

---

# Effect of temporal separation on synchronization in rhythmic performance

---

Chris Chafe, Juan-Pablo Cáceres, Michael Gurevich§

Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Stanford, CA 94305, USA; e-mail: cc@ccrma.stanford.edu; §Sonic Arts Research Centre (SARC), Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK

Received 14 May 2009, in revised form 18 April 2010

---

**Abstract.** A variety of short time delays inserted between pairs of subjects were found to affect their ability to synchronize a musical task. The subjects performed a clapping rhythm together from separate sound-isolated rooms via headphones and without visual contact. One-way time delays between pairs were manipulated electronically in the range of 3 to 78 ms. We are interested in quantifying the envelope of time delay within which two individuals produce synchronous performances. The results indicate that there are distinct regimes of mutually coupled behavior, and that ‘natural time delay’—delay within the narrow range associated with travel times across spatial arrangements of groups and ensembles—supports the most stable performance. Conditions outside of this envelope, with time delays both below and above it, create characteristic interaction dynamics in the mutually coupled actions of the duo. Trials at extremely short delays (corresponding to unnaturally close proximity) had a tendency to accelerate from anticipation. Synchronization lagged at longer delays (larger than usual physical distances) and produced an increasingly severe deceleration and then deterioration of performed rhythms. The study has implications for music collaboration over the Internet and suggests that stable rhythmic performance can be achieved by ‘wired ensembles’ across distances of thousands of kilometers.

## 1 Introduction

*Temporal separation* refers to the time it takes for the actions of one person to reach another while acting together. If the acts are aural in nature—music or speech—then time delay between the actors is a function of the speed of sound in the medium and the distance between them. From speech telecommunications literature concerned with *turn-taking interaction*, we know that conversation is possible even with one-way delays of up to 500 ms (Holub et al 2007). In contrast, for *synchronous rhythmic interaction*, it is the ability to simultaneously share, hear, and ‘feel’ the beat that counts. This is an aspect of musical interaction that places a much greater restriction on the range of acceptable time delays and has been a source of frustration for musicians attempting to use telecommunication media usually intended for voice. “How much delay is too much?” is a common question asked by performers who are increasingly using the Internet for real-time audio collaboration.<sup>(1)</sup>

The physical settings for playing music always impose a certain amount of temporal separation. A likely spacing between the outer members of a string trio, quartet, or quintet, lies within the range of 2 to 3 m or approximately 6 to 9 ms one-way delay

<sup>(1)</sup>The Internet presents intriguing possibilities for high-quality interaction but involves a wide range of time delays (Kapur et al 2005). A dramatic decrease in telecommunication delays happened in the early 2000s, when research groups including Stanford University and McGill University began testing IP network protocols for professional audio use, seeking methods for bi-directional WAN music collaboration. Long-distance acoustic delays were now closer to room-sized acoustic delays and ensemble performances began to feel acceptable. The new capability used computer systems which exchanged uncompressed audio through high-speed links like Internet2, Canarie, and Geant2 (significantly higher resolution and faster transmission than for standard digital voice communication media like telephone, VoIP, Skype, etc).

[given the usual semicircular arrangement and that the speed of sound is approximately  $3 \text{ ms m}^{-1}$  (Benade 1990)]. So, imagine the scenario encountered by two musicians trying to play synchronously at a distance five times greater, separated by 45 ms delay (they would be approximately 15 m apart).<sup>(2)</sup> In the simplest sense, player A is waiting for the sound of player B, who is waiting for the sound of player A, and the tempo slows down from this recursion.

By manipulating time delays experimentally, between pairs of subjects clapping together but in separate rooms, we previously observed a relationship between temporal separation and tempo (Chafe and Gurevich 2004). By analyzing the same data set, the rhythmic interaction dynamics can now be described. Different synchronization regimes and delay-coping-strategies come into play across the ‘delay-scape’ studied.

### 1.1 *Quantifying synchronization in rhythmic performance*

Micro-timing differences between seemingly well-synchronized players have been measured with near-millisecond accuracy in studies of instrumental performance. Asynchronization of a pair of voices is “the standard deviation of the onset time differences of simultaneous tones of those voice parts” (Rasch 1988, page 73). Instrumental trio performances (which were analyzed in terms of 3 pairs) showed a range of approximately 30 to 50 ms. Greater asynchronization was correlated with different levels of temporal separation for repeated performances by instrumental duos (Bartlette et al 2006).<sup>(3)</sup> An increase in asynchronization from 30 to over 200 ms for the delay range (6 to 206 ms) was measured and the results also depended on the choice of music, tempo, and instrument. Hand-clapping experiments, including the present work, have also been used to observe a rise of asynchronicity with delay. However, asynchronization has been lower (and upper-end delays lower), from 12 to 23 ms (for delays of 6 to 68 ms) (Farner et al 2009)<sup>(4)</sup> and 10 to 20 ms here (for delays from 3 to 78 ms).

