SYNCHRONIZATION IN RHYTHMIC PERFORMANCE WITH DELAY

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MUSIC
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Juan Pablo Cáceres Chomalí
March 2013

This dissertation is online at: http://purl.stanford.edu/jc235zv4623

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Christopher Chafe, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Jonathan Abel

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Takako Fujioka

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

In the last few years, musicians have been exploring ways to perform with people in different parts of a city, a country or the world. New technologies for the Internet have been developed that are already high-quality in terms of audio experience. There is, however, a persistent problem: telecommunications delay. Even though these latencies are presently down to approximately 150 milliseconds from opposites sites of the globe, we know that delays of 20 milliseconds are already problematic for musical performance.

This research presents an experimental analysis of the rhythmic strategies that humans use to stay synchronized with different delay conditions, similar to the ones encountered in present day networked music performance concerts. The vast majority of experimental studies of sensorimotor synchronization involve humans synchronizing to machines, usually some sort of tapping experiments. Likewise, rhythmic tracking has overwhelmingly focused on the analysis of one musical source, but not on the interaction of performing musicians. This research's focus is on the latter.

The results show that performers are able to adapt to different delay conditions anticipating each other's beats. This anticipation is believed to be an intrinsic aspect of beat perception. The results are compared to earlier studies and new metrics are proposed that consider the alternating interaction between musicians. This approach also adopts phase dynamics using stroboscopic mapping.

An integrated model for tracking and generating rhythmic interactions between two performers is also presented. This model uses coupled adaptive oscillators and includes an anticipation and reaction parameter that is shown to be critical to understand rhythmic synchronization with delay. We show how this model compares to

the experimental data and use it to explain some common observations in rhythmic performance with delay.

Acknowledgements

The Stanford's Center for Computer Research in Music and Acoustics (CCRMA) is an amazing institution, unique of its kind. I am extremely grateful to be part of this community and have completed my PhD research in this place.

It is impossible to highlight my gratitude enough towards my advisor and mentor Chris Chafe. He has always given me incredible amounts of freedom and trust, some of it deserved I hope. We did many research and ski trips together, and I will always treasure his views on technology, music, and life.

I would like to thank my thesis committee members Takako Fujioka, Jonathan Abel and Ge Wang for their help, insights, and patience during the last period of writing remotely from Santiago, Chile.

I am grateful for everything I learned from CCRMA's faculty in courses, seminars, and discussions. Julius Smith, Jonathan Berger, John Chowning, Malcolm Slaney, and Bill Schottstaedt are the greatest computer music faculty members one can hope for, and I feel very lucky to have been at Stanford with an active and always engaging Max Mathews.

Most of the dataset used in this research was generated by a study team at CCRMA, including students Nathan Schuett, Grace Leslie, Sean Tyan, and the CCRMA technical support staff. My gratitude goes to them for obtaining this incredibly rich set of data. I am lucky to have a friend like Michael Gurevich who provided part of the early analysis of these experiments, and it is always a pleasure to discuss music and technology with him.

I received a Fulbright scholarship to support part of my graduate studies at Stanford. I made lots of progress in research residencies at The Banff Centre, supported

by an iCore and a Banff Centre Research Residency Scholarships.

I made many lasting friendships during my stay at Stanford and San Francisco. The most amazing musical and technical projects I worked in where collaborations with Robert Hamilton. His work in virtual worlds is really inspiring and made us push the limits of networked music performance. David Willyard, with his persistence and hard work, made it possible to transfer our technology with MusicianLink, a startup company that also supported parts of this research. Fernando Lopez-Lezcano and Carr Wilkerson helped me not only with many and diverse computational requests, but I also had the privilege to teach courses and workshops with them. Alain Renaud and I developed diverse networked concerts, and I am glad I have a friend with that kind of enthusiasm and energy to work with.

I would like to thank all my friends and colleagues from whom I have learned so much during my time at Stanford. They are too many to provide a full list, but I want to express my gratitude in particular to Bruno Ruviaro, Per Bloland, Chryssie Nanou, Justin Yang, Luke Dahl, Woon Seung Yeo, Hiroko Terasawa, Matthew Wright, Sasha Leitman, Gautham Mysore, Mihir Sarkar, Ryan Cassidy, Miriam Kolar, Kyogu Lee, and Jeffrey Treviño.

The Stanford Music Department staff has always helped with all my difficult requests during all these years. I would like to express my appreciation in particular to Debbie Barney, Nette Worthey, and Mario Champagne for all their help and support in making this happen.

I want to express my full gratitude to my mother and father, and to my brother and sister. They have always unconditionally believed and supported me in all my life adventures. My friend Stefano Corazza has also been incredibly supportive in many different ways, and he has become like a brother to me.

Finally, I would like to thank my family, with all my heart, and especially my twins Natalio and Melina. Their light and energy kept me going even during incredibly hard periods of writing this dissertation. They are sweet and bright, and this work is dedicated to them.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> Latency is the basic building block of life and music.
>
> Pauline Oliveros, *The 156th Acoustical Society of America (ASA) Meeting, 2008, Miami*

When music performers play together, they have an uncanny ability to stay synchronized. We know that ensembles don't keep metronomic time—that will sound unhuman, machine-like—but are still able to follow each-other in incredibly complex musical situations. Furthermore, experimental studies show that rhythmic duos are able to synchronize in the presence of time delays way beyond the natural acoustic situations—although with some time-keeping disruptions [18, 19, 26, 30, 6].

It is believed that human synchronization with musical beats is uniquely human and is not encountered in other animals [49, chap. 7.5.3]. There seem to be some higher level brain structures that afford human rhythmic synchronization, specifically auditory and motor coupling.

The vast majority of experimental studies of sensorimotor synchronization involve humans synchronizing to machines, usually some sort of tapping experiments [57]. Likewise, rhythmic tracking has overwhelmingly focused on the analysis of one musical source, but not on the interaction of performing musicians. Several models for

rhythmic tracking exist (see Collins [21] for a comprehensive review of current beat-tracking systems), using Kalman Filters, Bayesian Networks, Coupled Oscillators, and others.

In recent years, distributed musical performance using internet networks has introduced—or made more conspicuous—a new parameter into the problem: latency. A signal can travel no faster than the speed of light. In practice, this means that geographically separated musical performances (say between the US and Europe) include delays that are way longer than the ones needed for accurate musical synchronization.

This setting presents a problem, but also an opportunity. We know that it gets harder to perform as the delays get longer, but we don't have a quantification or a comprehensive explanation for this phenomenon. We don't understand exactly how humans behave (musically) in this delayed medium. This opens a whole unexplored research area that will have to answer what other studies have quantified and described: how musicians manage to perform together, and even predict each-other expressive timing variations.

This research tries to address two related problems:

1. Understand, from data, the rhythmic strategies that humans use to stay synchronized with different delay conditions.

2. Model these strategies using the dynamic systems approach.

It is important to note that this research doesn't try to find specific brain structures of motor and auditory behavior. It models rhythmic coupling from a dynamic perspective assumption. Although there's some speculation that rhythmic coordination may be based on internal neural oscillations that entrain to an external sequence, there's no evidence for it [47].

In order to motivate the problem in a broader context, we discuss some applications that would be possible with a model of rhythmic synchronization with delay:

Distributed Musical Performance In distributed concerts, musicians are aware and feel the disruption of delay, but don't have strategies and a conceptual framework to deal technically and musically with it. A better understudying

of the sensorimotor aspects of music performance with delay can help to create music that is workable in this environment, or to establish strategies to minimize the adverse effects.

Automatic Accompaniment There's an intrinsic delay in any system that listens to human performance and try to create an automatic accompaniment. There are several systems already in place [21], but a model that explicitly includes delay can enhance these systems.

Distributed Rock Bands Music games like *Rock Band* [2] are a great success. Adding a remote performance feature is hard because of the delay involved. A system than understand musical delay and can predict it will enable this possibility.

A further understanding of expressive performance synchronization Although the main motivation of this research is to understand synchronization in delayed environments like the Internet, the insights obtained can also be used to as part of a more general understanding of expressive tempo variations in musical performance, and the internal mechanisms to keep a performance synchronized in the presence of large variations of tempo.

1.1 Contributions

The main contributions of this research are:

1. An analysis and interpretation of rhythmic performance with delay using an adapted version of the stroboscopic technique used in the synchronization literature. The role of the strobe alternates between performers depending on musical context.

2. An integrated model for tracking and generating rhythmic interactions between two performers using coupled oscillators, including an anticipation and reaction parameter that is shown to be critical to understand rhythmic synchronization with delay.

1.2 Outline

In Chapter 2 we introduce network music and its historical development, describe the main technologies used in real-time network concerts, and explain the challenges of musical performance with delay and some current strategies to cope with it.

In Chapter 3 we describe the experiment used for this dissertation and present some analytical tools to explains some observed phenomenon in the data.

Chapter 4 presents a coupled oscillator model for the experiment presented in the previous section. The performance of this model is compared to the experimental data and we explain why some differences may happen.

We conclude the dissertation in Chapter 5 with some closing remarks, suggest improvements to the model proposed, and explain possible future research and applications.

Chapter 2

Latency in the Context of Networked Performance

Internet multimedia communications have grown exponentially and become ubiquitous in the last decade. Audio and video transmission are now common, widely extended and adopted. The specific characteristics and demands of music performance are, however, very different from the conversational communication. "Real-time" has a different meaning in music and speech. While latencies of 200 milliseconds are considered good in speech transmission, musicians have already problems with delays on the order of 20 milliseconds.

The technical requirements are only part of the story in Internet network music. A distributed performance brings to the forefront some of the most intrinsic characteristics of music: time, space and its interrelation. This type of performance forces the musicians to think about these elements because they are made conspicuous; things that were taken for granted are now a "problem" that needs to be addressed from a different perspective and using distinct strategies.

We review in this chapter part of the history of telecommunications and sound, and how latency is currently addressed in distributed music performance.

2.1 Telecommunications and Sound: Historical Overview

The idea of transmission of music to remote locations goes back to the first developments of telephony and the Musical Telegraph [38] in the second half of the 19th Century. Ever since, different attempts to connect remote locations have explored a variety of interactions; these include Hindemith and Weill's radio plays [36], John Cage's transistor radio experiments, and the first computer network experiments [7, 33], some of which where done remotely.[1]

It is only since the massive expansion and speedup of Internet in the 1990's that high-quality, real-time, bi-directional network performance between separated geographical locations, has become possible. Bargar et al. [5] set several challenges for music and audio applications on *Internet2* [40]. These include digital libraries, three-dimensional audio, going beyond virtual reality, audio web-casting, music education and forensic applications, among others.

The last decade of the 20th Century saw the first concerts between different cities using video conferencing systems (typically ISDN-based) to send audio and video. During this time, however, the time delays on the performances ranged from 0.5 seconds to 30 seconds [63], making any king of real-time musical interaction virtually impossible.

The reasons for the long latencies in audio transmission included the slow compression algorithms and low bandwidth of the time. The first dramatic decrease in delays happened in the early 2000's, with tools developed by various research groups (including McGill University [67] and SoundWIRE at CCRMA [17, 62]). The basic approach is to send uncompressed audio (avoiding the latency introduced by compression encode/decode algorithms) through high speed links like *Internet2*. With these ingredients the round trip time (RTT) came much closer to the speed of light, the theoretical limit.

[1] An historical and aesthetic overview of Interconnected Musical Networks can be found at Barbosa [3] and Weinberg [66, 65].

2.2 Networked Music Performance Technologies

Systems for real-time, high-quality and low-latency audio over the Internet that take advantage of high-speed networks are available and have been used in the last several years for distributed concerts and other musical applications [56].

Wide Area Network connections inevitably introduce transmission delays between two or more hosts. Keeping delay to a minimum is one of the main goals when tuning network parameters. Delay is known to be disruptive in musical performance [19], so a sensible goal is to minimize it as much as possible. Often, there is a tradeoff with audio quality. The longer the latency, the better the audio (i.e., less dropouts) if facing problematic network conditions.

Jacktrip [13], the main application we have used to power our concerts in the last few years, has a design that achieves:

- The highest audio quality possible, by using uncompressed linear sampling and redundancy to recover from packet loss.
- Throughput maximization, which gets audio packets onto and off of the network as soon as the sound card can deliver them.
- Working with any number of channels (depending on available computer processing power and bandwidth).
- Flexibility in routing and mixing audio channels from and to the different hosts.

Other networking audio packages use different strategies for audio delivery and synchronization. These include the use of compressed audio, artificially increased delays that match one or more musical measures [32] (i.e., musicians play asynchronously with the output that was generated one or more measures before) and one-way recording techniques (with one location/performer at a time) where latency is not an issue [56]. NetJack uses a master/slave approach to synchronize audio clocks. This approach is most suitable for local area networks (LAN) where jitter is smaller [15]. Other systems (nStream, SoundJack and jack-tools) also deliver uncompressed audio [10].

Although some of these technologies minimize audio latency, for longer geographical distances we will still be limited by the theoretical limit of the speed of light. It is thus important to understand the problem of latency in the context of musical performance.

2.3 The Problem with Latency

Figure 2.1 shows an example of the current scenario in network performance. The sound produced by the performer on the left (ipsilateral) and the one on the right (contralateral) are shown both locally and remotely, with the present global time indicated. The signals represented with thin blue lines show the amount of latency introduced by the network. The final result on each location (the combination of both sounds) is different.

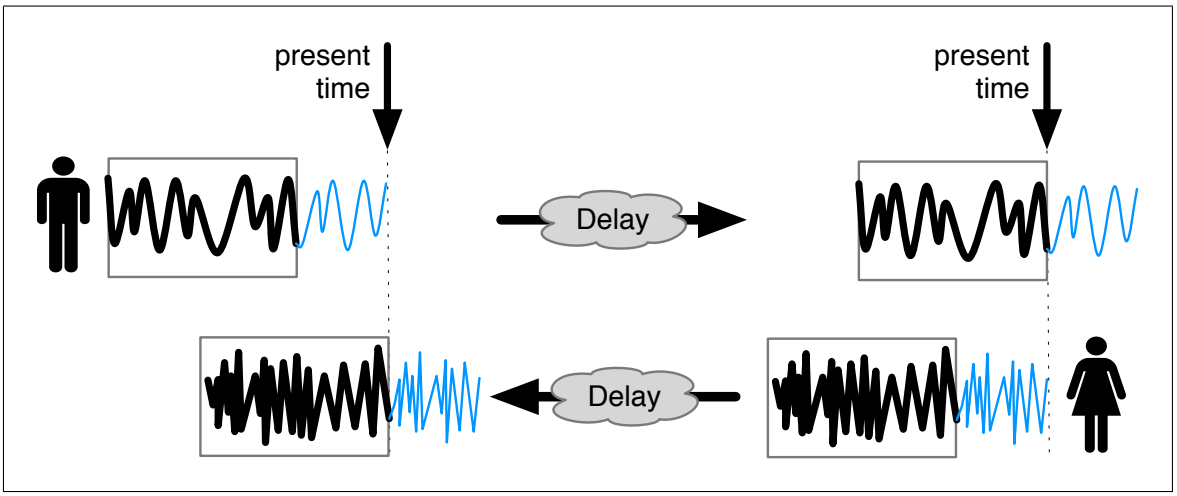

Figure 2.1: The current situation on the network.

Depending on the amount of delay, this scenario may or may not be a perceptual and performative problem. In the experiment analyzed in this research (it has come to be known as the "Clapping Experiment") [19], two subjects are asked to perform a simple rhythmic task, which is performed with signal latency in the range of 1–78 milliseconds. The delayed sound is just the one coming from the other performer. The subjects can always hear their own sound without any latency (without Delayed

Auditory Feedback, DAF). The results show that there is a negative linear relation between tempo acceleration and delay. For delays of more than 12 milliseconds, the longer the latency, the more severe the tempo deceleration in the performance. For shorter delays there is also a surprising tempo acceleration.

A different experiment documented online [12] and discussed in Section 4.6.1 shows the same results in a real musical situation with the *St. Lawrence String Quartet*. Neglecting any asymmetries, the performers in the string quintet (viola at Stanford, CA, the quartet at Banff, Alberta) were separated by a path of approximately 25 milliseconds (unidirectional). In the performance of standard classical repertoire (Mozart's String Quintet in G minor, K. 516), the deceleration effect was also experienced. Other observable effects included the power balance/unbalance between the bigger ensemble and the smaller one (the single viola). In this case, the viola was following the quartet, and the tempo struggle happened on the quartet side. The viola was much more comfortable during the performance.

Other studies have done similar experiment with comparable results [6]. Experiments exploring cases like Delayed Auditory Feedback and attack length dependent perceptual synchronization will be reviewed in Section 3.1.

2.4 Some Current Musical Strategies to Deal with Latency

2.4.1 Synchronization and Conducting Systems

Trying to deal with the time delay introduced by network latency has been the main focus for many researchers. Currently, the main approach is to add delay to one or more of the participants in the performance, so that the amount of delay can translate into musical terms.

