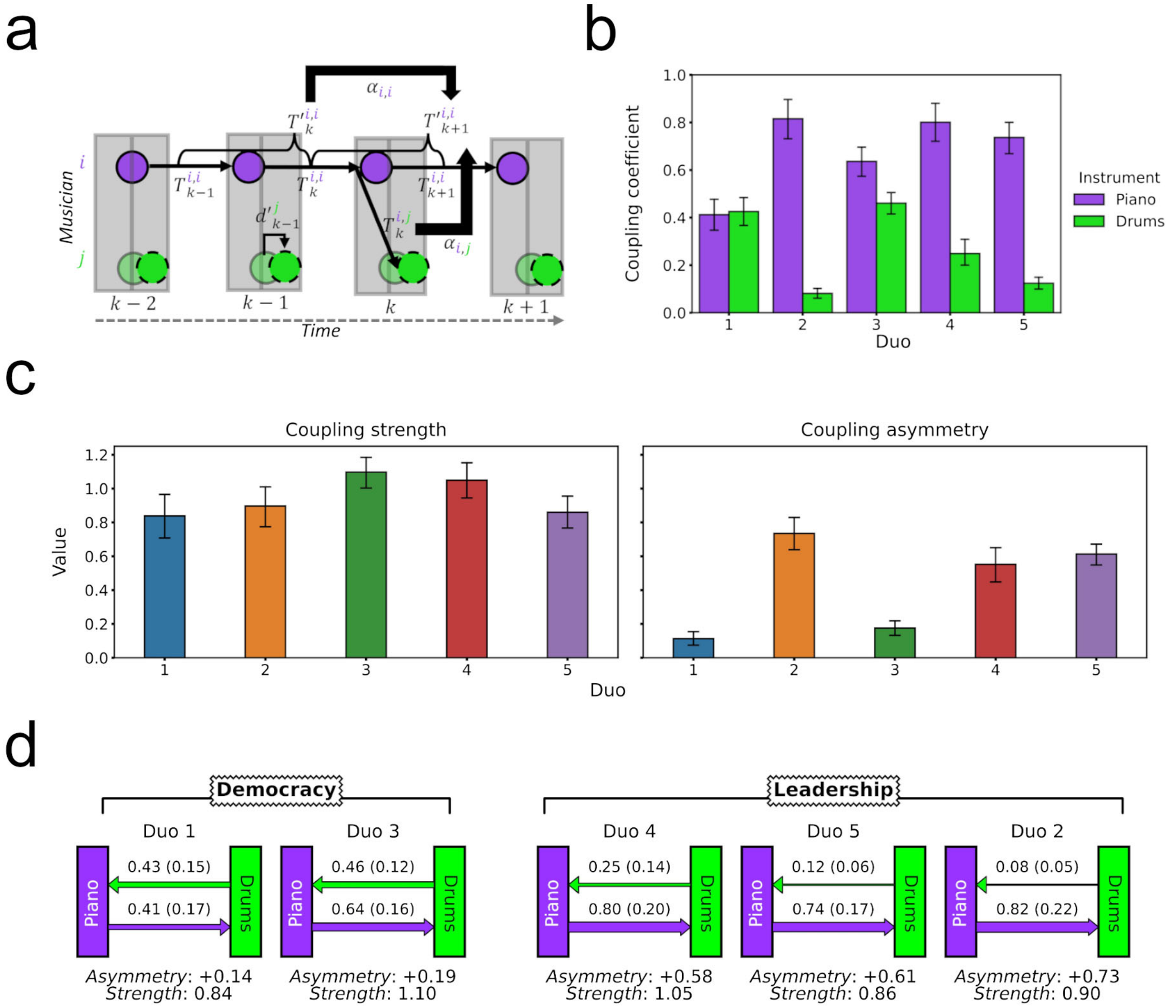

FIGURE 8. Measured coupling between the musicians in each duo. (a) A schematic diagram of the linear phase correction model. Quarter-note beats are given by colored circles: the circles with dashed borders show the position of beats after latency has been applied, representing when they would have actually been heard by a performer. The vertical gray rectangles indicate the metric grid of quarter notes, while horizontal and diagonal black lines show the interbeat interval between quarter notes, either from one performer or between performers. The braces indicate the difference between successive interbeat intervals played by the same musician. (b) The mean coupling coefficient obtained for each participant. (c) The mean coupling strength and asymmetry across each duo. Error bars in both (b) and (c) show 95% confidence intervals of the mean obtained via bootstrapping with 10,000 replicates. (d) Duos grouped by their respective coordination strategy, with the direction and degree of the coupling in each duo given by the color and thickness of the arrows respectively. Values above each arrow show the mean coupling coefficient, with parentheses indicating standard deviations. Duos are ordered, left-to-right, by average coupling asymmetry across all conditions.