The mean of the onset-time differences was a magnitude (absolute value) in the two delay studies cited. Our approach (and the earlier baseline performance study—Rasch 1988) has kept the sign of the difference in order to observe the lead/lag of one performer’s note onset with respect to another’s. This allows the analysis to observe micro-timing regimes which underlie tempo change.

## 2 Experiment

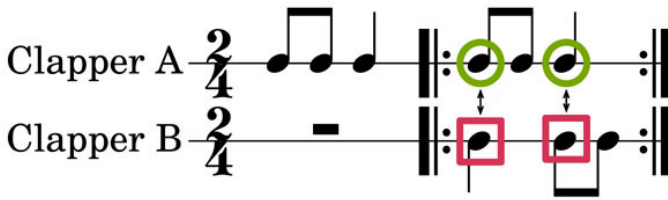
We examined performances by pairs of clappers under different delay conditions. A simple interlocking rhythmic pattern was chosen as the task (figure 1). The pattern had three properties which were conducive for the experiment: first, it comprised independent but equal parts rather than unison clapping (a kind of simple polyphony); second, it created a context free of ‘internal’ musical effects (Bartlette et al 2006); and third, the rhythm could be analyzed for lead/lag (the metrical structure’s phase advance could be individually monitored per part). The duo rhythm was easily mastered by a pool of subjects who were not selected for any particular musical ability.

Subjects were seated apart in separate studios and monitored each other’s sound with headphones (with no visual contact). 11 delay conditions in the range from  $d = 3$  to 78 ms (one-way) were introduced in the sound path (electronically) and were randomly varied per trial. The shortest delay,  $d = 3$  ms, is equivalent to having a subject clapping 1 m from the other’s ears. The longest delay,  $d = 78$  ms, corresponds to a separation of approximately 26 m, equivalent to a distance wider than many concert stages.

<sup>(2)</sup> Delay of approximately 45 ms is also what we encounter (Cáceres and Chafe 2010) between San Francisco and New York when transmitting uncompressed audio over the Internet2 network (<http://www.internet2.edu/about/>).

<sup>(3)</sup> Asynchronization, but in Bartlette et al (2006) it is called ‘coordination’.

<sup>(4)</sup> Asynchronization, but in Farner et al (2009) it is called ‘SD of lead’.



**Figure 1.** [In color online, see <http://dx.doi.org/10.1068/p6465>] Duo clapping rhythm used to test the effect of temporal separation. Subjects in separate rooms were asked to clap the rhythm together while hearing each other's sound delayed by a slight amount. Common beats in the duo clapping rhythm provide reference points for analysis of ensemble synchronization. Circles and squares represent synchronization points.

Recordings were processed automatically with an event-detection algorithm ahead of further processing to extract synchronization information.

A control trial was inserted at the end of each session in which the electronic delay was bypassed. The delay in this condition consisted only of the air delay from hand clap to microphone,  $d = 1$  ms.

## 2.1 Method

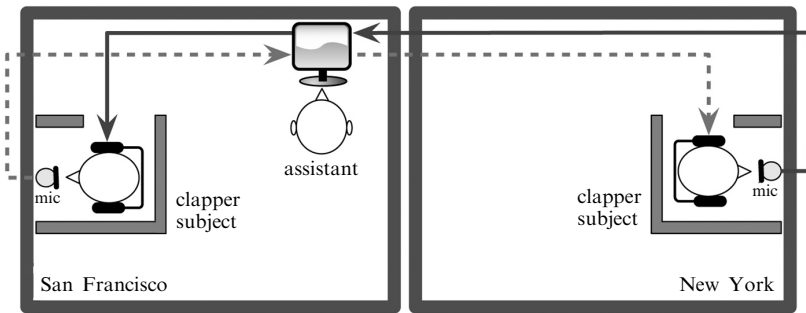
**2.1.1 Trials and control.** One-way delay was fixed to a constant value during a trial and applied to both paths. Delay was varied in 11 steps according to the sequence  $d_n = n + 1 + d_{n-1}$  which produces the set:

$$d_0 = 1; d_1 - d_{11} = \{3, 6, 10, 15, 21, 28, 36, 45, 55, 66, 78\} \text{ ms} .$$

The sequence was chosen in order to weight the distribution towards the low-delay region and gradually lengthen in the higher region, but it bears no special significance otherwise. Delays were presented in random order and each duo performed each condition once. Starting tempo in each trial was also randomly selected from one of three pre-recorded 'metronome' tracks of clapped beats at 86, 90, and 94 beats per minute (bpm). (Other pilot trials, not analyzed as part of the present experiment, were presented inside the random sequence block: 2 for diverse tempi, and 2 for asymmetric delays. Sessions began with one subject-against-recorded-track which also ran at the end of the block, also not included.) A final 1 ms trial using analog bypass mode was included as a control. The bypass was designed to obtain the lowest possible delay. Overall, one session took about 25 min to complete.