One of the first basic synchronization approaches over wide area networks (WAN) was proposed by Goto et al. [32], Goto and Neyama [31]. It relies on the *one-phase delay* concept used for metronomic music. The connection delay is increased by a constant corresponding to a repetitive musical structure, e.g., a 12-bar blues chord

progression. Each musician is then playing on top of a delayed version of all the others. NINJAM [1], a software application that implements this technique provides several audio examples on its website. These approaches have the limitation of only being useful for metronomic and repetitive (with loops) music only.

Bouillot [9] explores a similar approach with a distributed metronome and user-configurable delays. He also includes Delayed Auditory Feedback (DAF) in the synchronization process, as in Chew et al. [20]. This approach works to synchronize musicians but has the added and artificial cost of making musicians hear their own instrument delayed, like a church organ player.

Fober et al. [27, 28] propose an algorithm to compensate for the clock frequencies drifts when the receiver and sender clocks differ. Time-stamped packages (events, not audio) are sent and the receiver can use its local clock to order and synchronize packages.

In a different approach, an ad-hoc interactive environment is presented by Hajdu [35] for five musicians and one conductor. He streams just data messages. In this case synchronization is not a concern; the conductor has just the function of defining musical parameters but without any precise synchronization control. Several pieces are performed in the environment, including John Cage's *Five*.

2.4.2 Anticipation and Music Prediction

The literature that deals specifically with music prediction is very limited. In an early study Chafe [16] analyzes different performances of the same pianist in a Charles Ives passage. Just timing and key velocity are used as feature parameters. The study uses Mahalanobis distance [24, Appx. A] to cluster different performances. Real-time analysis/synthesis of the performance is proposed to classify and predict musical performance to overcome transmission delays. Prediction in a teleautonomous control in robotic applications is also suggested.

A recent study [59, 58] presents a system for recognition and prediction in Indian tabla performance. The system is unidirectional, and defines a transmitter and a

receiver. After preprocessing the input audio signal with the aid of contact microphones for a more accurate event detection, the system sends symbolic data over the network; no audio is transmitted. The symbolic data system is similar to the one encountered on speech recognition, with audio data associated with bols, the equivalent of phonemes in speech recognition. The K-Nearest neighbors algorithms [24, Ch. 4] is used to classify similar bols. The phrase prediction is implemented using simplified formal grammars [24, Ch. 8]. At the time of writing there is no demo or audio available from this project.

2.4.3 Supervision and Control in Engineering and Music

A complete review of the extensive literature on supervisory control is out of the scope of this dissertation. There are, however, some recent articles that deal specifically with the similarities and dissimilarities of supervisory control in engineering and music. Inagaki and Stahre [39] has a good description of the field in engineering as well as its relation with the recent literature in music. Human supervisory control emerged in the decade of 1960 as part of teleoperated lunar vehicles and manipulators research. In order to deal with the long transmission delays, the remote computer communicates with the human operator, scheduling short-range goals between orders. The local machine—human operated—would also mimic predicted behavior of the remote robot manipulator.

Johannsen [41] describes several activities in the arts and how they relate to the engineering concepts of supervisory control, which include conducting, playing musical instruments, sound design, information retrieval, and others. This research is however more concerned with a qualitative description of the musical field in a supervisory control framework rather than the application of its techniques to music. Inagaki and Stahre [39], in his description of an orchestra, also tries to obtain inspiration from the musical fields. Sheridan [61] has also a similar description of the orchestral hierarchical levels, and includes some considerations of virtual reality and telepresence in the context of music technology.

Chapter 3

Rhythmic Coordination with Delay: Experimental Synchronization Analysis

3.1 Effects of Delay in Rhythmic Synchronization

Several studies on perceptual aspects of synchronization have been attempted. Hirsch [37], in his pioneering study on perception of temporal order, identifies three ranges of perception; the short range extending from 0 to 20 milliseconds (phase perception), a middle range between 20 and 100 milliseconds (auditory patterns), and a longer range for 100 milliseconds and more (separate auditory events). Several later studies agree with these ranges to within a couple of milliseconds [23]. It is also important to be aware of the Hass (or precedence) effect [52, 29]; when identical sounds have arrival times that differ by 30–40 milliseconds, the sound is perceived as one coming from a specific location rather than two distinct sounds.

In actual musical situations, synchronization between different instruments is never perfect, and the tolerances in the performance of small ensembles, like wind and string trios, have been shown to be as high as 50 ms on average during a complete performance [55]. Several factors like instrumental timbre, length of articulations, masking, and reverberation help to make this situation perceptible. Computer

music rendering and simulation of performance has shown that variations on synchronization are perceived as "human like", whereas systematic perfect synchronous performances sound more "machine like". It is also interesting to note that as ensembles grow in number and physical size, asynchronization also grows. At a certain point the presence of a conductor becomes necessary to maintain synchronization for larger ensembles.

Delayed Auditory Feedback (DAF)—one's own sound delayed—experiments are also relevant in situations when the delay is large (more than 100 ms one way). A common effect of DAF is a slowing down on performances [30, 51]. This can be understood by performers "waiting" for the next sound to catch up, consequently making the performance slower with time. A similar effect has been observed on simple rhythmic tasks [19][1] and on the performance of Poulenc [20] and Mozart [6] repertoire between two musicians. This later study shows strategies by which musicians adapt to the delay. These strategies include playing ahead of each other, or deciding that one of the two musicians will following the other. The latter seems to be the natural tendency when there is an unbalance in the structure of the ensemble such that one side can dominate rhythmic timing. The weak side naturally follows the strong one, and we discuss examples of this below. Bartlette et al. [6] show that musicians rated the musicality of a performance as high even for delays of 46 milliseconds (unidirectional).

3.2 Experiment: Clapping Duos with Delay

3.2.1 Experiment description

NOTE: We have described this experiment in detail previously. The reader is advised to consult Chafe et al. [19] for details. A summary will be presented here.

We examined performances by pairs of clappers under different delay conditions. A simple interlocking rhythmic pattern was chosen as the task (Figure 3.1). Subjects in separate rooms were asked to clap the rhythm together while hearing each other's

[1] We discuss this study and use the same dataset in the present dissertation.

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.

The pattern has three properties which are conducive for the experiment: first, it comprises independent but equal parts rather than unison clapping (a kind of simple polyphony), second, it creates a context free of "internal" musical effects [6], and third, the rhythm can be analyzed for lead/lag (the metrical structure's phase advance can 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 (and with no visual contact). 11 delay conditions in the range from $d = 3$ to $78ms$ (one-way) were introduced in the sound path (electronically) and were randomly varied per trial. The shortest $d = 3ms$ is equivalent to having a subject clapping $1m$ from the other's ears. The longest $d = 78ms$ equates with a separation of approximately $26m$, equivalent to a distance wider than many concert stages. 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, $d = 1ms$.

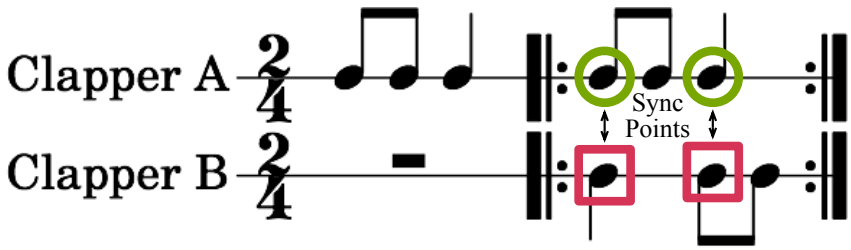

Figure 3.1: Duo clapping rhythm used.

3.2.2 Method

3.2.2.1 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 $d_n(\text{ms}) = n + 1 + d_{n-1}$ which produces the set:

$$d_0 = 1; d_n = \{3, 6, 10, 15, 21, 28, 36, 45, 55, 66, 78\}(\text{ms})$$

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 per trial was also randomly selected from one of three pre-recorded "metronome" tracks of clapped beats at 86, 90 and 94bpm. (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 1 subject-against-recorded-track which also ran at the end of the block, also not included.) A final $1ms$ 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 minutes to complete.

3.2.2.2 Number of subject pairs and trials

24 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.

3.2.2.3 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 3.2). One microphone (Schoeps BLM3) was located 0.3 meters 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. Glasses wearers 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 clapping hands to microphone introduced a time delay of about $1ms$ and is added into our reported delays. In other words, our reported $3ms$ delay comprises respectively, $1ms + 2ms$, 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 96kHz audio interface (M-Audio PCI Delta 66, Omni I/O). Custom software was written in C++ using the STK[2] set of open-source audio processing classes which were interfaced to the Jack [3] real-time audio subsystem. All delays were confirmed with analog oscilloscope measurement. Each trial was recorded as a stereo, 16bit, 96kHz sound file. The direct microphone signals from both rooms were synchronously captured to the two channels.

3.2.2.4 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

[2] http://ccrma.stanford.edu/software/stk/
[3] http://jackaudio.org/

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*). Their starting tempo was established by play back of a short clip (6 quarter-note claps) recorded at the target tempo. 3 starting tempi were used in random order (86, 90, 94bpm) in order to avoid effects of over-training to one absolute tempo. Trials proceeded in the following steps:

1. Room-to-room audio monitoring switches on.
2. A voice recording (saying "San Francisco" or "New York") plays only to the respective *initiator*, to cue them up.
3. A recording of clapped beats at the new tempo (functioning as a metronome) plays for 6 beats only to the *initiator*.
4. The *initiator* starts rhythm at will. The other subject has heard nothing until the point when they hear the *initiator* begin to clap.
5. The other joins in at will.
6. After a total of 36 seconds, the room-to-room monitoring shuts off, i.e., communication is cut, signaling the trial's end.

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.

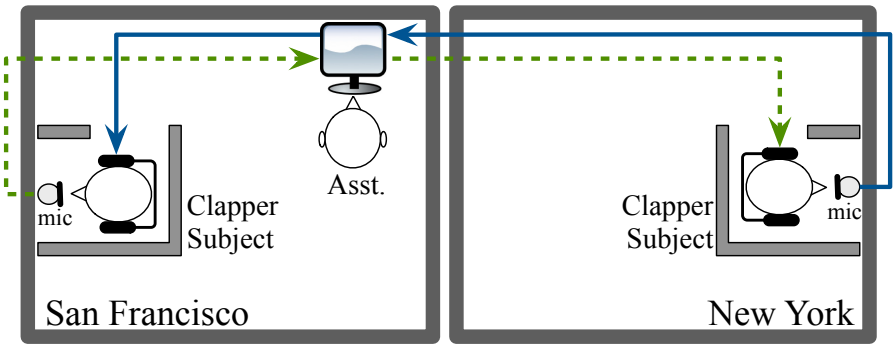

Figure 3.2: Floor plan. Rooms were acoustically and visually isolated and room reflections were minimized with sound absorbing panels. Electronic delay from mic to headphones was manipulated by computer.

3.2.3 Processing of Recordings

3.2.3.1 Recorded segments of interest

We were interested only in the sections of the recordings in which both clappers are 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 are involved. For the trial shown in Figure 3.3, clapper B (red squares) starts the trial and is followed by clapper A (green circles). Enclosed (circles and squares) notes correspond to the common beats which were automatically identified in a first pass on the raw data.

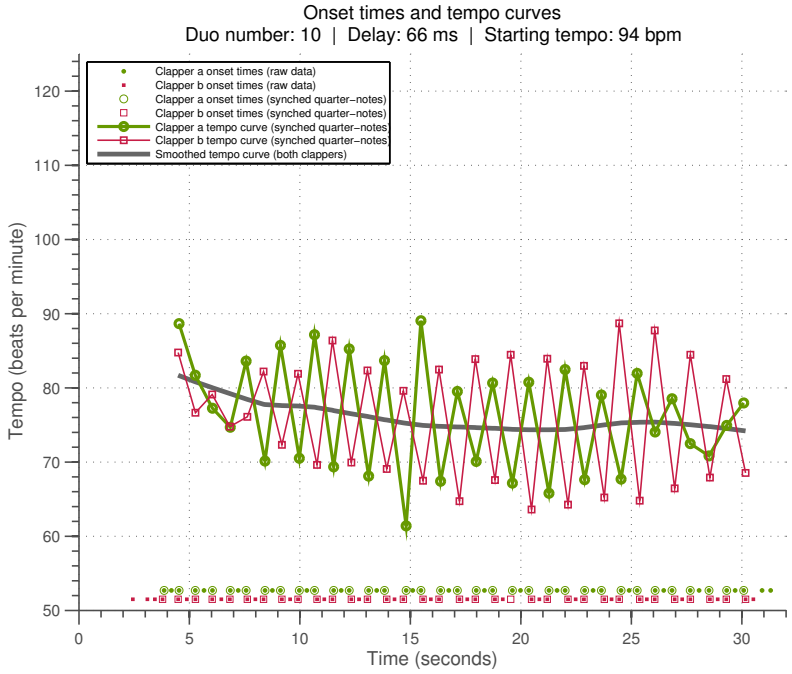

Figure 3.3: Example of onset times, synchronization points and tempo curves for one trial. A smoothed tempo curve is derived from the instantaneous tempi of both player's synchronized events.

3.2.3.2 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 using the "amplitude surfboard" technique [60] tuned to measure onsets with an accuracy of ±0.25ms. The extremely clean clapping recordings allowed false events (usually spurious subject noises) to be rejected using 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.

3.2.3.3 Validation

Recordings were automatically examined and only validated for inclusion in further analysis if they passed several automatic tests. Ninety-five trials contained more than one missing event per clapper and were discarded. If only one event was missing, it was automatically fixed through interpolation. Four 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.

3.2.3.4 Event labeling, tempo determination

Inter-onset intervals (IOI's) 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 and quarter-notes were present, the IOI's were clustered into two separate groups using the k-means clustering algorithm [8]. 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 (in bpm) was computed with:

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

$$\text{tempo}_{eighth-note} = \frac{60}{2 \cdot \text{IOI}}(\text{bpm})$$

3.2.3.5 Effect of Starting Tempo

ANOVA and multiple comparisons of the mean tempo at each of the three starting tempi (86, 90, 94bpm) 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.

3.2.3.6 Database

Figure 3.3 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 3.4 groups smoothed tempo curves for each condition (including the control). Data for the full set of trials is available online[4] for continuing analysis. The site also offers the algorithm code for the present analysis.

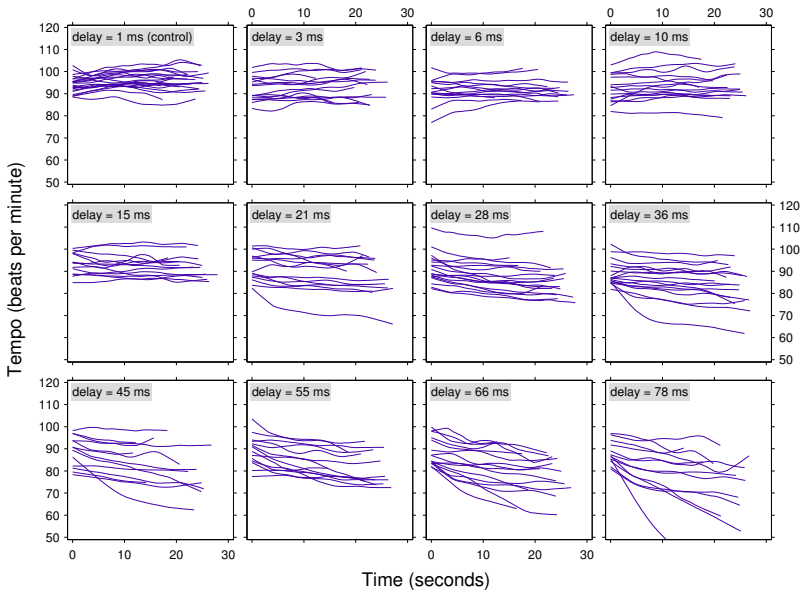

Figure 3.4: All trials' tempo curves grouped by delay. Tempo acceleration during a given performance is tracked by measuring inter-onset intervals as shown in Figure 3.3.

[4] https://ccrma.stanford.edu/groups/soundwire/research/temporal-separation-article-som/

3.3 Lead/Lag Synchronicity Analysis

3.3.1 Synchronization points

The assigned rhythm in Figure 3.1 creates points at which claps should be simultaneous, also highlighted by circles and squares in Figure 3.5. A clapper in "San Francisco" is green circles, a 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 $3ms$; approximately synchronous at $15ms$; lagging at $78ms$. Lead/lag is measured with respect to measure-length periodicity. Odd-numbered events have inverted (antiphase) sign. 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.

For the examples represented in Figure 3.6, the lead-lag factor was computed as follows:

$$\text{lead-lag} = (a_{4thsync}[1] - b_{4thsync}[1]) + (b_{4thsync}[2] - a_{4thsync}[2]) \tag{3.1}$$

where $a_{4thsync}[n]$ are green circles sync points (clapper A) and $b_{4thsync}[n]$ are red squares sync points (clapper B).