coefficients in each of our duos (Supplementary Materials Figure S11c). For 4 out of the 5 duos, the drummer had exerted significantly more influence on the pianist than the pianist had exerted on the drummer, with duo 1 being the only group where neither musician had emerged as significantly more influential than their partner (mean difference in coupling, duo 1 = -0.01, 95% CI: [-0.08, 0.05]). With regards to these differences in coupling between instruments, we refer to the comment in Chafe et al. (2010) that, during a networked performance, "the weaker side (in terms of rhythmic function) naturally follows the strong one" (p. 990); Chafe's example of a guitarist following a drummer in a networked performance bears resemblance to the pianists in our duos, who typically followed their drummer partners.

Coordination Strategies

We evaluated the coordination strategy employed by each ensemble by considering the strength (or "gain")

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

and asymmetry of their coupling, equivalent to the sum of and absolute difference between the two partner-coupling coefficients obtained from a single performance, respectively. We show the average coupling strength and asymmetry for each duo in Figure 8c, and in Supplementary Materials Figure S11d we present confidence intervals for the differences in mean coupling strength and asymmetry calculated across all independent pairwise combinations of duos (10 combinations total). No correlation was found between coupling strength and asymmetry, $r(128) = .13$, $p = .14$.

The members of duo 3 exhibited the strongest ensemble coupling out of all the groups studied, with a mean coupling strength of 1.10 (*SD* = 0.24, range = 0.56 to 1.47). Significant differences in mean coupling strength were found between this duo and all other groups apart from duo 4 (difference in mean coupling strength, duo 3/4 = -0.05, 95% CI: [-0.14, 0.05]). Duo 1 displayed the weakest coupling overall, with a mean coupling strength of 0.84 (*SD* = 0.28, range = 0.27 to 1.35). All in all, however, coupling strength did not differ to a particularly large extent between the duos, with the average sum of coupling coefficients falling within the range suggested by Vorberg (2005) to be required for a stable performance by two musicians.

Coupling asymmetry varied more across the duos and indicated the presence of two distinct coordination strategies. The coupling within duo 1 was the most symmetrical of all groups studied, with a mean coupling asymmetry of 0.14 (*SD* = 0.09, range = 0.00 to 0.30), indicating that the distribution of error correction was almost entirely equal between pianist and drummer. No significant differences were observed between the mean coupling asymmetry of this group and duo 3 (0.06, 95% CI: [-0.01, 0.12]; see also Figure 8c). Thus, and despite the drummer of this latter duo emerging as more influential than the pianist, we consider both duos 1 and 3 to best embody the same coordination strategy—that of egalitarianism or "democracy" (Wing et al., 2014), where both musicians had adapted to each other at equivalent (or near-equivalent) levels, such that neither could be said to clearly and definitively occupy a leadership role in the ensemble.

The self-reports from participants in these two groups reinforced our labeling of their interaction as democratic, involving attempts to maintain the reciprocal co-adaptation in timing typical of interpersonal action coordination (Nowicki et al., 2013). References were continually made by these participants to an inability to choose whether or not to lead the performance or follow their partner ("difficult to decide whether to plow on at correct speed when things go awry or to try and match [the pianist]": drummer, duo 1) and, even when such a decision was made, they were not necessarily able to manifest this in their performance ("this time I tried to resist and keep the initial tempo but it didn't work," ". . . tried to lead tempo again but gave up": pianist, duo 3). This occasionally contributed to situations where both participants directly disagreed as to who was attempting to lead the other, as seen in two remarks made about the same performance by duo 3: "felt like [the drummer] was following my tempo this time" (pianist), "I had to play a beat ahead of [the pianist]" (drummer). The overall sense amongst these groups was one of confusion about how best they should coordinate with their partner, leading to performances that felt more "like a battle of wills" (drummer, duo 1) than truly interactive.

Coupling in the remaining three duos was less balanced, with duos 2, 4, and 5 all displaying significantly greater coupling asynchrony than both duos 1 and 3 (Figures 8c, Supplementary Materials Figure S11d). We therefore considered these three duos to instead embody a "leadership" coordination strategy, where one musician had adapted significantly less to their partner than their partner had adapted to them (Goebl & Palmer, 2009; Konvalinka et al., 2010). As noted previously, it was the drummer in all three duos who emerged clearly as the leader here, with the pianist thereby acting as the follower. Duo 2 exhibited the most unbalanced coupling overall, with a mean coupling asymmetry of 0.73 (*SD* = 0.21, range = 0.28 to 1.19): these performers thus established the strongest leadership dynamic of all the groups we studied. There were no significant differences in mean coupling asymmetry between the remaining two leadership groups, duos 4 and 5 (0.06, 95% CI: [-0.01, 0.14]).