**2.1.2 Number of subject pairs and trials.** Twenty-four pairs of subjects participated in the experiment. Subjects were students and staff at Stanford University. A portion of the group was paid with gift certificates and others participated as part of a course in computer music. All subjects gave their informed consent according to Stanford University IRB policy. No subjects were excluded in advance. Individuals in the pool were paired up randomly into duos. Each duo performed all 11 conditions plus the control, once each.

**2.1.3 Acoustical and electronic conditions.** Acoustical conditions minimized room reverberation effects and extraneous sounds (jewelry, chair noise, etc). Subjects were located in two sound-isolated rooms (CCRMA's recording and control room pair whose adjustable walls were configured for greatest sound absorption). They were additionally surrounded by movable sound-absorbing partitions (figure 2). One microphone (Schoeps BLM3) was located 0.3 m in front of each chair. Its monaural signal fed both sides of the opposite subject's headphones. Isolating headphones, Sennheiser HD280 pro, were chosen to reduce headphone leakage to microphones. Wearers of glasses were required to remove their frames to enhance the seal. Volume levels were adjusted for users' comfort and ease of clapping. Direct sound was heard by leakage. The distance from



**Figure 2.** [In color online.] Floor plan. Rooms were acoustically and visually isolated and room reflections were minimized with sound-absorbing panels. Electronic delay from the microphones to headphones was manipulated by computer.

clapping hands to microphone introduced a time delay of about 1 ms and is added into our reported delays. In other words, our reported 3 ms delay comprises respectively, 1 ms + 2 ms, air and electronic delays.

A single computer provided recording, playback, adjustable delays and the automated experimental protocol with GUI-based operation. The setup comprised a Linux PC with 96 kHz audio interface (M-Audio PCI Delta 66, Omni I/O). Custom software was written in C++ using the STK<sup>(5)</sup> set of open-source audio processing classes which were interfaced to the Jack<sup>(6)</sup> real-time audio subsystem. All delays were confirmed with analog oscilloscope measurement. Each trial was recorded as a stereo, 16 bit, 96 kHz sound file. The direct microphone signals from both rooms were synchronously captured to the two channels.

**2.1.4 Protocol.** Two assistants provided an instruction sheet and read it aloud. Subjects could read the notated rhythm from the handout and listen to the assistants demonstrate it. New duos first practiced face-to-face. They were told their task was to “keep the rhythm going evenly” but they were not given a strategy nor any hints to help make that happen. After they felt comfortable clapping the rhythm together, they were assigned to adjacent rooms designated ‘San Francisco’ and ‘New York’.

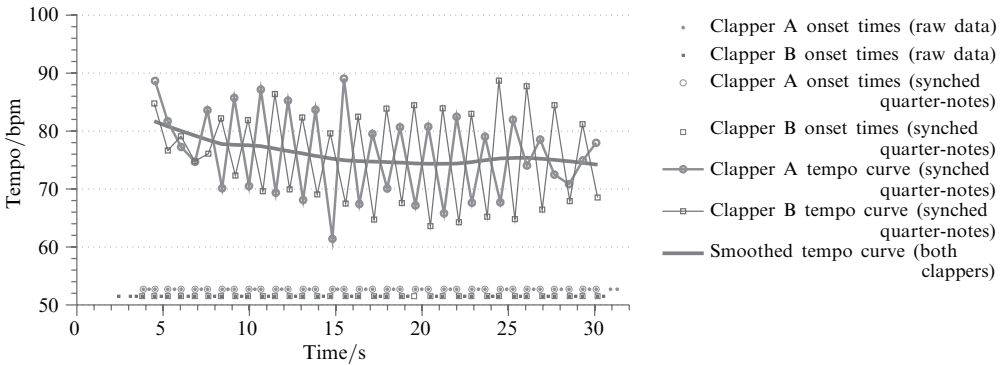
The presentation was computer-controlled. Each time a new trial began, one subject was randomly chosen by the protocol program to begin the clapping (that subject is henceforth referred to as the initiator). His/her starting tempo was established by playback of a short clip (6 quarter-note claps) recorded at the target tempo. 3 starting tempi were used in random order (86, 90, 94 bpm) in order to avoid effects of over-training to one absolute tempo. Trials proceeded in the following steps:

- (i) Room-to-room audio monitoring switches on.
- (ii) A voice recording (saying “San Francisco” or “New York”) plays only to the respective initiator, to cue him/her up.
- (iii) A recording of clapped beats at the new tempo (functioning as a metronome) plays for 6 beats only to the initiator.
- (iv) The initiator starts rhythm at will. The other subject has heard nothing until the point when he/she hears the initiator begin to clap.
- (v) The other joins in at will.
- (vi) After a total of 36 s, the room-to-room monitoring shuts off, ie communication is cut, signaling the end of the trial.

Assistants advanced the sequence of trials manually after each take was completed. Short breaks were allowed and a retake was made if a trial was interrupted.