This differs from previous studies which have measured the absolute value of asynchronization. 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 their partner. Figure 3.6 compares these means and their variances (95% error bars). At very small delays, performances are dominated by a tendency to lead. Increasing delay traverses two "plateaus:" first is the region with best synchronization, followed by a second plateau beginning at $28ms$ delay. At the greatest delays, lag increases dramatically.

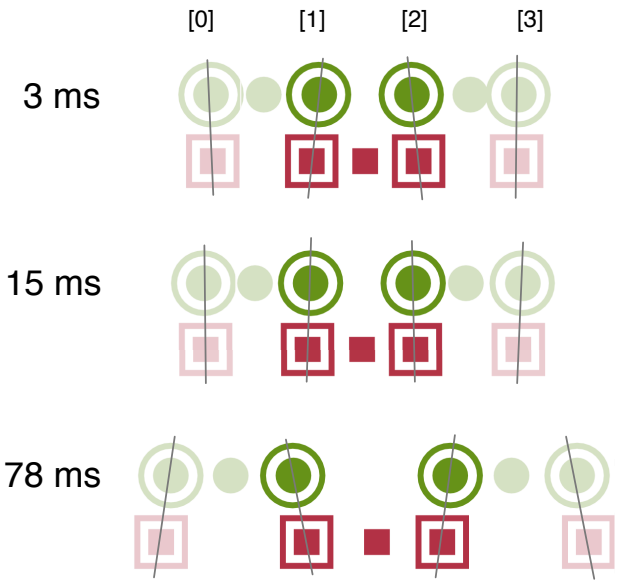

Figure 3.5: Lead/lag at different delays.

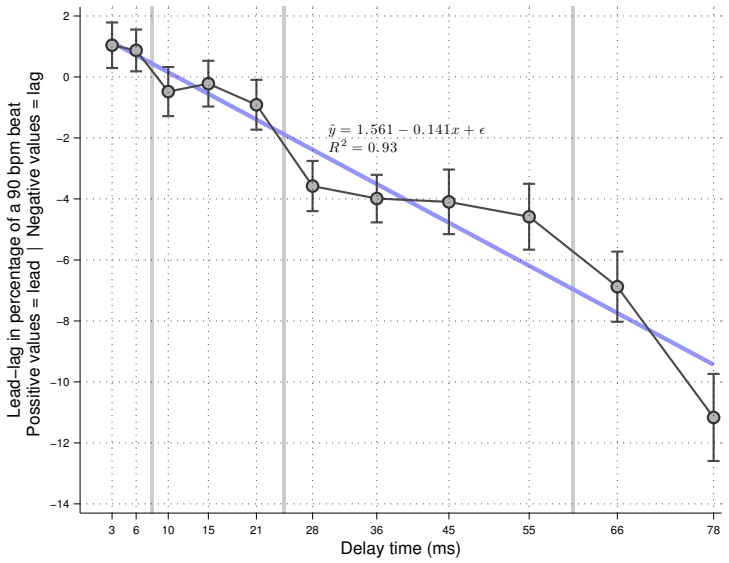

Figure 3.6: Onset asynchrony measured at all beat points for the set of delay conditions. Error bars show 95% confidence intervals.

3.3.2 Tempo Acceleration

A single measure of tempo acceleration (its mean) is compared for all performances[5]. In Figure 3.7 a linear model (green thick line) correlates well with data sampled at the given delay conditions. Error bars show 95% confidence intervals for the acceleration mean. Single blue dots represents acceleration mean for each individual trial. We can see that for very small delays, there is a tendency to accelerate. Longer delays produce deceleration as expected.

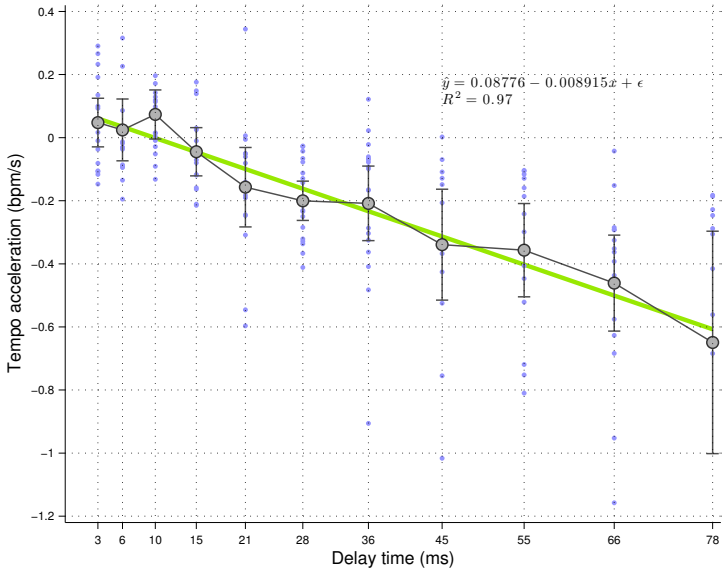

Figure 3.7: A single measure of tempo acceleration (its mean) is compared for all performances. Error bars show 95% confidence intervals.

3.4 Rasch Asynchronization

Rasch [55] defines *asynchronization of a pair of voices* (A_{Rasch} in this study) as the "standard deviation of the onset time differences of simultaneous tones of those voice

[5]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.

parts." The details for its computation are described in the cited paper, but we summarize the procedure for the clappers duo.

For each duo, we fist compute the mean onset times:

$$\bar{w} = \frac{a_{4thsyn} + b_{4thsync}}{2} \tag{3.2}$$

Then the relative onset times are:

$$\begin{aligned} v_a &= a_{4thsync} - \bar{w} \\ v_b &= b_{4thsync} - \bar{w} \end{aligned} \tag{3.3}$$

For duos the relative onset times is symmetric (i.e $v_a = -v_b$), so in our case it doesn't convey sign information. We are therefore interested only in the $\mathrm{A_{Rasch}}$, defined as the standard deviation of the onset times difference $d_{ab} = v_b - v_a$.

We compute $\mathrm{A_{Rasch}}$ for all trials (each duo has a single $\mathrm{A_{Rasch}}$) and then average the for each delay condition (Figure 3.8).

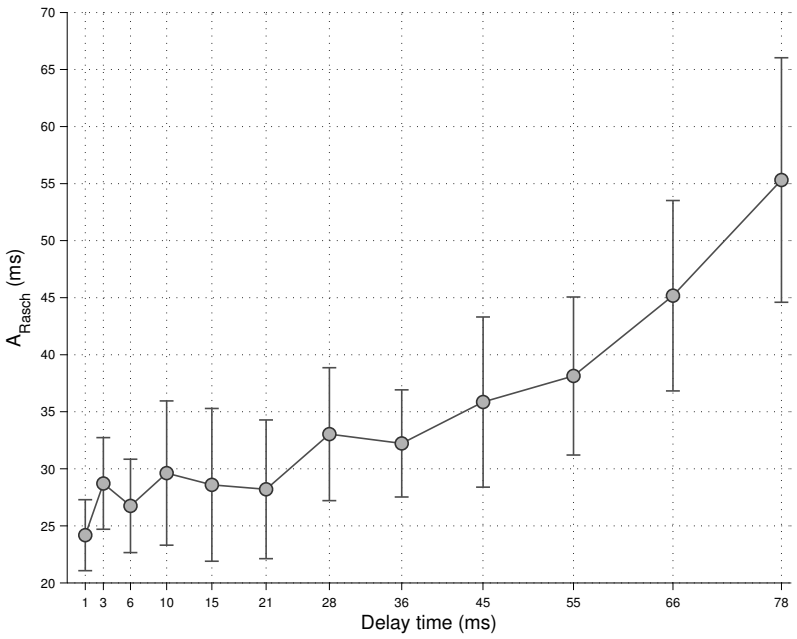

Figure 3.8: Rasch Asynchronization means. Error bars show 95% confidence intervals.

Rasch asyncronization doesn't reveal interaction dynamics, but shows, as expected, that asynchronization increases for longer delays. For "natural" delays (between 1 and 21 milliseconds), error bars overlap which suggest that asynchronization is very similar. This shows the need of a metric that considers the interaction dynamics and can be measured in relative and absolute terms.

3.5 Relative-Time *Alternating-Asynchronization* Analysis

All the studies discussed previously, including the lead-lag metric presented earlier (see Section 3.3), use *absolute time*. This means that there's a central computer that captures audio with the same time reference for all performers. This is fine for metrics that evaluate absolute synchronization. However, it is also important to examine how each performer understands time within his *local* frame of reference. The discussion that follows use *relative time* for its analysis. We will discuss how this is important in understanding the dynamics of rhythmic performance with delay.

The current understanding is that beat perception is anticipatory and not reactive [50]. This means that performers synchronizing to an external rhythmic stimuli play, on average, a little bit ahead of the expected beat, not after or exactly at the same time. For ensembles with delay, this anticipation or reaction is performed in relative time, i.e., the time frame of the each performer.

3.5.1 Asynchronization Analysis

In the synchronization literature, the *stroboscopic technique* [53, chap. 3.2 and 6.3] is commonly used in experimental synchronization analysis. The technique consists of observing the state of an oscillator (usually the phase, but we will observe times also) at the times dictated by a second oscillator (a strobe oscillator). We will then alternate the role of the observed and strobe oscillator in the metric discussed in what follows.

The interlocking nature of the rhythmic task of our experiment (Sec. 3.2.1) can be

understood as an alternation in the role of **stimulator** and **responder**. Therefore, in the context of rhythmic interaction, we define:

Stimulator Beat in which the performer has the rhythmic initiative, i.e., he is not waiting for another performer to synchronize.

Responder Beat in which the performer is waiting and trying to synchronize with the other performer.

This was already implicit in the *lead-lag* metric used in the previous analysis (Sec. 3.3), but we will now consider this alternation between the role of stimulator and responder in computing asynchronization in a more general way. The result is slightly different in our experiment but can be generalized for other rhythmic patterns.

Figure 3.9 illustrates how the *alternating-asynchronization* was measured for the clapping task. The pattern is defined as the sequence eighth-note–eighth-notes–quarter-notes (♪–♪–♩), numbered 1-2-3 (Figure 3.9 (a)). In that illustration, the first clapper (stimulator) has to synchronize to the second one (responder) and then the process is reversed. The *lead-lag* considered this as a pair. Now each synchronization is newly established in each beat 1 and depends only on who is "waiting" for the other performer. The interlocking clapping pattern makes it easy to define and understand who takes the role of stimulator and responder, but this would work also for more complex interactions as long as we can define this role separation.

We are interested in the sign of asynchronization. If the *stimulator* plays after the *responder*, asynchronization is positive, if it plays ahead, it's negative. Figure 3.9 (b) and (c) shows these different situations.

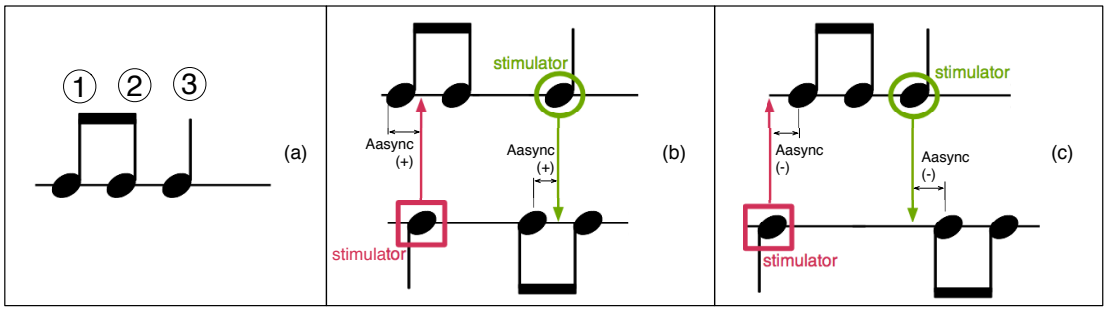

Figure 3.9: Alternating-asynchronization as measured for the interlocking pattern. (a) Pattern order definition. (b) Example of positive alternating-asynchronies. (c) Example of negative alternating-asynchronies.

We compute *alternating-asynchronization* across all duo trials in absolute and in relative time for each delay condition. The analysis produces alternating-asynchronization value that aggregates all trials. Figure 3.10 compares these means and their variances (95% error bars) in absolute time. Figure 3.11 compares these means and their variances (95% error bars) in *relative time*. Note that since in our experiment the delay was symmetric, this corresponds to approximately subtracting the delay from the mean.

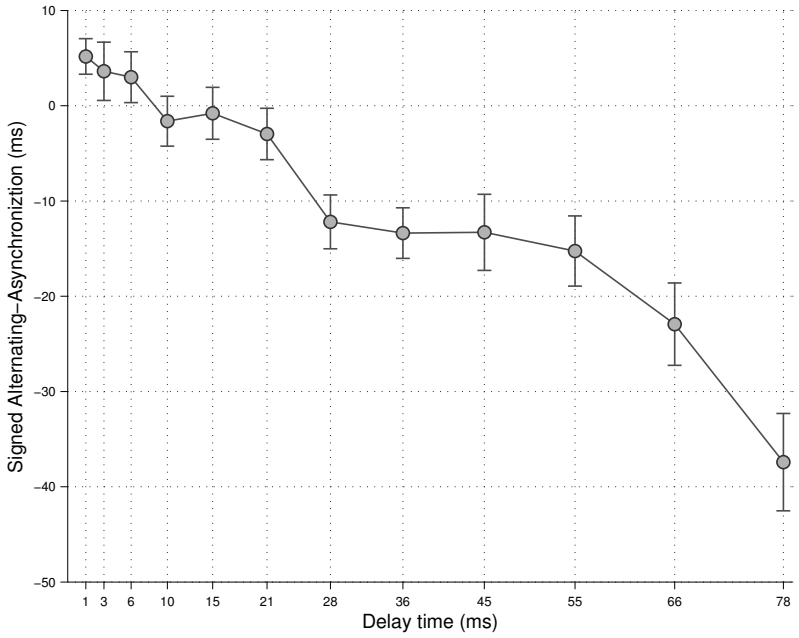

Figure 3.10: Alternating-asynchronization versus delays means. Error bars show 95% confidence intervals.

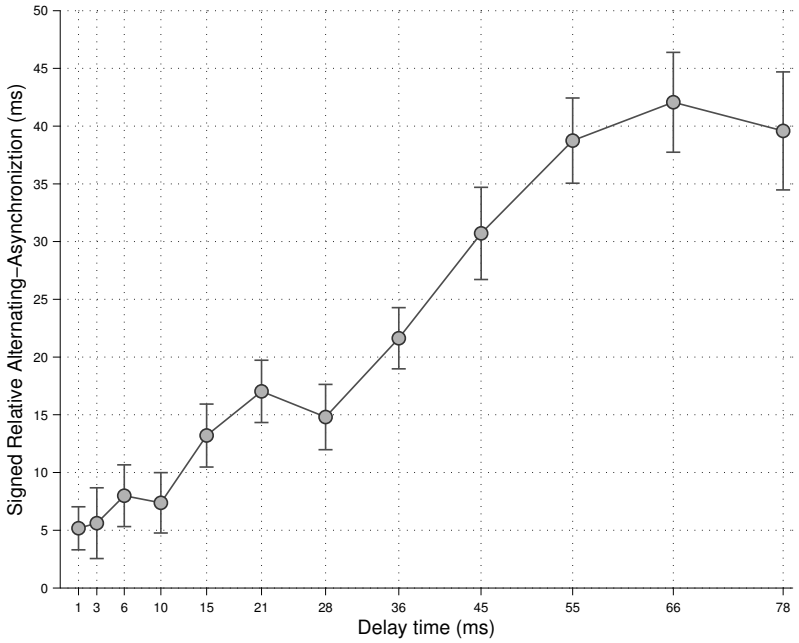

Figure 3.11: Relative alternating-asynchronization versus delays asynchronies means. Error bars show 95% confidence intervals.

3.5.2 Phase Analysis

We want to see if there is a relationship between phase correction and frequency adaptation. To investigate this, we use the same technique (strobe alternating-asynchronization) to compute the phase, ϕ, normalized to -0.5 and 0.5 and defined for the pattern shown in Figure 3.12.