Although fewer self-reports were provided by the participants in these three duos than those in the democracy groups, they were nonetheless revealing in demonstrating an awareness of the leader-follower relationship that they had established. One drummer made direct reference to this strategy, describing how they had ignored their partner while their partner had followed and adapted to them: "I worked out where I had to play in relation to [the pianist]. It seemed that it appeared to [the pianist that] we were playing in sync, however I displaced my beat by one triplet quaver against the pulse I got from [them]" (drummer, duo 4). We also note here that the performers in the three leadership duos demonstrated substantially greater disagreement in their reported success scores than the musicians in the two democracy duos, as demonstrated by lower values of Pearson's *r* (Supplementary Materials Figure S3b). This

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

again supports our claim that a divergence in ensemble role took place for the musicians in these groups that did not occur in the two democracy duos.

In Figure 8d, we visually depict the coupling networks established by all five duos and group them under either "democracy" or "leadership" headings.

SIMULATIONS DEMONSTRATE TRADE-OFFS IN COORDINATION STRATEGIES USED IN NETWORKED PERFORMANCES

Evidence for democratic and leadership coupling between performers can be found throughout the literature on musical performance (Goebl & Palmer, 2009; Jacoby et al., 2021; Timmers et al., 2014; Wing et al., 2014). The coordination strategy adopted in any ensemble likely depends on the appropriateness of this strategy for their performance situation and the style of music they play. For instance, mutual co-adaptation ("democracy") was found to be more effective in coordinating temporal alignment during real-time, face-to-face jazz improvisation by duos than in non-contingent, asynchronous "overdubbed" performances (Setzler & Goldstone, 2020). Beyond jazz, "leadership" coordination has been observed in Classical string quartets, where artistic leadership has typically been attributed to the first violin, with the remaining instruments taking up other roles (Timmers et al., 2014). Finally, studies of West African drum ensembles have found that rhythmic adaptation was distributed asymmetrically across performers and reflected their social organization (Jacoby et al., 2021).

We were interested in establishing whether our duos had chosen to employ a particular coordination strategy because it offered them an advantage in achieving a particular aesthetic or musical outcome in their performance: for example, enabling tighter synchronization with their partner or reducing the overall magnitude of any tempo change. We accomplished this by modeling a series of simulated networked performances in which the coupling patterns between musicians were systematically manipulated yet were otherwise derived from actual performance data obtained from each of our thirteen experimental conditions. Conducting simulations allowed us to have complete control over the coupling between musicians in a way that would not be possible when working with results from the corpus directly. This assisted in interrogating the specific effects of coupling on the objective factors shown to predict subjective evaluations of performance success in Figure 6. In addition, simulations also enabled us to explore alternative coordination strategies that were not displayed by any of the duos in the experiment.

We compared the following simulations across each of the thirteen conditions tested in the experiment: 1) a "democracy" coordination strategy, in which the coupling coefficients for both simulated performers were set equal to the average of all coupling responses obtained for one condition; 2) a "leadership" coordination strategy, in which the simulated pianist was coupled to the drummer to a degree equivalent with the mean pianist-drummer coupling observed for one condition, while drummer-pianist coupling was set to zero; and 3) a baseline "anarchy" coordination strategy, not followed by any duo in the experiment, in which each simulated musician acted independently of the other with all coupling coefficients set to zero.

To ensure consistency across simulations, self-coupling coefficients $\alpha_{i,i}$ were all set to the mean observed for that experimental condition, intercept terms $\alpha_{i,0}$ were set to 0, and the error term ε_k^i was set to 5 ms, found to add sufficient noise to artificial inter-beat intervals without adversely affecting the stability of the simulation. Five hundred individual simulations were conducted for every experimental condition across each of the parameters given above (1,500 simulations per condition, 19,500 simulations total), with tempo slope, ensemble asynchrony, and timing irregularity selected as the criteria for comparing between simulations: earlier in Figure 6, we described how increases in these factors were predictive of comparable decreases in subjective ratings provided by both musicians and listeners. In Figure 9a, we plot the median values obtained for these criteria across simulations conducted for each condition. Marker style and color are used to differentiate conditions with and without latency.

The coordination strategies we tested reveal the trade-off between minimizing both tempo drift and ensemble asynchrony when optimizing coordination in networked performance. Democracy was the best strategy for maximizing ensemble synchronization; however, simulations made using this strategy quickly slowed down and became increasingly unstable, as both simulated musicians matched their performance to each other's delays. Leadership, on the other hand, was the best strategy for minimizing tempo drift; however, simulations made using this strategy displayed substantially lower synchrony than democracy. Finally, while anarchy did lead to regular timing and no global drift in tempo, this came at the expense of unacceptable asynchrony between musicians, who became several seconds out-of-time by the end of the simulation as a result of no adaptation between them.

