

## INTERPERSONAL ENTRAINMENT IN MUSIC PERFORMANCE: THEORY, METHOD, AND MODEL

---

MARTIN CLAYTON, KELLY JAKUBOWSKI,  
TUOMAS EEROLA  
*Durham University, Durham, United Kingdom*

PETER E. KELLER  
*Western Sydney University, Sydney, Australia*

ANTONIO CAMURRI, GUALTIERO VOLPE, &  
PAOLO ALBORNO  
*University of Genoa, Genoa, Italy*

**INTERPERSONAL MUSICAL ENTRAINMENT**—TEMPORAL synchronization and coordination between individuals in musical contexts—is a ubiquitous phenomenon related to music’s social functions of promoting group bonding and cohesion. Mechanisms other than sensorimotor synchronization are rarely discussed, while little is known about cultural variability or about how and why entrainment has social effects. In order to close these gaps, we propose a new model that distinguishes between different components of interpersonal entrainment: *sensorimotor synchronization*—a largely automatic process manifested especially with rhythms based on periodicities in the 100–2000 ms timescale—and *coordination*, extending over longer timescales and more accessible to conscious control. We review the state of the art in measuring these processes, mostly from the perspective of action production, and in so doing present the first cross-cultural comparisons between interpersonal entrainment in natural musical performances, with an exploratory analysis that identifies factors that may influence interpersonal synchronization in music. Building on this analysis we advance hypotheses regarding the relationship of these features to neurophysiological, social, and cultural processes. We propose a model encompassing both synchronization and coordination processes and the relationship between them, the role of culturally shared knowledge, and of connections between entrainment and social processes.

*Received: September 9, 2019, accepted July 25, 2020.*

**Key words:** entrainment, sensorimotor synchronization, interpersonal coordination, cross-cultural, ethnomusicology

**E**NTRAINMENT DESCRIBES THE TEMPORAL dynamics of interacting rhythmic systems. The essence of interpersonal musical entrainment (IME) is the interaction and coordination of human beings mediated by sound and movement. Although interpersonal entrainment has been studied in various contexts including conversation (Richardson, Dale, & Shockley, 2008; Shockley, Santana, & Fowler, 2003), spontaneous clapping (Néda, Ravasz, Brechet, Vicsek, & Barabási, 2000), and sports (Cohen et al., 2010), music is an important focus for this study since it affords forms of coordination that are particularly precise, complex, periodic, and flexible, and that appear to be essential to social or ritual events that participants find richly and intensely meaningful (Clayton et al., 2005). Music and dance are important contexts in which the potential of interpersonal entrainment to generate affect and reinforce social bonds is exploited to the fullest degree. Understanding the processes underlying this phenomenon and the ways in which this coordination might vary cross-culturally is vital to advance understanding of interpersonal human coordination and collaborative action in general.<sup>1</sup>

The aim of this paper is to present analyses of interpersonal musical entrainment considering a range of musical and cultural factors, and in the final section to present a new model of IME that compares and contrasts two distinct components, labeled here *synchronization* and *coordination*. By *sensorimotor synchronization* (SMS) we refer to a largely automatic process of temporal alignment manifested especially with periodic

---

<sup>1</sup> The following abbreviations are used in this article: ADAM (ADaptation and Anticipation Model), BPS (Beat Perception and Synchronization), CRQA (Cross-Recurrence Quantification Analysis), CWT (Cross-Wavelet Transform), DAT (Dynamic Attending Theory), ES (Event Synchronization), IEMP (Interpersonal Entrainment in Music Performance, the title of the research project enabling this research), IME (Interpersonal Musical Entrainment), IOI (Interonset Interval), OF (Optical Flow), ROI (Region of Interest), SET (Scalar Expectancy Theory), SMS (Sensorimotor Synchronization), and WT (Wavelet Transform). The following abbreviations refer to these individual corpora analyzed in the section *Measuring Entrainment in Musical Ensembles*: NIR (North Indian Raga), UC (Uruguayan Candombe), MJ (Malian Jembe), CSS (Cuban Son and Salsa), TS (Tunisian Stambeli), ESQ (European String Quartet).

rhythms in the 100–2000 ms timescale; by *coordination*, we point to a separate process of structural alignment of individual parts, which extends over longer timescales and is more accessible to conscious appraisal and control. Synchronization (SMS) is the subject of a considerable body of literature, much of it based on experimental tapping studies (Repp, 2005; Repp & Su, 2013). Coordination is also the subject of a significant body of writing—although the phenomenon is less consistently named—for example, in studies of coordinated body movements (Clayton, 2007b; Eerola, Jakubowski, Moran, Keller, & Clayton, 2018; Ginsborg & King, 2009; Williamon & Davidson, 2002). Past studies have either concentrated on one or the other exclusively or have considered both without drawing attention to the distinction: few have considered both components or explored their relationship (see e.g., Bishop & Goebel 2017; Keller, 2014; Ragert, Schroeder, & Keller 2013). Building on the foundation of past research, we draw out the distinctions between these phenomena much more explicitly. We propose that doing so, and beginning to explore the relationship between the two components, brings greater clarity to the study of IME.