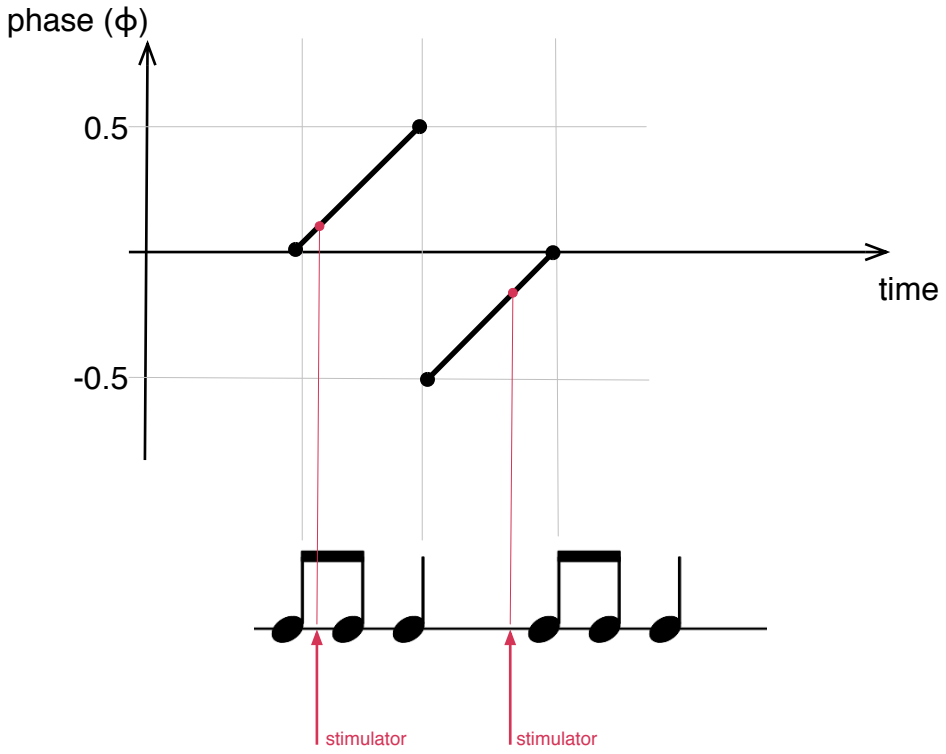

Figure 3.12: Phase stroboscopic analysis. For the rhythmic pattern used, phase is defined as above.

Alternating-asynchronization measured time differences, while phase analysis measures a *proportion* of alternating-asynchronization with respect to instantaneous tempo. For example, for slower performances, the same alternating-asynchronization (in time units) will result in a smaller ϕ. Figure 3.13 and Figure 3.14 shows means phase relations for all trials in absolute and relative terms.

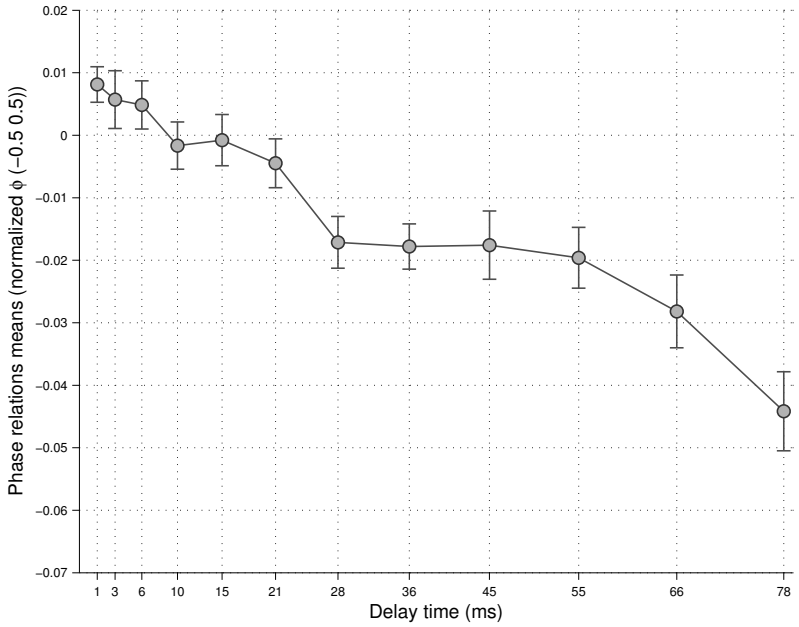

Figure 3.13: Mean phase relations. Error bars show 95% confidence intervals.

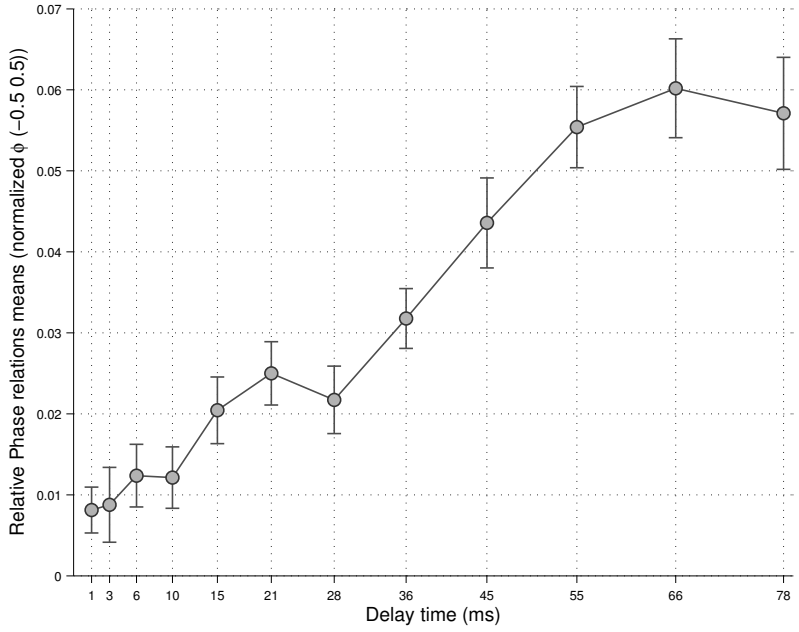

Figure 3.14: Relative mean phase relations. Error bars show 95% confidence intervals.

3.5.3 Internal Tempo Analysis

Asynchronization is measured relative to a second performer. We are also interested in measuring if there is an internal mechanism by which clappers have control of synchronization and tempo. We have already seen that tempo decreases with longer delays. The model "a waits for b, b waits for a" [34] that accounts for this ritard would produce decelerations much higher than the ones observed (see Section 4.4). Alternating-asynchronization already suggests that there are higher anticipation levels for longer delays. We now turn our attention to internal tempo control.

In the pattern used, subjects have more control over tempo in the first part of the pattern, after which they have to wait for the other clapper to perform his part. Therefore, we can measure a ratio between these two inter-onset-intervals (IOIs) to obtain a measure of this effect. Figure 3.15 shows the two IOI of reference we use for the computation of this metric.

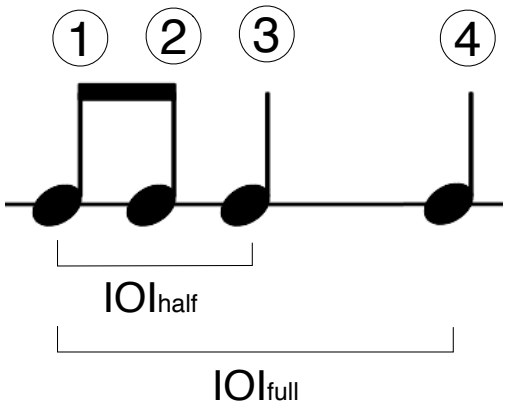

Figure 3.15: Internal IOIs definition

We then compute $IOI_{ratio} = \frac{IOI_{half}}{IOI_{full}}$ across all trials.

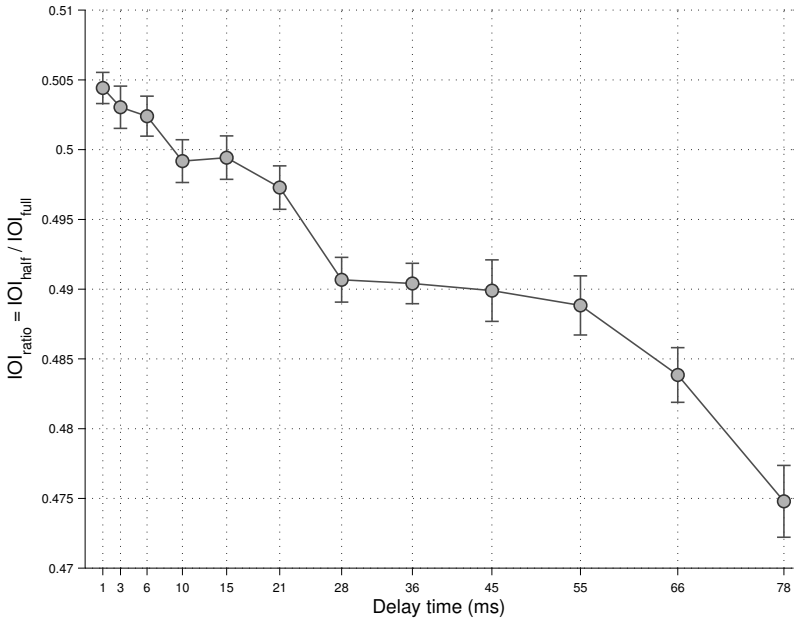

Figure 3.16: $\mathrm{IOI}_{\mathrm{ratio}}$ for all trials means. Error bars show 95% confidence intervals.

3.5.4 Individual duo imbalances

We are also interested in obtaining a metric for any internal imbalances for each duo. This can appear as one clapper leading versus the other following (reactive). We can get a metric of an internal duo imbalances by computing the mean difference between relative phases, i.e., the difference between relative mean phase of one clapper compared to the other. This will give us a metic for how much a clapper is leading with respect to the other for each trial.

Figure 3.17 reveals a slight increase with delay but only significant (most confidence intervals overlap) for longer delays (66 and 78). It is not significant enough though to establish a trend.

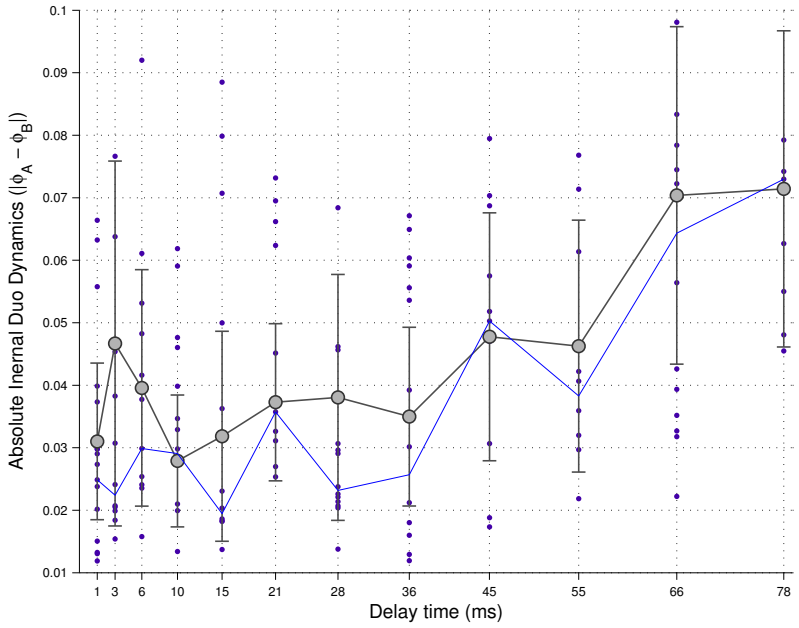

Figure 3.17: Individual duos imbalances per delay (dots) and means. Error bars show 95% confidence intervals. Thin blue line is median.

3.5.5 Role-based leading/reactive clappers

We want to analyze if the initiator who establishes the tempo by clapping first may have assumed an unintended musical role as leader. For this, we use the internal duo imbalance phase metric. Figure 3.18 shows the results with initiator with respect to follower, i.e., mean(initiator)-mean(follower). A positive result indicates that the initiator leads, a negative one that the follower leads. We see no significant impact on who assumes the initiator role (mean is very close to 0).

Figure 3.19 plots internal duo imbalance with A and B as references. Green circles indicate that Clapper A leads that trial, and red squares that Clapper B leads. We see that this reveals a pattern that the initiator versus follower analysis doesn't. This suggests that the roles of leader and follower are assumed by a clapper independently of whether or not he starts the trial to establish the tempo.

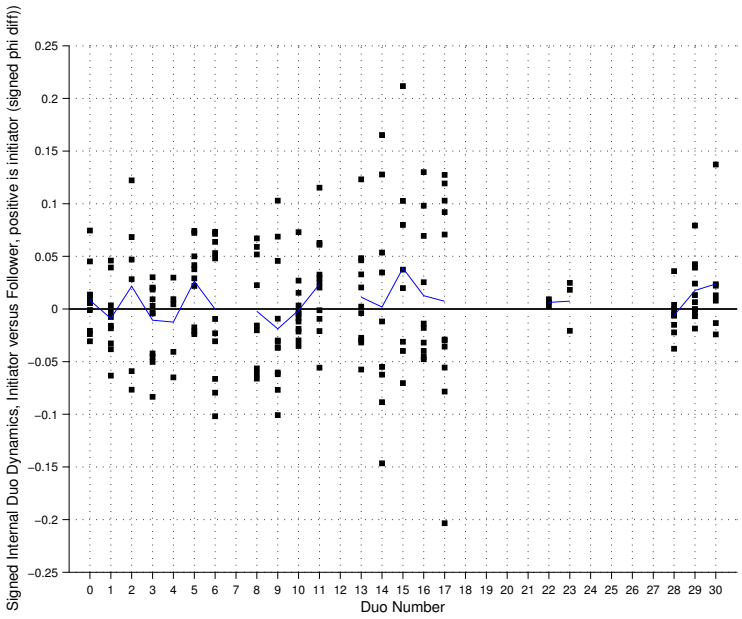

Figure 3.18: Individual internal duo dynamics (mean of phi per duo), initiator versus follower. Thin blue line is median.

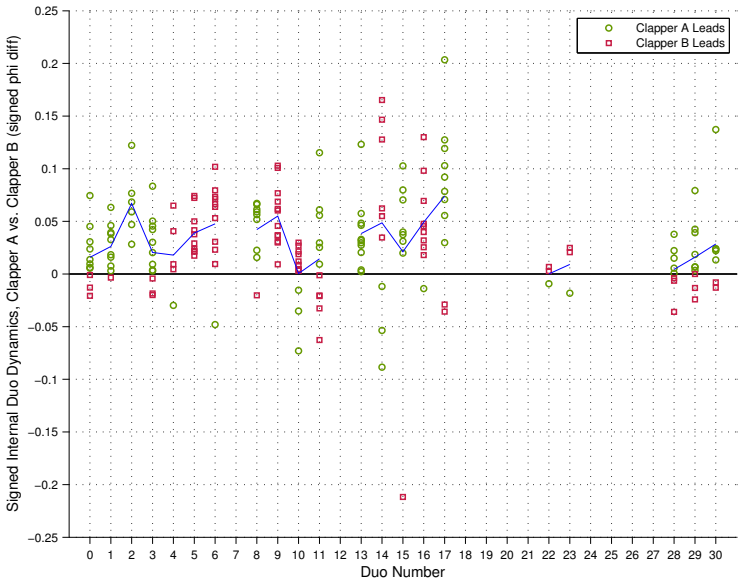

Figure 3.19: Individual internal duo dynamics (mean of phi per duo), Clapper A with respect to Clapper B. Thin blue line is median.

3.6 Discussion

3.6.1 Lead-lag versus alternating-asynchronization

Both lead-lag and alternating-asynchronization metrics preserve the sign in order to observe changing interactions. If we compare Figure 3.6 with Figure 3.10 with the naked eye, they follow the same trend. The actual numbers differ slightly though, because lead-lag needs a pair to measure asynchronization. Since not every duo is exactly even or odd, some beats are lost at the end in a number of duo. Alternating-asynchronization, in turn, measures every beat.

Furthermore, the lead-lag metric only measures interactions in pairs and in absolute time, but alternating-asynchronization gives also a relative *anticipation* number that is comparable to tapping experiments [50]. We discuss this in the next section.

3.6.2 Regimes

We discussed already the different regimes in Chafe et al. [19]. We reproduce the summary table here (Table 3.1).

Regime	Delay (ms)	Air equiv. (m)	Net equiv. (km)	Effect
1	3,6 (**8**)	<3	<500	acceleration
2	10,15,21 (**25**)	8	1700	"natural"
3	28,36,45,55 (**60**)	20	4000	deceleration
4	66,78	20<	4000<	deterioration

Table 3.1: Clapping regimes, actual sampled delays and interpolated transition values (**bold** in parenthesis).

When we look with alternating-asynchronization analysis, we see different regimes though (Table 3.2), and this may explain how deterioration happens (Figure 3.11). It is first important to note that the sign is always positive. This suggest that there's always a tendency to anticipate the beat, even for very short delays, which is consistent with the tapping literature.

The next feature that we see is that for longer delays, alternating-asynchronization stabilizes at approximately 40 milliseconds. This suggests there's an upper bound

beyond which no more anticipation is performed, and explains why deterioration starts at 55 milliseconds of delay. Since an ideal compensation (alternating-asynchronization) will be the same that as the delay and we have an upper bound of 40 milliseconds for alternating-asynchronization, the difference between delay and alternating-asynchronization will only increase as delay increases, producing the deterioration observed.

Regime	Delay (ms)	Effect
1	3,6,10	natural (produces acceleration in some cases)
2	15,21,28	first adaptation plateau
3	35,45	strong adaptation
4	55,66,78	lead (confidence) stabilizes

Table 3.2: Clapping regimes for relative alternating-asynchronization.