<sup>(5)</sup> <http://ccrma.stanford.edu/software/stk/>

<sup>(6)</sup> <http://jackaudio.org/>



**Figure 3.** [In color online.] Onset times, synchronization points and tempo curves for one trial [duo number 10, delay 66 ms, starting tempo 94 beats per minute (bpm)]. A smoothed tempo curve is derived from the instantaneous tempi of both player's synchronized events.

## 2.2 Processing of recordings

**2.2.1 Recorded segments of interest.** We were interested only in the sections of the recordings in which both clappers were performing together. Since the protocol allowed the initiator to clap solo for a variable length of time before the second one joined, we first identified the region in which both clappers were involved. For the trial shown in figure 3, clapper B (squares) starts the trial and is followed by clapper A (circles). Enclosed (circles and squares) notes correspond to the common beats which were automatically identified in a first pass on the raw data.

**2.2.2 Event detection.** An automated procedure detected and time-stamped true claps. Detection proceeded per subject (one audio channel at a time).

Candidate events were detected by the 'amplitude surfboard' technique (Schloss 1985) tuned to measure onsets with an accuracy of  $\pm 0.25$  ms. The extremely clean clapping recordings allowed false events (usually spurious subject noises) to be rejected by simple amplitude thresholding. A single threshold coefficient proved suitable for the entire group of sessions. The algorithm first found an amplitude envelope by recording the maximum dB amplitude in successive 50-sample windows, while preserving the sample index of each envelope point. A 7-point linear regression (the 'surfboard') estimated the slope at every envelope sample. Samples with high slope were likely to be event onsets. Candidate events were local maxima in the vicinity of samples with slopes that fell within some threshold of the maximum slope. In the event of several candidates in close proximity, the one with the highest amplitude was chosen. After an event was identified, there was a refractory period, during which another could not occur.

**2.2.3 Validation.** Recordings were automatically examined and only validated for inclusion in further analysis if they passed several automatic tests. 95 trials contained more than one missing event per clapper and were discarded. If only one event was missing, it was automatically fixed through interpolation. 4 trials were shorter than our minimum-length requirement (16 beats, which was 3 SD less than the mean length). If a duo failed to keep the offset relationship of the rhythm, that trial was discarded. If a duo did not satisfactorily perform the control trial, the entire session was discarded. Three duos did not pass. A total of 168 trial recordings were validated for further analysis.

**2.2.4 Event labeling, tempo determination.** Inter-onset intervals (IOIs) were calculated from the event onset times. Conversion from IOI to tempo in bpm (by combining two eighth-notes into one quarter-note beat) was ambiguous in the presence of severe deceleration and required that very slow eighth-notes be distinguished from quarter-notes.

Since only eighth-notes and quarter-notes were present, the IOIs were clustered into two separate groups by using the  $k$ -means clustering algorithm (Bishop 2007). The group of notes clustered with the shortest IOI was identified as eighth-notes and the one with the longest as quarter-notes. Conversion to tempo was computed with:

$$\text{tempo}_{\text{quarter-note}} = \frac{60}{\text{IOI}} \text{ bpm} ,$$

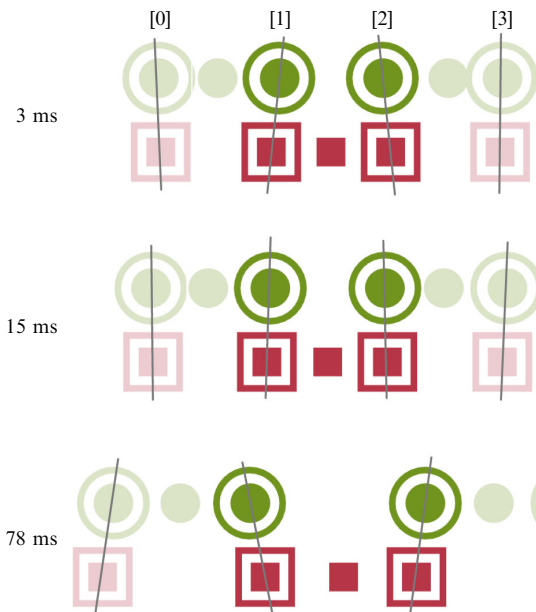
$$\text{tempo}_{\text{eighth-note}} = \frac{60}{2 \times \text{IOI}} \text{ bpm} .$$

**2.2.5 Effect of starting tempo.** ANOVA and multiple comparisons of the mean tempo at each of the three starting tempi (86, 90, 94 bpm) revealed no significant difference between these cases ruling out a dependence on absolute tempo. Data for all trials were shifted (proportionally) after event detection and labeling phases to a starting tempo of 90 bpm before further analysis.

**2.2.6 Database.** Figure 3 presents the results for one trial. The example shows raw onset times, common beat synchronization points, instantaneous tempo of each event in both clappers, and a smoothed common tempo curve. Figure 5 groups smoothed tempo curves for each condition (including the control). Data for the full set of trials are available online<sup>(7)</sup> for continuing analysis. The site also offers the algorithm code for the present analysis.