The simulated performances made under the control conditions were the exceptions to the above analysis as,

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

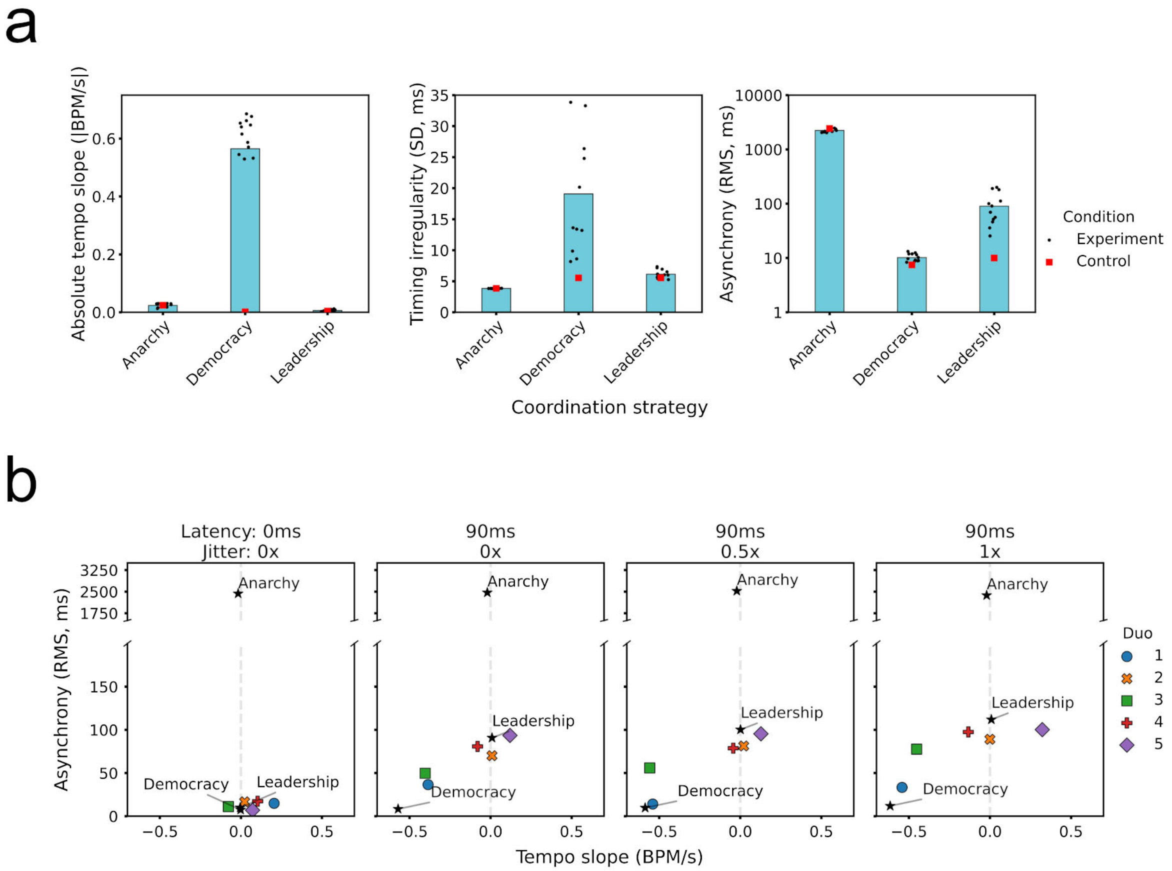

FIGURE 9. Simulation results. (a) Comparison of tempo slope, timing irregularity, asynchrony (note log scale) across simulated coordination strategies: each point shows the median value obtained from 500 individual simulations for one condition in the experiment, with the height of each bar representing the overall mean for that strategy across all conditions. Condition type is given by marker size, shape, and color, with red square markers showing results from control simulations, with no latency applied. (b) Simulated asynchrony and tempo slope coefficients obtained from the control and 90 ms latency conditions (all jitter scalings): duo number and coordination strategy are delineated by the style and color of the marker.

when no latency was applied, both democracy and leadership achieved similar results across all comparison criteria. This indicates that the use of either strategy can be considered optimal within real-time, non-delayed jazz performance, where the capacity of a musician to perceive their partner's performance would not normally be impeded. This may also explain why no duo displayed coordination equivalent to our anarchy coordination strategy in the actual experiment, as following this strategy in a "normal" performance still led to massive asynchronies.