3.6.3 The role of anticipation in delayed performance

We mentioned previously that beat perception is anticipatory and not reactive, and this results further confirm this theory. For all delays measured, there's always a rhythmic anticipation (Figure 3.11). For short delays this is consistent with tapping experiments, but we can also see that this anticipation increases for longer delays. Previous studies have hypothesized different strategies, that include:

- Intentionally pushing the beat
- Leading by ignoring the sound of part of the ensemble

We will discuss in the next section if there is a frequency adaptation "inside" the beat. But the current results suggest that there is a stronger phase coupling for longer delays.

3.6.4 Frequency adaptation inside the beat

Figure 3.16 shows different regimes for internal tempo of each clapper. This gives us a measure of how much frequency adaptation there is for each delay condition. We already discussed that there is a strong phase coupling (anticipation). The

different regimes are shown in Table 3.3. "Natural" delays show a symmetrical IOI_{ratio}. For shorter delays, the adaptation shows slower IOI_{half} with respect to IOI_{full} ($IOI_{ratio}<0.5$). This means that clappers are "taken by surprise" by the partner. For the second regime shown in table, $IOI_{ratio}\approx0.5$, which is the natural and expected condition of the pattern. Longer delays show a strong frequency adaptation inside the beat, with subjects accelerating IOI_{half} with respect to the IOI_{full} ($IOI_{ration}>0.5$). This reveals that besides the phase adaptation (anticipation) discussed previously, there's also a frequency adaptation inside the beat that each subject can control in order to also accelerate the performance and avoid more tempo degradation.

Regime	Delay (ms)	Effect
1	1,3,6	reactive adaptation
2	10,15,21	natural (symmetric)
3	28,35,45,55	frequency adaptation plateau
4	66,78	strong frequency adaptation

Table 3.3: Clapping regimes for IOI_{ratio}.

3.6.5 Time of adaptation

Phase adaptation (anticipation) with different delay conditions is almost instantaneous. A qualitative analysis of Figure 3.20 reveals that after the first beat, when duo subjects are not yet aware of the underlying delay, all conditions cluster together. The first synchronization point (Sync point 1) is smaller as expected, i.e., there is less anticipation to push the beat, but then immediately the means cluster around the overall mean (Figure 3.11). In other words, it takes only one cycle of the pattern for the adaptation to already happen. Before the first pattern, subjects don't know (or feel) what the delay is, but after this the adaptation starts and doesn't increase.

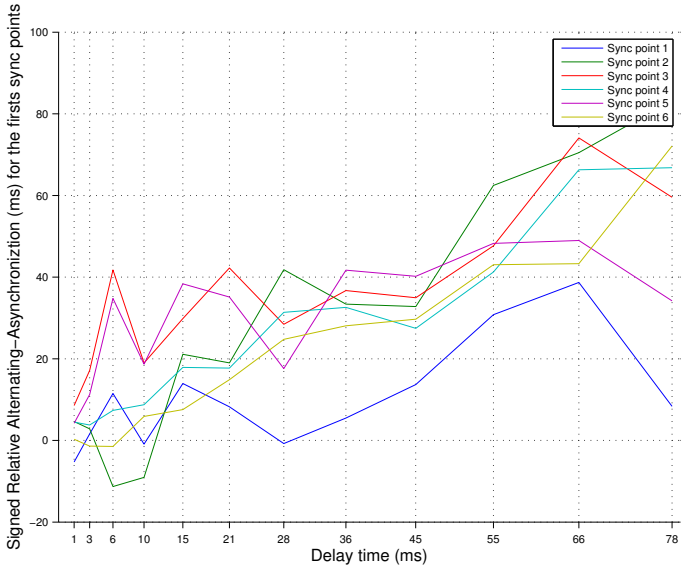

Figure 3.20: Mean relative alternating-asynchronization for the each of the first 6 sync points across duos. All conditions cluster together except the first synchronization point (Sync point 1) when duos are not yet aware of delay.

3.7 Conclusions

The experimental data analysis presented in this chapter reveals several features of the rhythmic interaction dynamics of human subjects. One of the first interesting discoveries was that for very small delay times, rhythmic interaction is not optimal as it would be expected. As was pointed out in Chafe et al. [18], there is a surprising acceleration for these very short delays. In Chafe et al. [19] and the present research, we have further investigated the causes of this effect.

We discussed that there is always a tendency to anticipate the beat of other performer. This is consistent with the tapping literature, and it is this anticipation for very small delays that produces this acceleration. We also see that humans are good at adapting when delay increases. For longer delays, anticipation also increases, but caps after 50 milliseconds.

The adaptation and the anticipation suggests an expectation mechanism in rhythmic interaction. In the next chapter, we will use adaptive coupled oscillators to model

these interaction dynamics. This model includes an expectation point that is consistent with the observed mechanics in this chapter. We will also incorporate an anticipation mechanism in these oscillators to consider the anticipation observed in the experimental data.

Chapter 4

Models for Rhythmic Coordination with Delay

4.1 Information Processing and Dynamic Systems Approaches to Rhythmic Synchronization

There are two main theoretical approaches for sensorimotor synchronization: information processing and dynamic systems theories [57, 54]. Information processing are usually associated with a system with a discrete-time series while dynamic systems theories model represent trajectories in a phase space.

Sensorimotor synchronization experimental studies have generally focused on humans synchronizing to a mechanical external stimulus. The setup is usually a tapping experiment in which subjects are instructed to follow the beat provided by a machine.

The present research focus is on the interaction of human performers. We turn now to a model that includes not only the tracking of rhythm, but also the interaction of both performers. We can use this model to obtain some insights on common observations in rhythmic performance with delay.

Before we describe the model for the synchronization between the two clappers with delay, we explain the adaptive oscillators used for rhythmic tracking. These

models are based on self-sustained adaptive oscillators. These oscillators have to synchronize with an external, discrete rhythmic stimulus, and thus provide an underlying model for meter perception.

4.2 Adaptive Oscillators to model meter perception

The class of oscillators used in this research are *adaptive oscillators* [45, 48, 64, 44, 46]. Other types of models, which are well described in Large [43], are linear models, and oscillator-level models [25]. We chose adaptive oscillators because they have been shown to work very well with discrete-time rhythms like the one analyzed in this research. Its underlying structure of expectation and adaptation provides an ideal way to include rhythmic generation and to couple two or more oscillators of its kind. Furthermore, this methodology can later be adapted to the more general canonical models and oscillator-level approaches.

Adaptive oscillators to track discrete pulses were originally described in Large and Kolen [45] and McAuley [48]. Several refinements have been proposed later [64, 44, 46], but the conceptual framework remains the same. They all model perception and attention in a way that is flexible enough to couple with variable tempo in music performance. They are all based on a circle map, an abstract model for nonlinear oscillation. They assume that musical rhythm is described by a series of discrete impulses. We first describe the mathematics of Sine circle maps and adaptive oscillators, and then show how to implement these in the context of the clapping experiment of this research.

4.2.1 Sine circle map

Circle maps are a good tool to do a qualitative description of oscillators with moderate external forcing. The details of this section draw from Pikovsky et al. [53, Chap. 7.3]. Large and Palmer [46] and Loehr et al. [47] also contains a summary description.

The equations of of motion for periodically forced oscillators can be reduced to a phase equation in the vicinity of the limit cycle:

$$\frac{d\phi}{dt} = \omega_0 + \varepsilon Q(\phi, t) \tag{4.1}$$

where ω_0 is the frequency of the self-sustained oscillator (without external forcing, the phase has to grow uniformly in time, i.e., $\frac{d\phi}{dt} = \omega_0$). For the specific case of a *sine circle map*, this equation becomes:

$$\phi_{n+1} = \phi_n + \eta + \varepsilon \sin \phi_n \tag{4.2}$$

with $\eta = \omega_0 T = \frac{2\pi}{T_0} T$, and T is the period of the forcing. This is a stroboscopic mapping with time interval T. When $\varepsilon = 0$, the system reduces to a stroboscopic observation:

$$\phi_{n+1} = \phi_n + \omega_0 T \tag{4.3}$$

If $\frac{T}{T_0} = \frac{p}{q}$ is rational, the system is periodic with period p.

This type of description can also be adapted for non-periodic external stimulus, specifically for the case of an external rhythm entraining an oscillator.

4.2.2 Adaptive oscillators conceptual framework

We propose a general conceptual framework for modeling rhythmic synchronization between musicians. The proposed methodology is independent of the specific adaptive oscillator model chosen. For consistency and simplicity we will use one of the original models proposed for adaptive oscillators, the Large & Kolen adaptive-oscillator [45].

Figure 4.1 illustrates the structure of an adaptive oscillator. They combine a limit-cycle oscillator with an energy pulse. The function of the pulse is to open a

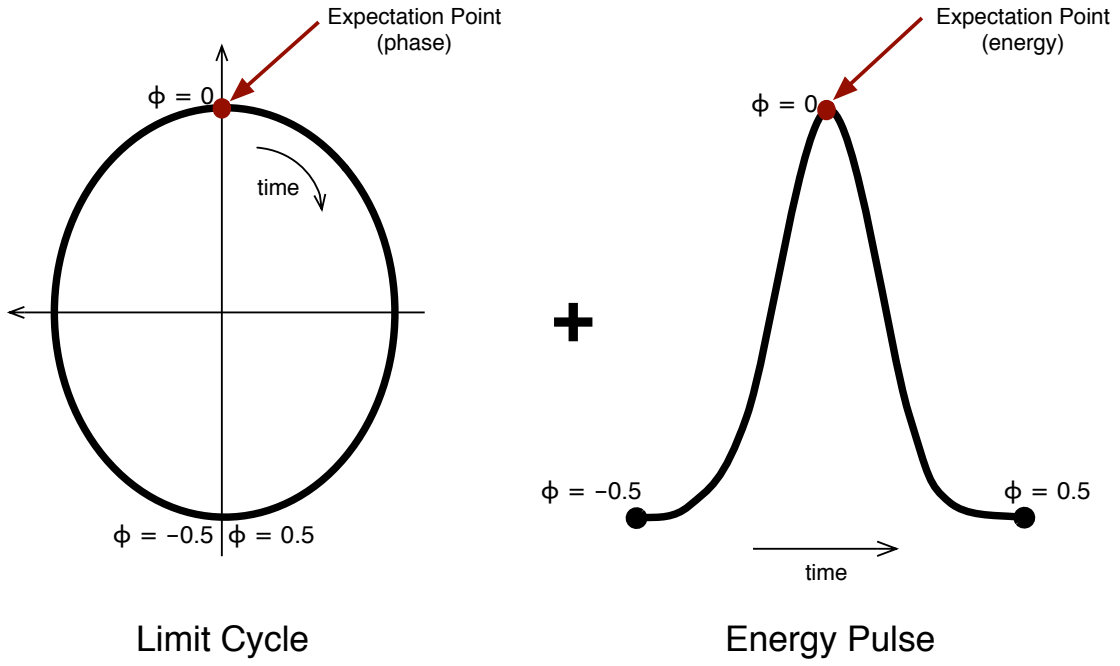

Figure 4.1: Conceptual framework for adaptive oscillators.

window of expectancy during which an external stimulus will be "heard". Anything that happens outside of the energy pulse is ignored in the adaptation. We now explain in detail the the Large & Kolen adaptive oscillator.

4.2.3 Large & Kolen Oscillator

The Large & Kolen adaptive-oscillator is described in Large and Kolen [45] and Large [42]. Although Large and Kolen [45] presents more details, it contains some errors and inaccuracies, and we prefer to follow the formulation presented in Large [42].

The basic periodic oscillatory unit adjusts its phase and period to lock with a discrete pulse's input stimulus $s(t)$. $s(t)$ represents the onsets of individual event notes. In our experiments, the onsets represent individual claps, with $s(t) = 1$ at each event and 0 at any other time.

The phase of the periodic oscillator is defined as:

$$\phi = \frac{t - t_x}{p}, \qquad with \qquad t_x - \frac{p}{2} \leq t < t_x + \frac{p}{2} \tag{4.4}$$

where t_x is the expectation point. Events that occur before t_x are early, and events that occur after are late.

Adaptive oscillators track input events by generating their own theoretical beat that instead of being a discrete event, has a width during which events are "listened". These are called *output pulses*, and are defined as:

$$o(t) = 1 + \tanh \gamma(\cos 2\pi\phi(t) - 1) \tag{4.5}$$

The parameter γ is an output gain, and it controls the width or skirt of the oscillator output. The oscillator will only entrain to signals that occur within its receptive field. A small γ means a wider receptive field, with the oscillator tolerating more variability in the signal. A large γ generates a narrow receptive field. In this case the oscillator will entrain to signals that are very close to it's internal period. This is useful in cases for rhythms that have a more complex internal structure to the beat, like syncopation and others. In our analysis, we will keep γ constant since we want to track each event.

Oscillators couple with an external stimulus by adjusting their phase. Adaptive oscillators can also adjust their frequency. Perfect synchronization happens when t_x is perfectly aligned with the input events $s(t)$. Therefore tracking is implemented based a "modified gradient descent procedure" [45, pag. 15] that dynamically minimizes the difference events and expectation points. Like standard gradient descent, the delta-rule is based on the first derivative of an error function:

$$E(t) = s(t)(1 - o(t)) \tag{4.6}$$

This error function $E(t)$ is non-zero only when there is a stimulus ($s(t) > 0$). It

also has a minimum when $o(t) = 1$. Minimization of this function produces perfect alignment between t_x and $s(t)$. The true gradient descent rule is modified in order to obtain proportional phase adjustments regardless of the nominal period of the oscillator.

The phase-tracking delta rule is implemented with:

$$\Delta t_x = \eta_1 s(t) \frac{p}{2\pi} \operatorname{sech}^2 \gamma(\cos 2\pi\phi(t) - 1) \sin 2\pi\phi(t) \tag{4.7}$$

And the period-tracking delta rule:

$$\Delta p = \eta_2 s(t) \frac{p}{2\pi} \operatorname{sech}^2 \gamma(\cos 2\pi\phi(t) - 1) \sin 2\pi\phi(t) \tag{4.8}$$

The coupling strengths, η_1 and η_2, determine the speed of adaptation for phase and period, respectively. The adaptation only occurs when events $s(t)$ are within the receptive field. When events arrive early (before t_x), the implication is that the oscillator is going too slow, so this causes a negative phase shift and a shortening of the period. Events after t_x imply that the oscillator has to slow down, so the phase is positively shifted and the period increased.

A simple example of this adaptive oscillator is shown in Figure 4.2. The stimulus is a periodic impulse with a period of 0.66 seconds. The oscillator starts with a period $p = 0.7$ seconds, $\eta_1 = 0.8$, $\eta_2 = 0.2$ and $\gamma = 2$. We see how the oscillator adapts it's periods for every event $s(t)$, until they are both aligned at 0.66 milliseconds. The oscillator is adaptive because after the tracking, in the absence of future events, stays at it's new period of 0.66 milliseconds and doesn't return to it's original period like many other kinds of oscillators.

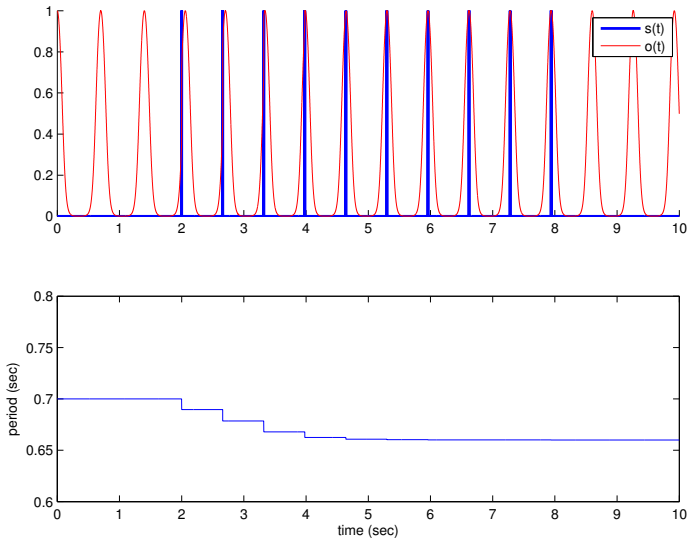

Figure 4.2: Example of a Large-Kolen oscillator responding to a simple periodic impulse stimulus. (adapted from Large and Kolen [45]).

4.3 Coupled Adaptive Oscillators for Rhythmic Tracking and Firing

In this section we use adaptive oscillators not only to track beats, but also as a generative (performing) mechanism. The underlying assumption is that *tracking* (following some rhythms) and *firing* (performing a beat) are part of the same self-sustained oscillator. Adaptive oscillators assume that performers have an internal expectation point where the next beat is more likely to occur. We expand this notion using the same adaptive oscillator for generative beat performance, i.e., performing the beat.

The tracking point is an idealized metronomic firing point, but musicians don't play metronomically. Instead, they are able to push the beat, or in turn, follow the expressive tempo variations of another performer in the ensemble. Depending on how confident the performer is, or on his intention of accelerating or decelerating the beat, the firing point will be before or after the expected point.