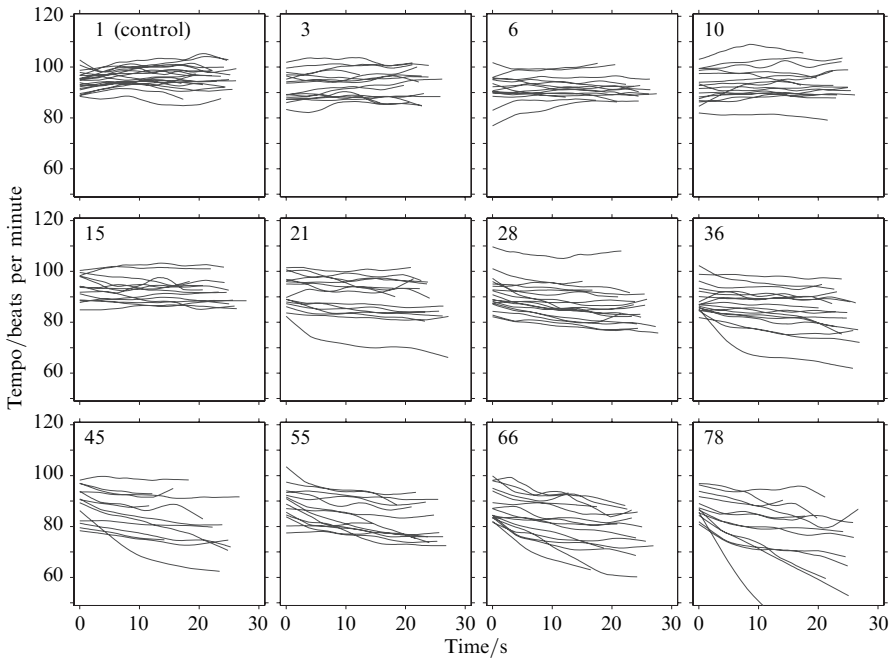
### 2.3 Synchronicity analysis

**2.3.1 Synchronization points.** The assigned rhythm in figure 1 creates points at which claps should be simultaneous, also highlighted by circles and squares in figure 4. Disparities at these synchronization points were calculated to show the amount of anticipation (lead) or lateness (lag) of each player's enclosed (circles and squares) event with respect to the other's.



**Figure 4.** [In color online.] Lead/lag at different delays. Clapper in ‘San Francisco’ is green circles, clapper in ‘New York’ is red squares. Ideally, each vertically adjacent pair of events is simultaneous. Leading or lagging by one subject with respect to the other at these points is related to delay: leading at 3 ms; approximately synchronous at 15 ms; lagging at 78 ms. Lead/lag is measured with respect to measure-length periodicity. Odd-numbered events have inverted (antiphase) sign.

<sup>(7)</sup> <http://ccrma.stanford.edu/groups/soundwire/research/temporal-separation-article-som/>



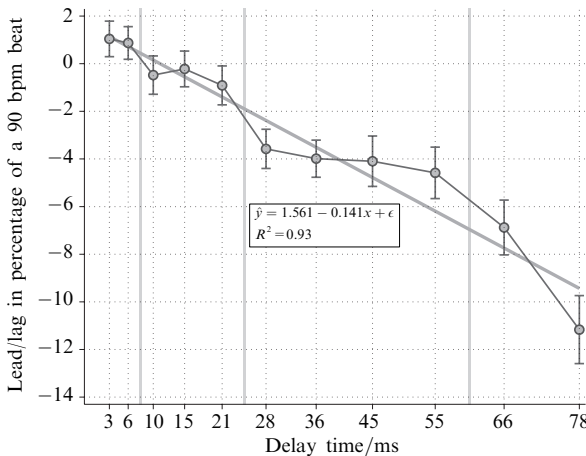
**Figure 5.** All trials' tempo curves grouped by delay. Tempo acceleration during a given performance is tracked by measuring inter-onset intervals as shown in figure 3. The delays (in ms) are shown in top left corner of each graph.

For the examples represented in figure 4, the lead/lag factor was computed as follows:

$$\text{Lead/lag} = (a_{\text{sync}}[1] - b_{\text{sync}}[1]) + (b_{\text{sync}}[2] - a_{\text{sync}}[2]) \quad (1)$$

where  $a_{\text{sync}}[n]$  are sync points (circles) (clapper A) and  $b_{\text{sync}}[n]$  are squares sync points (clapper B).

This differs from previous studies in which the absolute value of asynchronization was measured. The sign of the quantity is preserved in order to observe changing interaction dynamics. For each delay condition, the analysis produced a mean lead/lag value that aggregates all trials, all synchronization points and each player with respect to his/her partner. Figure 6 compares these means and their variances (95% error bars).



**Figure 6.** Onset asynchrony measured at all beat points for the set of delay conditions. At very small delays, performances are dominated by a tendency to lead (positive values). Increasing delay traverses two 'plateaus': first is the region with best synchronization, followed by a second plateau beginning at 28 ms delay. At the greatest delays, lag increases dramatically (negative values).