Ultimately, in a networked performance environment, our simulations indicate that it is not possible for an ensemble to find a coordination strategy that achieves both maximum synchronization between the performers and a minimum of global tempo drift. These two parameters exist on opposite sides of a trade-off; a choice must be made to optimize in favor of one feature, with performances suffering in other aspects.

Finally, in Figure 9b we compare the median tempo slope and asynchrony values obtained from 500 simulations conducted using the coordination strategies described above with the same number of simulations using the coupling patterns displayed by the duos studied in our experiment (discussed earlier with relation to our control analyses: see Supplementary Materials Figure S10b). Here we plot the control condition and the 90 ms latency conditions as illustrative examples; see Figure S12 for full plots including the remaining conditions.

The results confirm our assumption that networked performance cannot be optimized fully—no duo achieved minimal asynchrony without also slowing down, for instance. They also validate our description

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

of the coordination strategy employed by each duo as either democratic or leadership-based, insofar as results from simulations using the coupling established by each duo best approximated those obtained from the strategy they were claimed to follow in Figure 8d.

Discussion

The purpose of this study was to identify the possible methods that can be used to coordinate spontaneous group interaction over a network. We collected data from five duos of jazz pianists and drummers improvising music together while variable network latency was applied using a testbed. We identified two coordination strategies from the linear modeling of rhythmic adaptation in their performances. A leadership strategy, where one participant adapted to their partner but the other did not, resulted in a stable tempo but high asynchrony between the performers; a democratic strategy, where both participants adapted to each other at equivalent rates, achieved less asynchrony at the expense of tempo drift. Analysis of subjective performance evaluations indicated that high levels of tempo change and asynchrony were both associated with worse evaluations, as provided by the musicians themselves and a sample of naive listeners blind to the networked conditions.

Our findings demonstrate how remoteness presents new complexities and challenges to successful interaction. While both leadership and democratic coordination can demonstrably achieve good results in real-time performances (Wing et al., 2014), neither strategy can be considered optimal when network latency is present. Rather, ensembles must prioritize either maintaining a steady tempo or achieving low levels of asynchrony in their performance and coordinate their joint action in a manner contingent with achieving that goal.

Our results also highlight the musical qualities that different ensembles value when they perform together, as the participants in our experiment were told only that they should interact with their partner as they would in a "real" performance. One drummer who established mutual co-adaptation (democracy) with their partner described how "plowing on at right tempo didn't really seem like an option," as "cohesiveness [was] prob[ably] more important than tempo accuracy"; their performances "sound[ed] better when we slow down to meet each other," and this even enabled "quite a fun heavy groove when we got the hang of it" (drummer, duo 1). Tempo change was not inherently undesirable, for this ensemble at least, so long as it enhanced synchronization and afforded new possibilities for musical creativity.

Our results demonstrate how the strategies used to coordinate joint action in an ensemble can reflect genre-specific demands in music performance. Across all of the three remaining duos who established asymmetric (leadership) adaptation, it was the drummer who emerged as the most influential performer. This instrument typically has responsibility for maintaining musical time in any jazz ensemble (see Supplementary Materials section §2.1.); when latency is present it disrupts this sense of shared time, so it is perhaps unsurprising that the drummer assumed the role of leader and the pianist yielded this to them. These roles were allocated implicitly and without discussion in all three groups: so, in one sense, the asymmetric relationships they adopted were still "democratic," insofar as the individual roles adopted by each performer were consensually (albeit tacitly) allocated in accordance with genre-specific norms and the demands of the performance context.

Similar concerns to those involved in networked musical performance are at play whenever spoken conversations are coordinated over teleconferencing platforms. Temporal periodicity acts as a pragmatic resource to enhance communication (Rothermich et al., 2012) in speech and to facilitate coordination in both spontaneous musical and speech interaction (Pfänder & Couper-Kuhlen, 2019; Robledo et al., 2021). Musicians improvise simultaneously with each other and coordination becomes a continuous, mutual process; while temporal coordination in much of speech interaction concerns organization of turn transitions between participants in a conversation (Cech & Condon, 2004), it is also evident in the timing of backchannel (interjections and gestures provided by the non-floor holder) in relation to the ongoing flow of the floor-holding speaker's turn (Beňuš et al., 2011; Noguchi et al., 2000).

These temporal features are less likely to be reliably accessible as communicative and pragmatic cues in remote contexts than in face-to-face contexts, often leading to large increases in turn transition time (Boland et al., 2022) and decreasing the effectiveness of backchannel (Fox Tree et al., 2021). To prevent this, groups engaged in teleconferencing can adopt asymmetric roles. This may involve one individual acting as a moderator, leading the conversation by speaking confidently, intervening in discussions, and also by selectively muting and unmuting the microphones of other speakers to facilitate smoother turn transitions. This suggests that "leadership" is likely to be an effective strategy in coordinating timings in transactional or task-oriented communicative interactions.