In this conceptual framework, the same adaptive oscillator serves two functions, expectation, and as a measure of confidence, i.e., when a player performs based on

his expectation.

Figure 4.3 shows the temporal receptive field of the Large & Kolen adaptive-oscillator, with the firing points example. The position of the firing point within the limit-cycle is a measure of confidence or leading in rhythmic synchronization. When performers are confident and want to push the beat, there is an anticipation in the firing point ($\phi < 0$). When performers are reactive, they are waiting for the *stimulator* and are less confident ($\phi > 0$).

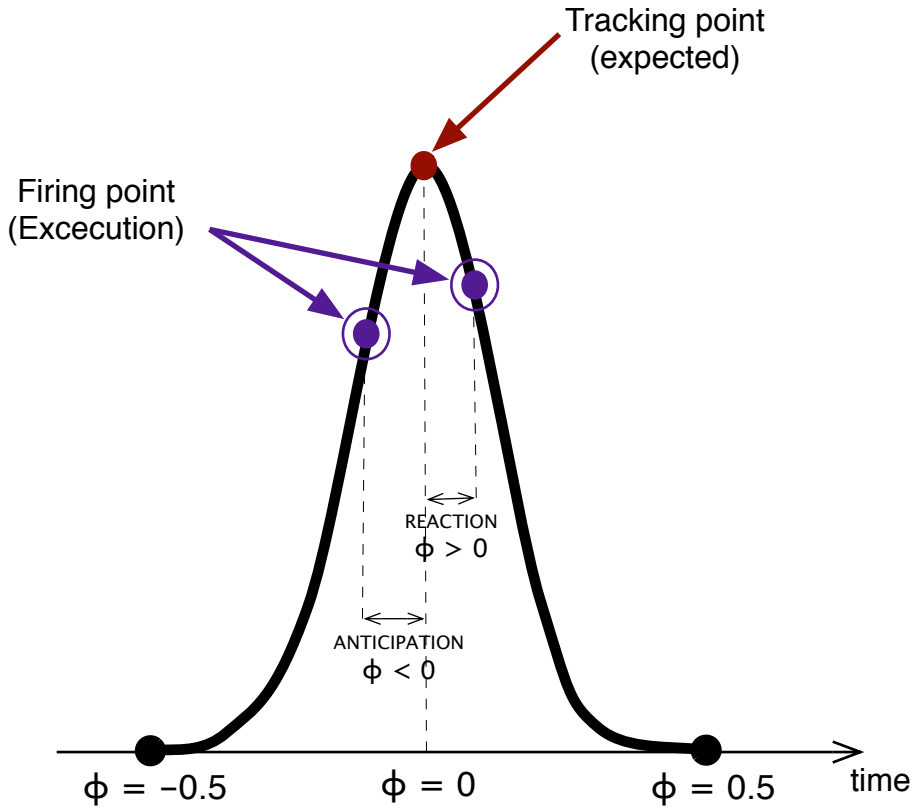

Figure 4.3: Integrated framework for tracking and firing rhythm. The figure shows the receptive field of an adaptive oscillator with tracking (expected) point and firing (execution) point.

4.3.1 Example: coupling two Large-Kolen oscillators

Figure 4.4 and Figure 4.5 show an example of a coupling between two oscillators using the rhythmic pattern of this research (see Figure 3.1). In this example, the firing point and the tracking point are the same ($\phi = 0$). Delay between oscillators is 20 ms.

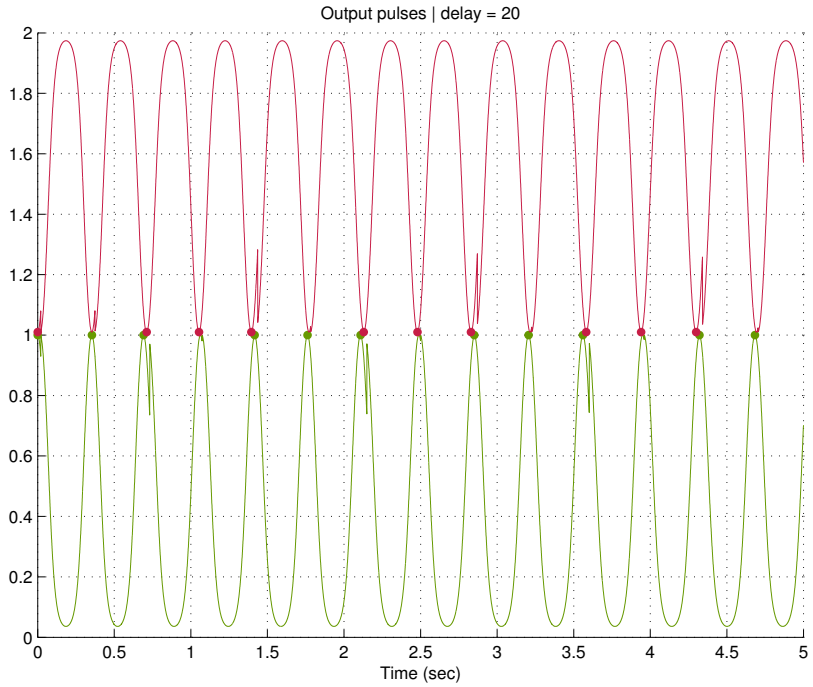

Figure 4.4: Synchronization example between two oscillators: output pulses.

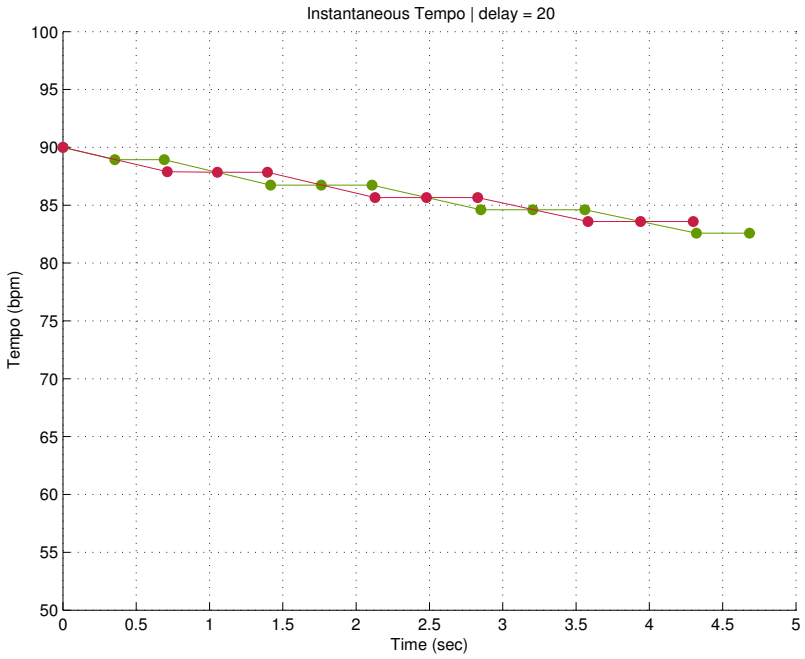

Figure 4.5: Synchronization example between two oscillators: instantaneous tempo.

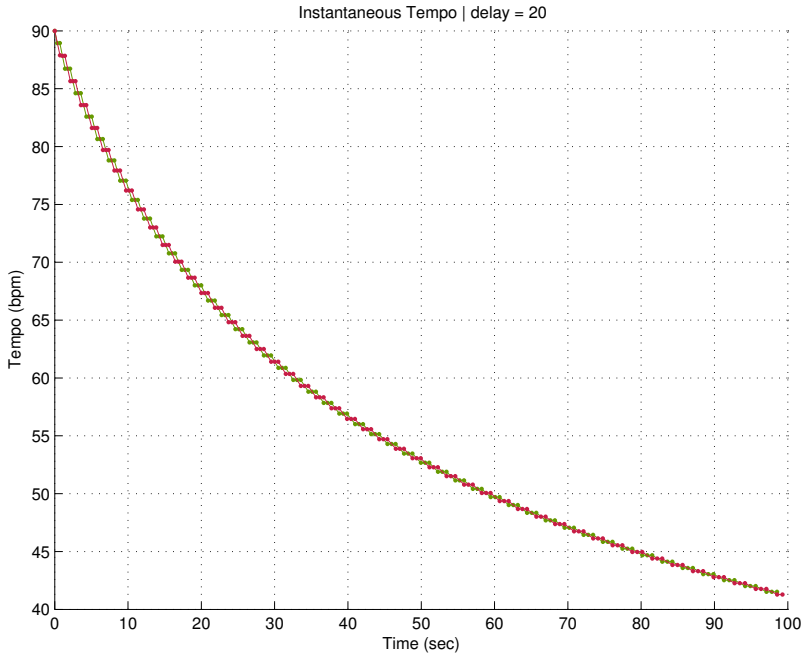

Figure 4.6: Synchronization example between two oscillators: instantaneous tempo.

We see in this example of coupled oscillators with no anticipation ($\phi = 0$) that tempo will slow down indefinitely, since expected points are always late due to the delay. The period of the oscillator will adapt and slow down until tempo is zero in the limit. In the next section we will discuss the different predicted tempos with several delay conditions, and how it is different with and without anticipation in firing.

4.4 Predicted Tempos

A simple memoryless model that instantly perceives tempo has been propose by Gurevich et al. [34]. In this model, each performer incorporates the perceived delay as a tempo decrease. Each performer waits for the other to clap and then immediately adjusts its tempo. There is no delay in the adjustment and no expectation point, so this model is purely theoretical and cannot be implemented computationally. Tempo M will therefore decrease according to:

$M(n) = \frac{60}{T_0+nd}$

where $T_0 = 60/M_o$ is the starting period in seconds, n is the quarter-note number and d is the delay time.

We compare the predictions of this model with the coupled oscillators model when there is no anticipation, i.e., firing and tracking are the same point.

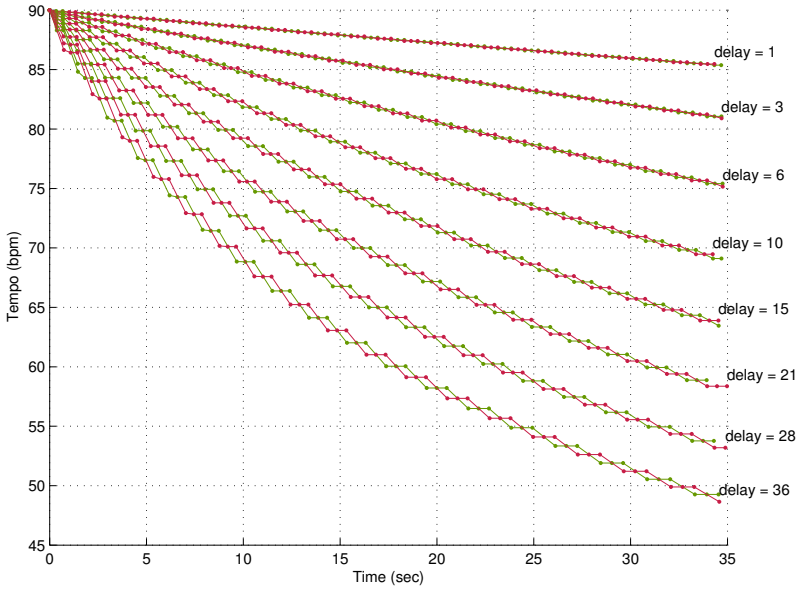

Figure 4.7: Synchronization example between two oscillators: predicted instantaneous tempo for all delay conditions.

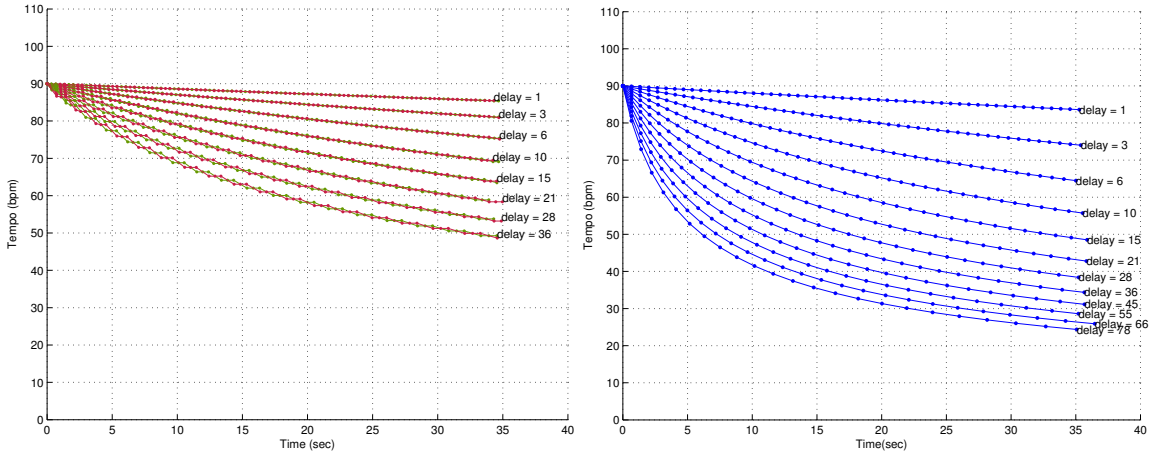

Figure 4.8: Comparison between oscillator model (left) and Gurevich model (right): predicted instantaneous tempo for all delay conditions.

In both models (Figure 4.8), tempo will decay indefinitely (eventually reaching 0). Humans outperform both models with a much more stable performance, with less tempo degradation. However, the nonlinearity of the coupled oscillators model

makes it significantly better than the simple oscillator model. The decay happens much slower. The nonlinearlity and expectation framework of the model account for this. In the next section we will discuss improvements to the model in its tracking and performance (See Figure 4.11 for a mean acceleration comparison).

4.5 Anticipation and reaction in rhythmic tracking

We want to understand the performance of a coupled oscillator model that incorporates distinct firing and tracking points (see Section 4.3). Figure 4.9 compares the signed relative alternating-asynchronization means for the experimental data with the theoretical compensation (anticipation) that would be needed to keep a constant tempo for the different delay conditions. The theoretical compensation is equal to the delay condition in the performance, i.e., each clapper anticipates by exactly the delay time. This will produce a steady tempo with no deterioration. When the experimental data (relative alternating-asynchronization) is greater than the theoretical compensation (deceleration in the Figure 4.9), tempo will speed up. When the opposite happens (acceleration in the Figure 4.9) tempo will slow down. This is consistent with the experimental data of Figure 3.7 and may account for at least part of the compensation process.

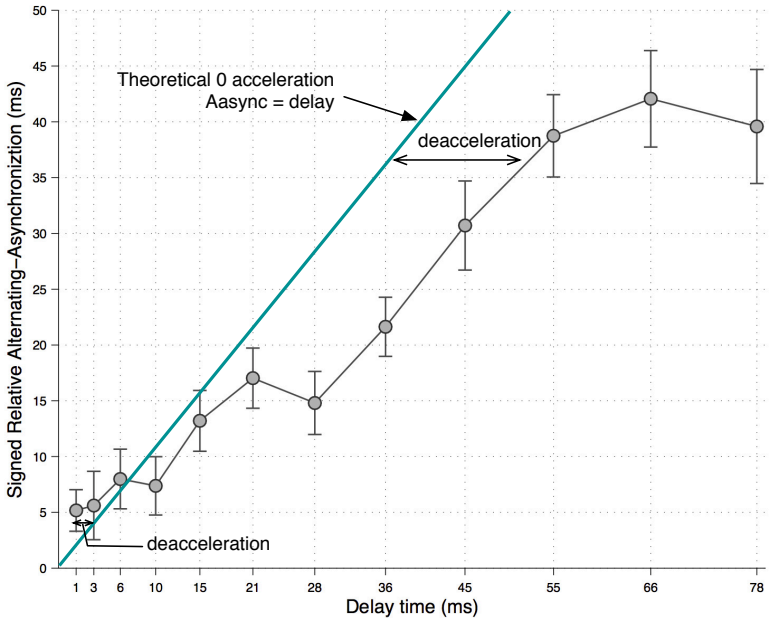

Figure 4.9: Relative alternating-asynchronization means and theoretical 0-acceleration values

We use the experimental data values from Figure 4.9 to simulate and predict tempo changes with the coupled oscillator model with anticipation (Figure 4.10). Compared to the previous model with no anticipation, we now see the acceleration for very small delays, and a better performance for longer delays. As predicted by the data, the ideal delay (zero tempo deterioration) will be between 6 and 15 ms.