### 3 Discussion

#### 3.1 Role-based lead or lag

The experiment tested identical musical roles playing identical musical parts. One possible confound to this symmetry is that the initiator who establishes the tempo by clapping first (see section 2.1.4) may have assumed an unintended musical role as leader. In Rasch (1988), instrumental trios had varying degrees of role differentiation which depended on the type of music (homophonic, polyphonic). For example, a string trio played compositions that established their musical roles as melody, inner voice, bass. Relative lead/lag differed between the roles. Melodic parts led, bass was second, and inner voices lagged. The study postulates that this is likely to be a property of performance of homophonic compositions. Recorder performances of polyphonic trios were also analyzed and, since the more nearly equal roles of Early Music tend to be supported by the bass, the bass led.

Roles were randomized in successive trials so that on average a given subject would be equally likely to be ‘initiator’ or ‘follower’. If being initiator induced a role that would affect relative lead, that aspect is equally distributed between subjects and any effect would be equally distributed across conditions. The question whether the protocol created any difference between initiator and follower can be examined by comparing individual trials, but is not studied here.

#### 3.2 Regimes

Clapping together functions differently across the ‘delay-scape’ studied. Our pilot work (Schuett 2002) postulated two qualitatively different regimes: ‘true ensemble’ performance and a delay coping strategy of ‘leader/follower’. The former broke down at delays within the range 20 to 40 ms (as indicated by a rise in a-isochronization at the beat level). When the latter strategy was explicitly engaged, the breakdown threshold increased to somewhere in the range of 50 to 70 ms. A study replicating the task (Farner et al 2009) also noted a first threshold of 25 ms [after which a-isochronization<sup>(8)</sup> at the measure-level increased], followed by a second threshold in the range of 35 to 50 ms [after which the magnitude of note onset timing differences<sup>(9)</sup> at the measure level increased]. Four regimes in the lead/lag analysis can be identified in figure 6 and are summarized in table 1, where equivalent air and network distance delays are also listed.

**Table 1.** Clapping regimes, actual sampled delays, and interpolated transition values (bold in parentheses). Delays are grouped by lead/lag level.

Regime	Delay/ms	Air equivalent/m	Net equivalent/km	Effect
1	3, 6 ( <b>8</b> )	<3	<500	acceleration
2	10, 15, 21 ( <b>25</b> )	8	1700	‘natural’
3	28, 36, 45, 55 ( <b>60</b> )	20	4000	deceleration
4	66, 78	>20	>4000	deterioration

3.2.1 *Shortest delays: Tendency to anticipate, acceleration (0 to 8 ms).* The clapping studies have identified a regime of tempo acceleration at very low delays [at 1, 3, 6, 10 ms in Chafe et al (2004) and 6 ms in Farner et al (2009)]. A linear model here in terms of lead/lag versus delay,

$$\hat{y} = 1.561 - 0.141x + \epsilon, \quad (2)$$

shows zero lead/lag occurring just above 8 ms ( $y$ -intercept in figure 6). It can be concluded that there exists an intrinsic tendency to anticipate and that this amount of delay is required to balance it out. See Repp (2005) for a review of negative mean

<sup>(8)</sup> A-isochronization, but in Farner et al (2009) it is called ‘imprecision’.

<sup>(9)</sup> Magnitude of note onset timing differences, but in Farner et al (2009) it is called ‘mean lead’.



asynchrony (NMA) that has been the subject of many tapping (with metronome) studies. The finding that “the NMA is thus a phenomenon peculiar to nonmusicians tapping in synchrony with a simple metronome” (page 973) should be re-evaluated in light of its possible apparent existence in mutually coupled behavior (Pikovsky et al 2003).

3.2.2 *‘Natural delays’: Best synchronicity, first plateau, stable tempo (8 to 25 ms)*. Synchronicity is best when a pair of clappers can mesh their rhythms without interference from delay. Each clapping subject is its own oscillator but is mutually coupled to the other. The two together form a more complex system with interaction dynamics which can remain stable across this range of delays. Even with a threefold increase in delay, the regime is characterized by constant, minimum lead/lag. Zero crossings for linear regressions of tempo acceleration are in this region (Chafe et al 2004; Farner et al 2009). Again, from the world of music performance, delay of the same order would be created by the spacing of musicians gathered in comfortably close proximity.

3.2.3 *Challenging delays—‘strategize or decelerate’: Second plateau, deceleration, and mitigating strategies (25 to 60 ms)*. The amount of clapping deceleration continues its monotonic increase with delay. However, lead/lag drops to a new stable value in this region which can be explained by a change to the mutually coupled interaction dynamics.