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

Our results have clear practical implications for the future development of network platforms used for musical performances. It would be feasible for a platform to apply our model in real-time and use this to provide feedback to musicians about predicted changes in their tempo and synchronization levels during a remote performance. This could consist of alerts when their mode of playing together may cause them to decelerate or drift out of time with each other (depending, perhaps, on pre-defined rules), similar to the warning messages currently implemented for unstable internet connections. We suggest that this feature could potentially improve user retention, as prior research has shown that encountering the negative effects of latency can impact willingness to participate in future remote performance (Chew et al., 2005; Driessen et al., 2011).

One limitation of this study concerns sample size. Recruiting professional musicians for experimental research involves an additional financial burden over recruiting amateurs, leading to smaller sample sizes and issues with statistical power. Developing proficiency in musical improvisation, however, takes many years of dedicated training (Berliner, 1994), meaning that the optimal way to research improvisation will always involve the recruitment of highly skilled practitioners, whose performances can then be isolated in an experimental environment. Corpus analyses of interaction in existing recordings would, however, provide a complementary perspective on the dynamics we model here and may be a direction for future research to explore.

A second limitation concerns our choice not to include a bassist in our participant-groups, as this instrument is typically included in the jazz rhythm section, alongside piano and drums (see Supplementary Materials section §2.1. for a description of the role played by this instrument). As a non-fretted stringed instrument, accurately converting the performance of a double bass to MIDI with a degree of latency that is acceptable for real-time music-making is difficult, however. An interesting direction for future research would involve designing a testbed system that uses audio signal processing techniques to simulate variable latency rather than MIDI, enabling the modeling techniques developed here to be applied to larger ensembles.

A third limitation concerns our use of generalist (e.g., Zoom) rather than specialist telecommunications platforms when measuring network latency. Recent technological advances have been able to reduce latencies during networked musical performance to below the minimum threshold tested in this research (e.g., Drioli et al., 2013), with exciting implications for musicians. Latency, however, will always be present to some degree during networked performance, especially for musicians situated far away from each other geographically, so we still consider it necessary to study how it can be accommodated. Future research, however, may involve using these specialist technologies when modeling latency and jitter, unlike our use of a generalist platform.

Taken together, our results provide the first demonstration that error correction, a core component of the human facility for temporal coordination, can be optimized to compensate for the lack of perceived simultaneity that arises when joint action occurs over a network. While face-to-face conversations and musical performances are increasingly becoming feasible as restrictions related to the COVID-19 pandemic ease, remote facilitation is likely to remain an essential part of modern life in the future, meaning that comprehending the ways this may impact successful human interaction has never been more crucial.

Author Note

For the purposes of Open Access, the author has applied a Creative Commons Attribution License (CC-BY) to any Author Accepted Manuscript arising. The design and analysis of the studies reported in this paper were not preregistered. All data and research materials (with no exceptions) are posted under a permissive license to a trusted third-party repository (accessible using the stable link https://doi.org/10.5281/zenodo.7773824). The complete analysis scripts, software, and code book created by the author have also been made publicly accessible to enable readers to replicate the analyses (analysis code: https://github.com/HuwCheston/Jazz-Jitter-Analysis, testbed software: https://github.com/HuwCheston/AV-Manip, online perceptual study software: https://github.com/HuwCheston/2023-duo-success-analysis). The authors declare that they have no conflicts of interest. Funding for this work was provided by a project incubation award from Cambridge Digital Humanities, project page https://www.cdh.cam.ac.uk/research/projects/newmusicsoftwareplatform/.

Author contributions are as follows—Huw Cheston (hwc31@cam.ac.uk): conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, funding acquisition, writing – original draft, writing – review & editing; Ian Cross (ic108@cam.ac.uk): conceptualization, writing – review & editing, supervision; Peter Harrison (pmch2@cam.ac.uk): conceptualization, software, data curation, funding acquisition, writing – review & editing, supervision.

Correspondence concerning this article should be addressed to Huw Cheston (hwc31@cam.ac.uk).

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

References

Aagaard, J. (2022). On the dynamics of Zoom fatigue. *Convergence: The International Journal of Research into New Media Technologies*, *28*(6), 1878–1891. https://doi.org/10.1177/13548565221099711

Bartlette, C., Headlam, D., Bocko, M., & Velikic, G. (2006). Effect of network latency on interactive musical performance. *Music Perception*, *24*(1), 49–62. https://doi.org/10.1525/mp.2006.24.1.49

Beňuš, Š., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, *43*(12), 3001–3027. https://doi.org/10.1016/j.pragma.2011.05.011

Berliner, P. (1994). *Thinking in jazz: The infinite art of improvisation*. University of Chicago Press.