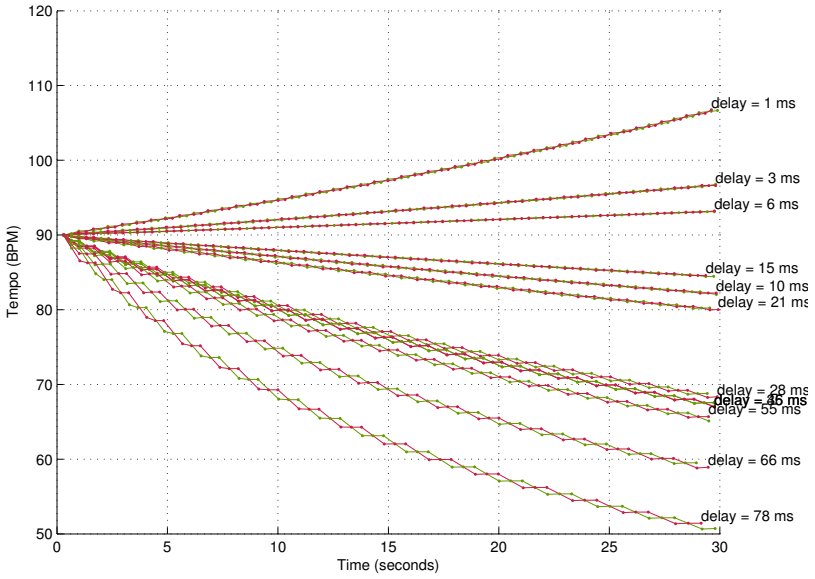

Figure 4.10: Predicted tempos for all delay conditions with coupled oscillators with anticipation. Anticipation values are taken from dataset analysis.

Figure 4.11 compares the mean acceleration from the experimental data with the mean acceleration produced by the coupled oscillators model and the Gurevich model. The oscillator model uses the relative alternating-asynchronization means as the anticipation value (in milliseconds) for the compensation. The figure shows that humans outperform this model at both extremes, for very small delays and for longer delays.

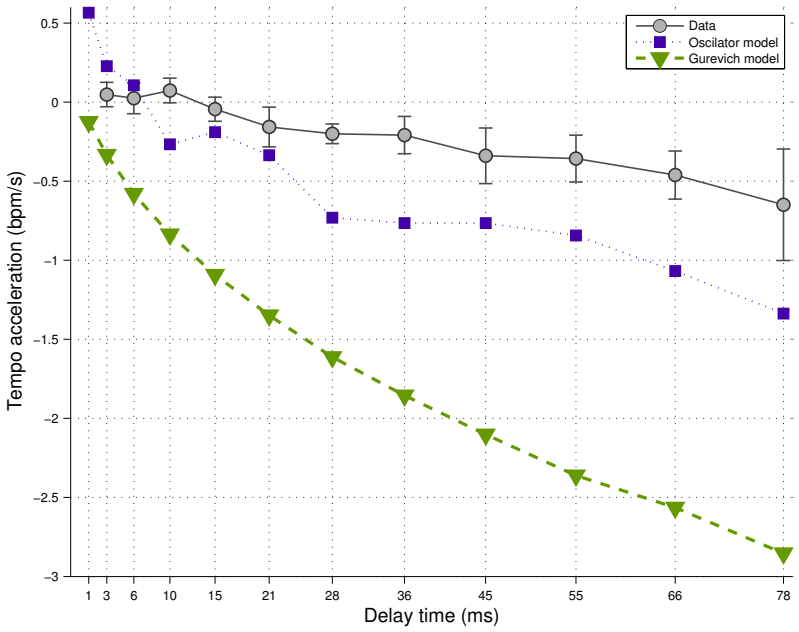

Figure 4.11: A single measure of tempo acceleration (its mean) is compared for experimental data (solid line), oscillator model with anticipation (dotted line) and Gurevich model (dashed line).

4.6 Tempo stabilization

All models considered so far will continue to increase or decrease tempo indefinitely. However, it has been shown that tempo tends to stabilize during a performance. This can happen after an initial acceleration or deceleration, depending on delay and starting tempo [11]. We have also observed this phenomenon in network performance, and in particular with some informal experiments done with the St. Lawrence String Quartet. A description of this experiment follows.

4.6.1 St. Lawrence String Quartet experiment[1]

Live music experiments have been made with the help of the St. Lawrence String Quartet, Stanford University's Ensemble-in-Residence. We present here a test between The Banff Centre, Alberta, Canada and CCRMA, Stanford University (June 2006) of a split string quintet. The quartet, located at Banff, was extended by a viola player performing from CCRMA. The sound was configured in such a way that only the room acoustics of the hall at Banff were heard. In order to achieve this effect, the viola player was placed in CCRMA's semi-anechoic room. Four speakers were positioned simulating the physical locations of the other members of the quintet (violin 1, violin 2, viola 1, violoncello) surrounding the viola performer. Each instrument in Banff was picked up using a directional microphone that was fed into the individual speakers in the semi-anechoic room, forming the quintet on the Stanford side. Two extra microphones where used to capture the acoustics of the hall. A microphone on the viola player at Stanford was sent to a speaker at Banff, forming the quintet on that side.

During this test, the audio RTT was measured and was approximately 50 ms. Neglecting any asymmetries, the performers where separated by a path of 25 ms (unidirectional). They performed Mozart's String Quintet in G minor, K. 516.

4.6.1.1 Reactions to the delay

Qualitative analysis of the performance, recordings and comments from the musicians are illustrative of issues that surface in such circumstances.

During a section between the viola (Stanford) and the violoncello (Banff) which constituted a rhythmic unison, the performance naturally slowed down. This effect can be predicted from the present study. Interestingly though, when used consciously by the musicians, it led to what was called the perfect ritard. The effect of "viola waits for cello" and "cello waits for viola" led to this controlled ritard, down to a certain point where the tempo stabilized. The amount of delay and the fact that this part of the performance included only two instruments (i.e., in "rhythmic power"

[1]Audio files from this experiment can be found at https://ccrma.stanford.edu/groups/soundwire/research/slsq/

equilibrium) can explain this phenomenon. The musicians were also able to maintain tempo when done consciously, i.e., when they were aware of the delay effect, they could define strategies to be able to play on time (without ritard). Strategies used are consistent with ones discussed earlier and in Bartlette et al. [6]. Again, these effects where observed when there was an equal balance in "ensemble power" or "rhythmic power" (only two musicians performing and with equally significant roles in terms of setting rhythm).

Another interesting effect observed dealt with the opposite situation; an unbalanced power between the two groups. Since four musicians were at Banff but only one at Stanford, the natural tendency (or strategy) of the Stanford viola player was to just follow the quintet. The quartet at Banff had to compensate and "struggle" with the slightly delayed sound coming in from Stanford. This struggle was clearly heard during the test; at one moment, the performers in Banff asked to "turn off" the viola (speaker) coming from Stanford. In that case, what was heard at Stanford was an impeccable performance. This is a real-time case of music minus one.

4.6.2 Tempo stabilization with phase analysis

To understand why tempo stabilizes after some degradation, we postulate that tempo anticipation and reaction are relative to instantaneous tempo and not absolute measures. It has also been shown that musicians tolerate more feedback delay for slower tempos [4]. In Section 3.5.2 we discussed how phase analysis measures *proportion* of alternating-asynchronization with respect to instantaneous tempo. For example, for slower performances, the same alternating-asynchronization (in time units) will result in a smaller ϕ, i.e., a smaller proportion of the beat cycle. If instead of using a fixed tempo anticipation, we use a relative one in terms of the instantaneous tempo and a fixed delay, tempo will naturally stabilize into an "optimal" condition.

Each combination of delay condition and phase anticipation will have a theoretical stable tempo. This tempo will be when the absolute anticipation (in milliseconds) matches the delay. The stable tempo will be:

$$p_{stable} = -\frac{delay}{\phi_{fire}}$$

For example, using a $delay = 28ms$ and $\phi_{fire} = -0.06$, the theoretical stable tempo will be 64.3 bpm. Using these same parameters with the coupled oscillator model and coupling parameters $\eta_1 = 0.2$ and $\eta_2 = 0.2$, we see that the stable tempo converges to a value close to the theoretical. The difference can be accounted by the nonlinearlity of the coupling. We will see in the next section how the coupling parameters influence the convergence tempo.

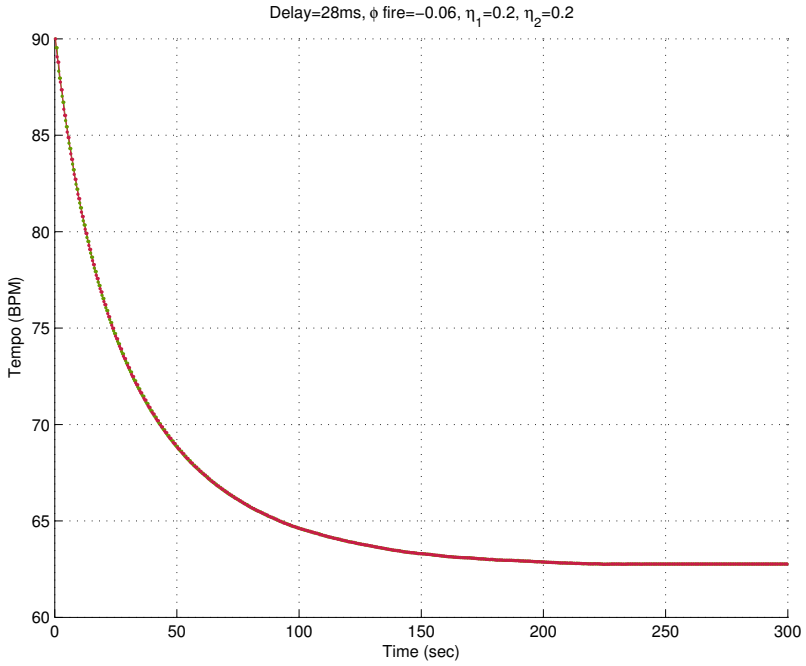

Figure 4.12: Tempo stabilization example with coupled oscillators

4.6.3 Phase and frequency coupling parameters

The coupled oscillators discussed in this research have two relevant coupling parameter, phase coupling (η_1) and frequency adaptation (η_2). The speed of adaptation and the stability of the coupling depends on the value of these coupling parameters. Figure 4.13 show an example with larger coupling strengths , $\eta_1 = 0.3$ and $\eta_2 = 0.8$. We remind the reader that the previous theoretical tempo stabilization was 64.3 bpm. We see that in this example, for these coupling parameters, the stabilization occurs at a much higher tempo (around 73 bpm), and that it converges faster than for the

example shown in Figure 4.12. There is also a continuous adaptation after tempo has stabilized, illustrated by the "zigzagging" of the tempo curves. This seems consistent with the behavior of the human subjects, but it should be pointed out that this is not the finer granularity for internal sub-beat proportions from inside the pattern observed in Figure 3.16. To account for that effect, the model would have to include a second order adaptation that is not in the scope of this research.

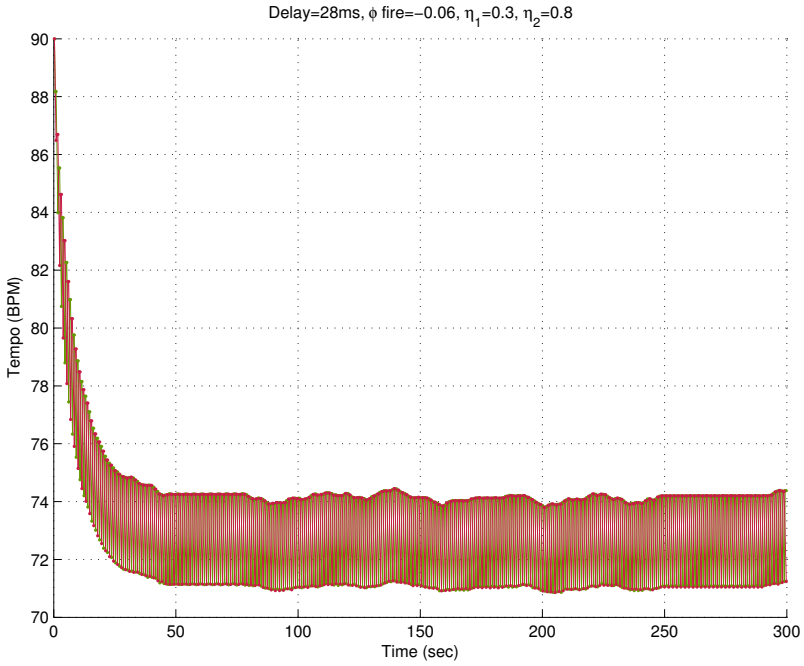

Figure 4.13: Synchronization example between two oscillators: instantaneous tempo

Figure 4.14 and 4.15 show the tempo convergence for different combinations of coupling strengths η_1 and η_2. A different set of experimental data is needed to tune these parameters, but we can comment on some of the observed behavior. When η_1 is small, tempo converges to very close to the theoretical value, independent of the value of η_2. This means that η_1 has an effect on the speed of the adaptation, but not on the value of the stable tempo itself. With larger η_2, however, convergence tempo increases dramatically, and in extreme cases is even higher than the starting tempo of 90 bpm.

We discussed in Section 3.5.5 that human subjects duos are not symmetric, that

sometimes a subject is in a stronger role, that of leader. A future direction in this research is to understand such interaction dynamics and to analyze in detail duos in order to discover interaction patterns and imbalances that can be modeled with asymmetric coupling strengths.

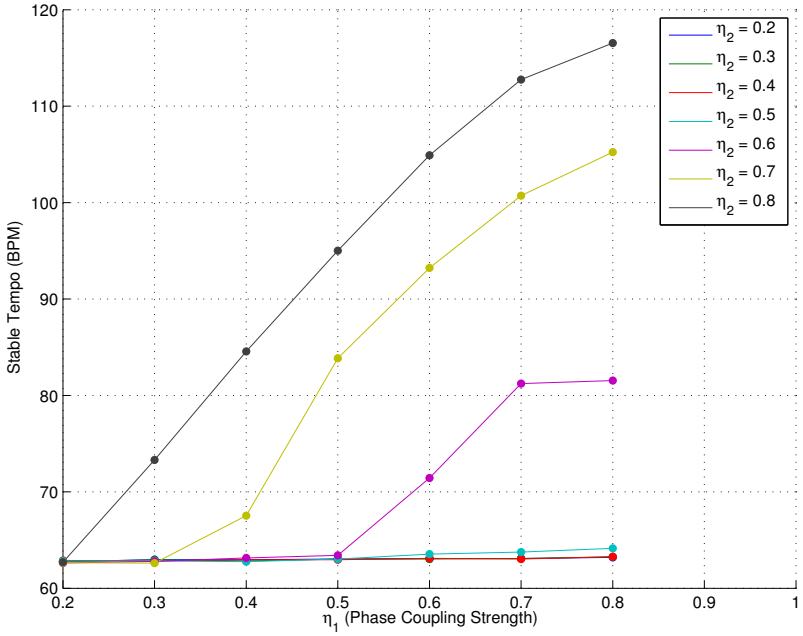

Figure 4.14: Coupling parameters dependency for delay=28ms and $\phi_{fire} = -0.06$

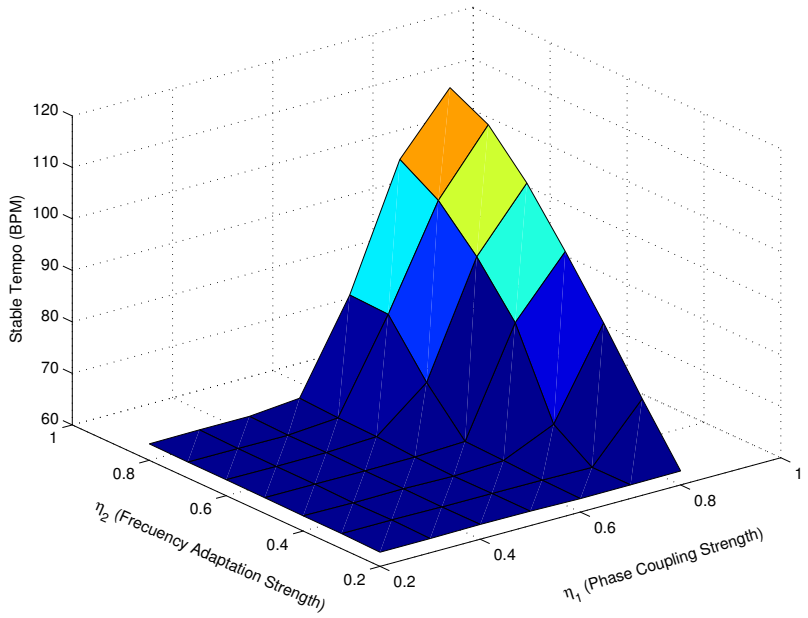

Figure 4.15: 3D plot for coupling parameters dependency for delay=28ms and $\phi_{fire} = -0.06$

4.7 Conclusions

We presented a model of coupled oscillators for the experimental data from Chapter 3. Specifically, we used the Large-Kolen adaptive oscillators, and integrated into it tracking (listening) and firing (performing) of the beat. We included in this model an anticipation and reaction parameter that is consistent with the experimental data in the literature and in this research. Humans anticipate the beat of the other performer and always play a little bit ahead of it. This effect becomes stronger as delay increases, but caps at an average of around 40 milliseconds. This seems to be a strong upper limit, but experiments with larger delays than the ones presented must be performed in order to validate this theory.