It has been hypothesized that strategies will consciously or unconsciously be engaged at higher delays (Farner et al 2009). Conscious strategies include intentionally leading by pushing the beat, or leading by ignoring the sound of part of the ensemble (in which case the ‘detached’ actors must follow). Either strategy has the effect of eliminating the recursion that comes with higher delay. Strategies can be conscious and imposed (Bartlette et al 2006) or evanescent (short-lived, combined, and/or distributed between actors). No explicit strategy was used in this study.

It can be further hypothesized that leader/follower strategy is employed when there is an imbalance in the structure of the ensemble. This idea comes from experiences in which the authors have participated in performing over a range of distances and delays, and with a wide variety of music. The weaker side (in terms of rhythmic function) naturally follows the strong one (whose rhythmic role, instrument type, and/or number of players dominate). For example, a guitarist playing with a drummer in this regime will tend to follow. An experimental Internet performance of the first movement of the Mozart G-minor string quintet (K516) (with the St Lawrence string quartet in Banff, Alberta plus a former member playing second viola in Stanford, California) found that a separation of approximately 30 ms (25 + 5/network + air) required a leader/follower strategy, otherwise it introduced perceptible variance in fast rhythmic passages. A counter-strategy, when two of the players consciously let the lag accumulate, promoted an effortless ritardando (intentional deceleration). In this sound clip, one hears that as the tempo slows it eventually settles on a point where stability is achieved (at a much slower tempo).<sup>(10)</sup>

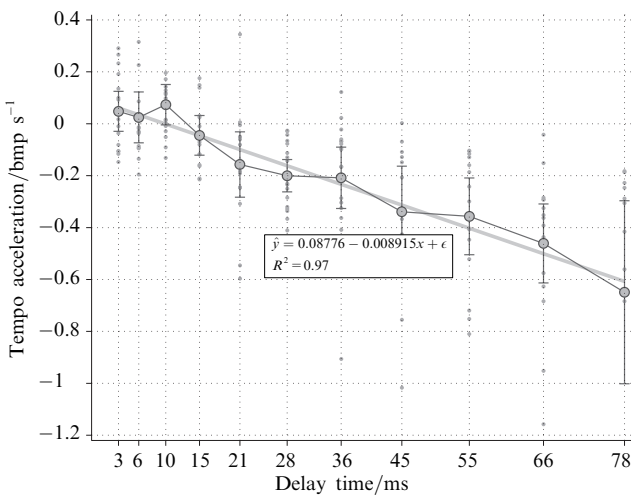
3.2.4 *Conditions above challenging delays: Deterioration, where playing accuracy rapidly falls off*. The ‘edge of playability’ is reached when strategies no longer suffice to maintain a mutually coupled regime of any sort, and for this experiment it lies between delays at 55 and 66 ms. Beyond this edge is a regime with sharply increasing lag and asymmetry. This limit is in agreement with Farner et al (2009), but lower than 86 ms which was still ranked with ‘high musicality’ by duo performers (Bartlette et al 2006). This discrepancy, between our ‘edge of playability’ and the higher limit for music played by instrumental duos, could be explained by differences in the task. Music has larger temporal structures than the ostinato rhythm of our clapping task. Our suspicion is that

<sup>(10)</sup> Recordings of the experiment are available online at <http://ccrma.stanford.edu/groups/soundwire/research/slsq/>

larger musical strategies, eg phrasing, intentional accelerations, and their arrivals, improve synchronization when delay is an issue. Past experiences are guides for these hypotheses which remain to be tested experimentally. Repp (2005) also mentions that “... synchronization with expressively timed music is easier than synchronization with a monotone sequence that has the same time pattern ...” (page 985).

### 3.3 Relation between lead/lag interaction dynamics and tempo

Describing the interaction dynamics of ‘mutually coupled’ musicians is the next step in developing an understanding of ensemble behavior. An indication of characteristics which need to be included in such a description comes from contrasting lead/lag with tempo acceleration. The linear model [equation (2)] fits the lead/lag analysis with  $R^2 = 0.93$ , whereas linear regression of tempo acceleration<sup>(11)</sup> fits better:  $R^2 = 0.97$  (see figure 7). This difference is important. If the inflections in the lead/lag data are an indication of modes in lead/lag interaction dynamics, we can begin to model possible ‘mechanics’ which change across the span of temporal separation.



**Figure 7.** A single measure of tempo acceleration (its mean) for all performances. A linear model (thick line) correlates well with data sampled at the given delay conditions. Error bars show 95% confidence intervals for the acceleration mean. Single small dots represent acceleration mean for each individual trial.

Our clapping experiment tested only one tempo, moderately fast, at 90 bpm (with slight offsets introduced in the experimental protocol to avoid over-training to an absolute tempo, see section 2.1). Experimentation with other, significantly different tempi will be required to include tempo in any model.