Boland, J. E., Fonseca, P., Mermelstein, I., & Williamson, M. (2022). Zoom disrupts the rhythm of conversation. *Journal of Experimental Psychology: General*, *151*(6), 1272–1282. https://doi.org/10.1037/xge0001150

Cáceres, J.-P., & Chafe, C. (2010). JackTrip: Under the hood of an engine for network audio. *Journal of New Music Research*, *39*(3), 183–187. https://doi.org/10.1080/09298215.2010.481361

Carôt, A., & Werner, C. (2009). Fundamentals and principles of musical telepresence. *Journal of Science and Technology of the Arts*, 26–37. https://doi.org/10.7559/CITARJ.V1I1.6

Cech, C. G., & Condon, S. L. (2004). Temporal properties of turn-taking and turn-packaging in synchronous computer-mediated communication. *Proceedings of The 37th Annual Hawaii International Conference on System Sciences, 2004*. 10 pp. https://doi.org/10.1109/HICSS.2004.1265282

Chafe, C., Cáceres, J.-P., & Gurevich, M. (2010). Effect of temporal separation on synchronization in rhythmic performance. *Perception*, *39*(7), 982–992. https://doi.org/10.1068/p6465

Cheston, H. (2022). 'Turning the beat around': Time, temporality, and participation in the jazz solo break. *Conference on Interdisciplinary Musicology 2022 Proceedings*, Edinburgh, United Kingdom.

Chew, E., Zimmermann, R., Sawchuk, A. A., Papadopoulos, C., Kyriakakis, C., Tanoue, C., et al. (2005). A second report on the user experiments in the distributed immersive performance project. *5th Open Workshop of MUSICNETWORK: Integration of Music in Multimedia Applications*, 8.

Clayton, M., Jakubowski, K., Eerola, T., Keller, P. E., Camurri, A., Volpe, G., & Alborno, P. (2020). Interpersonal entrainment in music performance. *Music Perception*, *38*(2), 136–194. https://doi.org/10.1525/mp.2020.38.2.136

Demos, A. P., & Palmer, C. (2023). Social and nonlinear dynamics unite: Musical group synchrony. *Trends in Cognitive Sciences*, S1364661323001225. https://doi.org/10.1016/j.tics.2023.05.005

Doffman, M. (2014). Temporality, awareness, and the feeling of entrainment in jazz performance. In M. Clayton, B. Dueck, & L. Leante (Eds.), *Experience and meaning in music performance* (pp. 62–85). Oxford University Press.

Doherty-Sneddon, G., O'Malley, C., Garrod, S., Anderson, A., & Langton, S. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, *3*(2), 105–125. https://doi.org/10.1037/1076-898X.3.2.105

Driessen, P. F., Darcie, T. E., & Pillay, B. (2011). The effects of network delay on tempo in musical performance. *Computer Music Journal*, *35*(1), 76–89. https://doi.org/10.1162/COMJ_a_00041

Drioli, C., Allochio, C., & Buso, N. (2013). Networked performances and natural interaction via LOLA: Low latency high quality A/V streaming system. In P. Nesi & R. Santucci (Eds.), *Information technologies for performing arts, media access, and entertainment: Second International Conference, ECLAP 2013, Revised Selected Papers* (*Vol. 7990*). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40050-6

Fox Tree, J. E., Whittaker, S., Herring, S. C., Chowdhury, Y., Nguyen, A., & Takayama, L. (2021). Psychological distance in mobile telepresence. *International Journal of Human-Computer Studies*, *151*, 102629. https://doi.org/10.1016/j.ijhcs.2021.102629

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11. https://doi.org/10.1016/j.tics.2003.10.016

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*(5), 427–438. https://doi.org/10.1525/mp.2009.26.5.427

Grant, K. W., van Wassenhove, V., & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. *Speech Communication*, *44*(1–4), 43–53. https://doi.org/10.1016/j.specom.2004.06.004

Harrison, P. M. C., Marjieh, R., Adolfi, F., van Rijn, P., Anglada-Tort, M., Tchernichovski, O., et al. (2020). *Gibbs sampling with people* (arXiv:2008.02595). arXiv. http://arxiv.org/abs/2008.02595

Holub, J., Kastner, M., & Tomiska, O. (2007). *Delay effect on conversational quality in telecommunication networks: Do we mind?* 2007 Wireless Telecommunications Symposium, Pomona, CA. https://doi.org/10.1109/WTS.2007.4563311

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

Jacoby, N., Polak, R., & London, J. (2021). Extreme precision in rhythmic interaction is enabled by role-optimized sensorimotor coupling: analysis and modelling of West African drum ensemble music. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1835), 20200331. https://doi.org/10.1098/rstb.2020.0331

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Keil, C. (1987). Participatory discrepancies and the power of music. *Cultural Anthropology*, *2*(3), 275–283.