When performing with delay, one goal is to be able to keep a steady tempo and avoid timing degradation. The model presented performs better than previous attempts, but humans are still better at this task. Second-order adaptation in a finer inter-beat granularity that are not accounted in the presented model may explain this.

Our model includes an anticipation parameter that is static, and a better understanding of this prediction mechanism is a desirable goal to improve the performance of the model.

A larger set of experiments needs to be performed in order to asses in more detail the implications of the model. In particular, a test between a robot clapper (the model) and a human subject would give insights on interactions which are still missing and need to be accounted for in the oscillator coupling.

Chapter 5

Conclusions and Future Work

5.1 Overview

Musical Networked Performance has motivated the research of the effects of time delay in music performance. We are in an exciting time when telecommunications are approaching the theoretical limit of the speed of light, and can be as fast as a couple hundred milliseconds between opposite parts of the planet. Music requirements are much more stringent than conversation, and these very small delays are still problematic for music performance. To understand and quantify the strategies humans use to deal with these types of latencies are the goals of this research.

We analyzed a set of experimental data in which duo subjects perform a simple interlocked pattern, under a set of delay conditions that ranged from 1 to 78 milliseconds. This range is commonly encountered in today's transcontinental concerts. The analysis of the data reveals that there's an innate tendency to anticipate the beat of the other performer in rhythmic interaction. This explains why for very small delays (delays that may be considered "unnatural" from an evolutionary point of view) there is an acceleration. For longer delays, humans increase the adaptation mechanism (alternating asynchronization) to avoid tempo degradation. We saw that there is a limit to the amount of adaptation, however, and that this caps in average at around 50 milliseconds. This explains also the tempo degradation observed for longer delays.

To model delayed interaction dynamics, we presented a coupled oscillator model

using adaptive oscillators. In this model we incorporated rhythmic execution ("firing" in oscillator terminology) to the tracking mechanism of the oscillator. With this addition we can couple two oscillators using the same rhythmic pattern that was discussed in the experimental analysis (Chapter 3). Firing in the oscillator incorporates an anticipation and reaction parameter. This addition is shown to improve the tempo degradation that happens with delay, but humans outperform the model. The model presented does not account for finer granularity of internal sub-beat proportions inside the pattern, which is possibly an explanation of why humans are better at it.

We also show how the coupled oscillator synchronization mechanism with anticipation can explain the tolerances for slower tempi, and we speculate that relative anticipation (in phase units) may have a role in why tempo stabilizes after some time and doesn't degrade indefinitely as in previous models.

We finished by opening a window to further research that is needed to better understand the interaction dynamics and the coupling behavior, showing how larger frequency adaptation coupling strengths have a great influence in the long-term dynamics interaction of the pattern analyzed.

5.2 Future Research and Directions

An immediate extension of the current research would be to implement and try a realtime oscillator with the anticipation mechanism discussed to deploy an experiment between humans and a machine. This experiment will reveal the strengths and limitations of the current model, and will facilitate obtaining more data to validate and fine tune some of its parameters.

We discussed how the internal sub-beat proportions may have an influence in the overall interaction of a performing duo. To further understand this, it is necessary to experiment with different types of musical meters, e.g., ternary and others. This should illuminate similarities and differences of the strategies adopted that are not explicit in our dataset. This new data may help to create a model that incorporates multiple oscillators to track different metrical levels.

Our model doesn't consider other acoustical/musical parameters like attack duration, loudness, reverberating environments, accented patterns, and others that are relevant to the psychoacoustics of synchronization. These should be incorporated in the future.

Speech synchronization [22] is also an emerging field of research. The main difference between speech and rhythmic synchronization seems to be the how the system (system in this case is the interacting humans) deals with error. In rhythmic synchronization, humans are able to recover usually very fast. By contrast, in speech synchronization errors are catastrophic, with the whole performance coming to a halt. In speech synchronization the interlocking mechanism may be more fluid but is also less robust. There's much promising research that can be done to understand synchronization in this different context.

Finally, although complete musical prediction is still in its infancy and may never be completely possible, the anticipation and expectation oscillator interaction dynamics suggest that humans already do this kind of operation, at least rhythmically, and having a machine performing a similar operation would be feasible. The delay can be used in musical ways [14], and music can be specifically created in a predictive way to "beat" the telecommunications delay. This has practical implications for applications like distributed musical games (like RockBand), but also opens new musical and technical windows that real-time network performance makes us think and dream about.

Bibliography

[1] NINJAM: Realtime music collaboration software, 2008. URL http://www.ninjam.com/.

[2] Rockband, 2013. URL http://www.rockband.com/.

[3] A. Barbosa. Displaced soundscapes: A survey of network systems for music and sonic art creation. *Leonardo Music Journal*, 13:53–59, 2003.

[4] A. Barbosa and J. Cordeiro. The influence of perceptual attack times in networked music performance. In *Audio Engineering Society Conference: 44th International Conference: Audio Networking*, 11 2011. URL http://www.aes.org/e-lib/browse.cfm?elib=16133.

[5] R. Bargar, S. Church, A. Fukuda, J. Grunke, D. Keislar, B. Moses, B. Novak, B. Pennycook, Z. Settel, J. Strawn, P. Wiser, and W. Woszczyk. AES white paper: Networking audio and music using Internet2 and next-generation Internet capabilities. Technical report, AES: Audio Engineering Society, 1998.

[6] C. Bartlette, D. Headlam, M. Bocko, and G. Velikic. Effect of network latency on interactive musical performance. *Music Perception*, 24(1):49–62, 2006.

[7] J. Bischoff, R. Gold, and J. Horton. Music for an interactive network of microcomputers. *Computer Music Journal*, 2(3):24–29, Dec. 1978.

[8] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 1 edition, Oct. 2007. ISBN 0387310738.

[9] N. Bouillot. nJam user experiments: Enabling remote musical interaction from milliseconds to seconds. In *NIME '07: Proceedings of the 7th international conference on New Interfaces for Musical Expression*, pages 142–147, New York, New York, 2007. ACM. doi: http://doi.acm.org/10.1145/1279740.1279766.

[10] N. Bouillot and J. R. Cooperstock. Challenges and performance of high-fidelity audio streaming for interactive performances. In *NIME '09: Proceedings of the 9th international conference on New Interfaces for Musical Expression*, pages 135–140, Pittsburgh, PA, USA, 2009.

[11] T. Brochier, J. B. L. Smith, and C. Elaine. Can an overly slow initial tempo counteract the deceleration caused by auditory delay? In *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami, Oct. 2011.

[12] J.-P. Cáceres. SoundWIRE research: St. Lawrence String Quartet, 2008. URL http://ccrma.stanford.edu/groups/soundwire/research/.

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

[14] J.-P. Cáceres, R. Hamilton, D. Iyer, C. Chafe, and G. Wang. To the edge with china: Explorations in network performance. In *ARTECH 2008: Proceedings of the 4th International Conference on Digital Arts*, pages 61–66, Porto, Portugal, 2008. ISBN 978-989-95776-3-3.

[15] A. Carôt, T. Hohn, and C. Werner. Netjack—remote music collaboration with electronic sequencers on the Internet. In *Proceedings of the Linux Audio Conference*, Parma, Italy, 2009.

[16] C. Chafe. Statistical pattern recognition for prediction of solo piano performance. In *Proceedings of International Computer Music Conference*, Thessaloniki, Greece, Sept. 1997.

[17] C. Chafe, S. Wilson, R. Leistikow, D. Chisholm, and G. Scavone. A simplified approach to high quality music and sound over IP. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, pages 159–164, Verona, Italy, Dec. 2000.

[18] C. Chafe, M. Gurevich, G. Leslie, and S. Tyan. Effect of time delay on ensemble accuracy. In *Proceedings of the International Symposium on Musical Acoustics*, Nara, Japan, 2004. Kyoto: Musical Acoustics Research Group, The Acoustical Society of Japan.

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

[20] E. Chew, A. Sawchuk, C. Tanoue, and R. Zimmermann. Segmental tempo analysis of performances in performer-centered experiments in the distributed immersive performance project. In *Proceedings of International Conference on Sound and Music Computing '05 (SMC05)*, Salerno, Italy, Nov. 2005.

[21] N. Collins. *Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems*. PhD thesis, University of Cambridge, 2006.

[22] F. Cummins. Periodic and aperiodic synchronization in skilled action. *Frontiers in Human Neuroscience*, 5:170, 2011. doi: 10.3389/fnhum.2011.00170. URL http://www.frontiersin.org/human_neuroscience/10.3389/fnhum.2011.00170/abstract.

[23] P. L. Divenyi. The times of Ira Hirsh: Multiple ranges of auditory temporal perception. *Seminars in Hearing*, 25(3):229–239, 2004. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=16479266&query_hl=22&itool=pubmed_ExternalLink.

[24] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2 sub edition, Oct. 2000. ISBN 0471056693.

[25] D. S. Eck. *Meter Through Synchrony: Processing Rhythmical Patterns with Relaxation Oscillators*. PhD thesis, Indiana University, 2002.

[26] S. Farner, A. Solvang, A. Sæbø, and U. P. Svensson. Ensemble hand-clapping experiments under the influence of delay and various acoustic environments. *J. Audio Eng. Soc.*, 57(12):1028–1041, 2009.

[27] D. Fober, Y. Orlarey, and S. Letz. Real time musical events streaming over internet. In *Web Delivering of Music, 2001. Proceedings. First International Conference on*, pages 147–154, 2001.

[28] D. Fober, S. Letz, and Y. Orlarey. Clock skew compensation over a high latency network. In *Proceedings of International Computer Music Conference*, pages 548–552, Gothenburg, Sweden, 2002.

[29] M. B. Gardner. Historical background of the Haas and/or precedence effect. *Journal of the Acoustical Society of America*, 43:1243–1248, June 1968.

[30] A. Gates, J. L. Bradshaw, and N. C. Nettleton. Effect of different delayed auditory feedback intervals on a music performance task. *Perception & Psychophysics*, 15(1):21–25, 1974.

[31] M. Goto and R. Neyama. Open remoteGIG: An open-to-the-public distributed session system overcoming network latency. *Transactions of Information Processing Society of Japan*, 43(2):299–309, 2002. (in Japanese).

[32] M. Goto, R. Neyama, and Y. Muraoka. RMCP: Remote music control protocol - design and applications. In *Proceedings of International Computer Music Conference*, pages 446–449, Thessaloniki, Greece, 1997.

[33] S. Gresham-Lancaster. The aesthetics and history of the Hub: The effect of changing technology on network computer music. *Leonardo Music Journal*, 8: 39–44, 1998.

[34] M. Gurevich, C. Chafe, G. Leslie, and S. Tyan. Simulation of networked ensemble performance with varying time delays: Characterization of ensemble accuracy. In *Proceedings of International Computer Music Conference*, Miami, 2004.

[35] G. Hajdu. Quintet.net: An environment for composing and performing music on the Internet. *Leonardo*, 38(1):23–30, Feb. 2005. ISSN 0024-094X.

[36] S. Hinton. Jasager, Der. In L. Macy, editor, *Grove Music Online*. 2007. URL http://www.grovemusic.com. (Accessed 26 March 2007).

[37] I. J. Hirsch. Auditory perception of temporal order. *The Journal of the Acoustical Society of America*, 31(6):759–767, June 1959.

[38] D. A. Hounshell. Elisha Gray and the telephone: On the disadvantages of being an expert. *Technology and Culture*, 16(2):133–161, Apr. 1975.

[39] T. Inagaki and J. Stahre. Human supervision and control in engineering and music: similarities, dissimilarities, and their implications. *Proceedings of the IEEE*, 92:589–600, 2004. ISSN 0018-9219. doi: http://dx.doi.org/10.1109/JPROC.2004.825876.

[40] Internet2. Internet2, 2010. URL http://www.internet2.edu/.

[41] G. Johannsen. Human supervision and control in engineering and music - foundations and transdisciplinary views. *Journal of New Music Research*, 31(3):179–190, September 2002. doi: http://dx.doi.org/10.1076/jnmr.31.3.179.14187.

[42] E. W. Large. Beat tracking with a nonlinear oscillator. In *Working Notes of the IJCAI Workshop on AI and Music*, pages 24–31, Montreal, 1995.

[43] E. W. Large. Resonating to musical rhythm: Theory and experiment. In S. Grondin, editor, *Psychology of Time*, chapter 6, pages 189–231. Emerald Group Publishing Ltd, Nov. 2008. ISBN 0080469779.

[44] E. W. Large and M. R. Jones. The dynamics of attending: How people track time-varying events. *Psychological Review;Psychological Review*, 106(1):119–159,

1999. ISSN 1939-1471(Electronic);0033-295X(Print). doi: 10.1037/0033-295X.106.1.119.

[45] E. W. Large and J. F. Kolen. Resonance and the perception of musical meter. *Connection Science*, 6(2):177–208, 1994. ISSN 0954-0091. doi: 10.1080/09540099408915723.

[46] E. W. Large and C. Palmer. Perceiving temporal regularity in music. *Cognitive Science*, 26(1):1–37, 2002. ISSN 1551-6709. doi: 10.1207/s15516709cog2601_1. URL http://dx.doi.org/10.1207/s15516709cog2601_1.

[47] J. D. Loehr, E. W. Large, and C. Palmer. Temporal coordination and adaptation to rate change in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, 37(4):1292–1309, 2011. ISSN 1939-1277(Electronic);0096-1523(Print). doi: 10.1037/a0023102.

[48] J. D. McAuley. *On the Perception of Time as Phase: Toward an Adaptive-Oscillator Model of Rhythm*. PhD thesis, Indiana University, 1995.

[49] A. D. Patel. *Music, Language, and the Brain*. Oxford University Press, USA, 1 edition, Dec. 2007. ISBN 0195123751.

[50] A. D. Patel, J. R. Iversen, Y. Chen, and B. H. Repp. The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research*, 163 (2):226–238, May 2005. ISSN 0014-4819. doi: 10.1007/s00221-004-2159-8. URL http://www.ncbi.nlm.nih.gov/pubmed/15654589. PMID: 15654589.

[51] P. Q. Pfordresher and C. Palmer. Effects of delayed auditory feedback on timing of music performance. *Psychological Research*, 16:71–79, 2002.

[52] J. Pierce. Hearing in time and space. pages 89–103. MIT Press, Cambridge, MA, USA, 1999. ISBN 0262531909.

[53] A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press, 1 edition, May 2003. ISBN 052153352X.

[54] J. Pressing. The referential dynamics of cognition and action. *Psychological Review*, 106(4):714–747, 1999. ISSN 0033-295X (Print); 1939-1471 (Electronic). doi: doi:10.1037/0033-295X.106.4.714.

[55] R. A. Rasch. Timing and synchronization in ensemble performance. In J. A. Sloboda, editor, *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, chapter 4, pages 70–90. Oxford University Press, New York, 1988.

[56] A. B. Renaud, A. Carôt, and P. Rebelo. Networked music performance: State of the art. In *Proceedings of the AES 30th International Conference*, Saariselkä, Finland, 2007.

[57] B. H. Repp. Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6):969–992, 2005.

[58] M. Sarkar. Tablanet: a real-time online musical collaboration system for indian percussion. Master's thesis, MIT Media Lab, Aug. 2007.

[59] M. Sarkar and B. Vercoe. Recognition and prediction in a network music performance system for Indian percussion. In *NIME '07: Proceedings of the 7th international conference on New Interfaces for Musical Expression*, pages 317–320, New York, NY, USA, 2007. ACM.

[60] A. Schloss. *On the Automatic Transcription of Percussive Music: From Acoustic Signal to High Level Analysis*. PhD thesis, Stanford University, 1985.

[61] T. Sheridan. Musings on music making and listening: supervisory control and virtual reality. *Proceedings of the IEEE*, 92:601–605, 2004. ISSN 0018-9219. doi: http://dx.doi.org/10.1109/JPROC.2004.825879.

[62] SoundWIRE Group. SoundWIRE research group at CCRMA, Stanford University, 2010. URL http://ccrma.stanford.edu/groups/soundwire/.

[63] A. Tanaka. Interaction, experience and the future of music. In K. O'Hara and B. Brown, editors, *Consuming Music Together: Social and Collaborative*

Aspects of Music Consumption Technologies, volume 35 of *Computer Supported Cooperative Work*, pages 267–288. Springer, London, 2006.

[64] P. Toiviainen. An interactive MIDI accompanist. *Computer Music Journal*, 22 (4), Dec. 1998. ISSN 0148-9267. doi: 10.2307/3680894. URL http://www.jstor.org/stable/3680894.

[65] G. Weinberg. The aesthetics, history, and future challenges of interconnected music networks. In *Proceedings of International Computer Music Conference*, pages 349–356, Göteborg, Sweden, 2002.

[66] G. Weinberg. Interconnected musical networks: Toward a theoretical framework. *Computer Music Journal*, 29(2):23–39, 2005.

[67] A. Xu and J. R. Cooperstock. Real-time streaming of multichannel audio data over Internet. In *Proceedings of the 108th Convention of the Audio Engineering Society*, pages 627–641, Paris, 2000.