### 3.4 Summary

Synchronous rhythmic behavior imposes strict bounds on temporal separation between actors. Their best synchronized trials fall within ‘natural’ time delays, ie delays within a narrow range associated with travel times across the usual spatial arrangements of clapping groups, ensembles, etc. Longer-duration aspects are presumed to have less stringent requirements which could be quantified through future experiment. Would, for example, temporal tolerances for musical versions of turn-taking (call-and-response) be akin to conversational turn-taking? Do longer-term rhythmic shapes interact with the requirements which we have derived only from strictly rhythmic tasks like the present one. The surprise benefit of focusing so closely on just the clapping rhythm has been the discovery that the most stable performances required a small amount

<sup>(11)</sup>This curve is computed with a smoothed tempo curve, merging both clappers into one curve. Smoothing is computed with a “local regression using weighted linear least squares and a 2nd degree polynomial model” (MATLAB’s smooth function included in the *Curve Fitting Toolbox*). Then, to obtain a single quantity representing a trial’s overall acceleration, the average of the derivative of the tempo curve is used.

of delay, without which we measured a tendency to accelerate. It suggests that to anticipate is a part of human rhythmic production (by 8 ms in our experimental context) and agrees with a similar tendency in related tasks (Repp 2005).

Extrapolating to network music performance, the hand-clapping experiment indicates an upper limit which corresponds roughly to a path length of 1700 km with present North American research Internets provided by Internet2 and Canarie, as test-beds (Gueye et al 2006). We provide our experimental findings as a glimpse into human factors which are key for evaluating this rapidly changing technology. The sound of Internet performance can evoke an 'in the room' experience. But when delay interferes, it has the odd quality that it is literally 'unheard'. Distant partners do not sound distant, they just get harder to play with. Their sound seems proximate because the usual distance cues are missing. Only by understanding the interaction of temporal separation synchronization can players understand how distance affects their collective rhythm.

**Acknowledgments.** Many thanks to our study team at CCRMA, including students Nathan Schuett, Grace Leslie, Sean Tyan, and the CCRMA technical support staff. Grant support from Stanford's Media-X program funded the 2004 data collection and Alberta's iCore Visiting Professor Program, the 2009 analysis. Stephen McAdams' comments on early drafts are gratefully acknowledged.

## References

- Bartlett C, Headlam D, Bocko M, Velikic G, 2006 "Effects of network latency on interactive musical performance" *Music Perception* **24** 49–62
- Benade A H, 1990 *Fundamentals of Musical Acoustics* Second revised edition (New York: Dover Publications)
- Bishop C M, 2007 *Pattern Recognition and Machine Learning* First edition (New York: Springer)
- Cáceres J P, Chafe C, 2010 "JackTrip: Under the hood of an engine for network audio" *Journal of New Music Research* **39** forthcoming, doi:10.1080/09298215.2010.481361
- Chafe C, Gurevich M, 2004 "Network time delay and ensemble accuracy: Effects of latency, asymmetry", in *Proceedings of the AES 117th Convention* (San Francisco, CA: Audio Engineering Society)
- Chafe C, Gurevich M, Leslie G, Tyan S, 2004 "Effect of time delay on ensemble accuracy", in *Proceedings of the International Symposium on Musical Acoustics (Nara, Japan)* (Kyoto: Musical Acoustics Research Group, The Acoustical Society of Japan)
- Farner S, Solvang A, Sæbø A, Svensson U P, 2009 "Ensemble hand-clapping experiments under the influence of delay and various acoustic environments" *Journal of the Audio Engineering Society* **57** 1028–1041
- Gueye B, Ziviani A, Crovella M, Fdida S, 2006 "Constraint-based geolocation of internet hosts" *IEEE/ACM Transactions Network* **14** 1219–1232
- Holub J, Kastner M, Tomiska O, 2007 "Delay effect on conversational quality in telecommunication networks: Do we mind?", paper presented at the Wireless Telecommunications Symposium WTS 2007, Pomona, CA
- Kapur A, Wang G, Davidson P, Cook P, 2005 "Interactive network performance: a dream worth dreaming?" *Organised Sound* **10** 209–219
- Pikovsky A, Rosenblum M, Kurths J, 2003 *Synchronization: A Universal Concept in Nonlinear Sciences* First edition (Cambridge: Cambridge University Press)
- Rasch R A, 1988 "Timing and synchronization in ensemble performance", in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition* Ed. J A Sloboda (New York: Oxford University Press) pp 70–90
- Repp B H, 2005 "Sensorimotor synchronization: A review of the tapping literature" *Psychonomic Bulletin & Review* **12** 969–992
- Schloss A, 1985 "On the automatic transcription of percussive music: from acoustic signal to high level analysis" PhD thesis, Stanford University
- Schuett N, 2002 "The effects of latency on ensemble performance" Undergraduate honors thesis, Stanford University

ISSN 0301-0066 (print)

ISSN 1468-4233 (electronic)

# PERCEPTION

VOLUME 39 2010

[www.perceptionweb.com](http://www.perceptionweb.com)

**Conditions of use.** This article may be downloaded from the Perception website for personal research by members of subscribing organisations. Authors are entitled to distribute their own article (in printed form or by e-mail) to up to 50 people. This PDF may not be placed on any website (or other online distribution system) without permission of the publisher.