Kello, C. T., Bella, S. D., Mede, B., & Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of The Royal Society Interface*, *14*(135), 20170231. https://doi.org/10.1098/rsif.2017.0231

Kilchenmann, L., & Senn, O. (2015). Microtiming in swing and funk affects the body movement behavior of music expert listeners. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01232

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping. *Quarterly Journal of Experimental Psychology*, *63*(11), 2220–2230. https://doi.org/10.1080/17470218.2010.497843

London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford University Press.

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in Python. *Proceedings of the 14th Python in Science Conference*, 18–24. https://doi.org/10.25080/Majora-7b98e3ed-003

Monache, S. D., Buccoli, M., Comanducci, L., Sarti, A., Cospito, G., Pietrocola, E., & Berbenni, F. (2019). Time is not on my side: Network latency, presence and performance in remote music interaction. *Proceedings of the XXII CIM Colloquium on Music Informatics*, *8*.

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

Noguchi, H., Katagiri, Y., & Den, Y. (2000). An experimental verification of the prosodic/lexical effects on the occurrence of backchannels. *6th International Conference on Spoken Language Processing (ICSLP 2000)*, (*Vol. 2*), 628-631–0. https://doi.org/10.21437/ICSLP.2000-347

Nowicki, L., Prinz, W., Grosjean, M., Repp, B. H., & Keller, P. E. (2013). Mutual adaptive timing in interpersonal action coordination. *Psychomusicology: Music, Mind, and Brain*, *23*(1), 6–20. https://doi.org/10.1037/a0032039

Olmos, A., Brulé, M., Bouillot, N., Benovoy, M., Blum, J., Sun, H., et al. (2009). Exploring the role of latency and orchestra placement on the networked performance of a distributed opera. *12th Annual International Workshop on Presence*, 10.

Pfänder, S., & Couper-Kuhlen, E. (2019). Turn-sharing revisited: An exploration of simultaneous speech in interactions between couples. *Journal of Pragmatics*, *147*, 22–48. https://doi.org/10.1016/j.pragma.2019.05.010

Pras, A., Schober, M. F., & Spiro, N. (2017). What about their performance do free jazz improvisers agree upon? A case study. *Frontiers in Psychology*, *8*, 966. https://doi.org/10.3389/fpsyg.2017.00966

Robledo, J. P., Hawkins, S., Cornejo, C., Cross, I., Party, D., & Hurtado, E. (2021). Musical improvisation enhances interpersonal coordination in subsequent conversation: Motor and speech evidence. *PLOS ONE*, *16*(4), e0250166. https://doi.org/10.1371/journal.pone.0250166

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, *50*(2), 232–244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Rottondi, C., Buccoli, M., & Zanoni, M. (2015). Feature-based analysis of the effects of packet delay on networked musical interactions. *Journal of the Audio Engineering Society*, *63*(11), 864–875. https://doi.org/10.17743/jaes.2015.0074

Rottondi, C., Chafe, C., Allocchio, C., & Sarti, A. (2016). An overview on networked music performance technologies. *IEEE Access*, *4*, 8823–8843. https://doi.org/10.1109/ACCESS.2016.2628440

Schober, M. F., & Spiro, N. (2014). Jazz improvisers' shared understanding: A case study. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00808

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*, Austin, Texas.

Setzler, M., & Goldstone, R. (2020). Coordination and consonance between interacting, improvising musicians. *Open Mind*, *4*, 88–101. https://doi.org/10.1162/opmi_a_00036

Timmers, R., Endo, S., Bradbury, A., & Wing, A. M. (2014). Synchronization and leadership in string quartet performance: A case study of auditory and visual cues. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00645

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025

Vorberg, D. (2005). Synchronization in duet performance: Testing the two-person phase error correction model. *Tenth Rhythm Perception and Production Workshop*, Alden Biesen, Belgium.

Vorberg, D., & Wing, A. M. (1996). Modeling variability and dependence in timing. In H. Heuer (Ed.), *Handbook of perception and action. 2: Motor skills*. Academic Press.

Wing, A. M., Endo, S., Bradbury, A., & Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *Journal of The Royal Society Interface*, *11*(93), 20131125. https://doi.org/10.1098/rsif.2013.1125

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Downloaded from http://online.ucpress.edu/mp/article-pdf/42/1/48/833292/mp.2024.42.1.48.pdf by University of Cambridge user on 28 March 2025