The work of bringing the complementary components of synchronization and coordination together is necessarily an interdisciplinary endeavor, since different disciplines have focused to a greater extent on one or the other. SMS clearly lies within the purview of experimental psychology, relating as it does to topics such as time perception and motor control. Although individuals synchronize with each other in dyads and larger groups, much of the literature analyzes individuals responding to pacing signals (Repp, 2005; Repp & Su, 2013): this focus is carried through into descriptions of musical entrainment that assume a single listener entraining to a musical stimulus, rather than a group of individuals mutually adjusting in a social context. The phenomenon we term coordination concerns the management of performance and shared understanding of musical structure and process and is often the topic of conscious control and explicit negotiation between individuals. While also of interest to psychology in discussion of topics such as prosociality and entitativity, and more broadly under the rubric joint action (Cirelli, Einarson, & Trainor, 2014; Keller, 2008; Keller & Appel, 2010; Kirschner & Tomasello, 2010; Wiltermuth & Heath, 2009), coordination also attracts interest from anthropologists, sociologists, and ethnomusicologists—disciplines more concerned with social processes. As we will demonstrate, the relationship between synchronization and coordination is not always dichotomous: indeed, we argue that a thorough understanding of IME

depends on consideration of both and of their interrelationship. One of the contributions we aim to make is to avoid the reduction of IME, a complex and multi-layered phenomenon, to SMS: if we assume that the demonstrated prosocial and bonding effects of entrainment are to be explained purely in terms of synchronization dynamics, for example, we may entirely miss a vital component of the process.

If these two components of our model have different disciplinary foundations, this carries through into different approaches to data collection and analysis. SMS has long been studied using time series data, whether from tapping experiments or extracted from audio performance recordings. Different analysis methods may be preferred depending on whether a linear or nonlinear mechanism is assumed—asynchrony measures vs phase analysis—but for most purposes these two approaches would produce analogous results. It is less clear from the literature how to analyze the longer-scale processes of coordination through which musical structures are aligned. Since coordination seems often to be manifested in coordinated movement, however, previous studies provide some precedents as to how this may be achieved using time-series data derived from video or motion capture data (Clayton, Jakubowski, & Eerola 2019; D'Ausilio et al., 2012; Eerola et al., 2018; Glowinski, Badino, Ausilio, Camurri, & Fadiga, 2012; Glowinski et al., 2013; Jakubowski et al., 2017). In the comparative data analysis we present in the section *Measuring Entrainment in Musical Ensembles* of this paper, synchronization is studied using an asynchrony-based approach pioneered by Rasch in the 1980s: the novelty of our analysis is in applying this method to diverse corpora, increasing the reproducibility of onset detection and comparing six very different musical traditions. Our coordination analysis deploys cross-wavelet transform (CWT) analysis on performance videos to explore the hypothesis that performer movements become more coordinated around musical points of change.

We aim, then, to explore empirically these two components of IME—synchronization and coordination—in diverse musical genres. We explore the distinction and its phenomenal, mechanistic, and sociological origins. We compare and contrast these two ideas, which are presented as dichotomous for historical and rhetorical reasons, with the ultimate aim of transcending the simple dichotomy and creating a nuanced model that bridges the dimensions of timescale, methodological approach, and disciplinary focus. We explicitly compare the music of different cultures in order to begin to explore what role “culture” may play in IME. We carry

this study out with explicit reference also to distinct disciplinary histories, bringing those intellectual trajectories into contact with the aim of fostering interdisciplinary cooperation (Jacoby et al., 2020).

In the next section, we review literature relating to core aspects of IME, especially as they relate to musical performance. Following an introduction to the topic in the section *Background and Operational Definitions*, the section *Foundations of Musical Entrainment* addresses IME's foundations from evolution and development through neuroscience and physiology before summarizing existing models of synchronization; the section *Social and Cultural Dimensions of Musical Entrainment* addresses social and cultural aspects of IME and discusses the role they should play in an expanded model of IME. We argue: a) that entrainment in musical contexts is an important component of interpersonal coordination in general; and b) that the diversity of such contexts means that in order to advance understanding, the phenomenon needs to be studied cross-culturally and its relationship to other social and cultural processes explored. We then outline measures of interpersonal entrainment in musical production (derived from both auditory and visual data, section *Measuring IME*). Using asynchrony calculations of a large and diverse collection, we present the first published cross-cultural summary and comparison of the synchronization between co-performers and explore the factors that account for variation in this data (section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis*). We also explore movement coordination between performers, using video data, in another novel cross-cultural analysis (section *Interpersonal Coordination in Music Ensembles: Continuous Data From Ancillary Movements*). Through these exemplar analyses we help to clarify which aspects of IME may be culturally variable, and which may vary little or not at all with culture. Finally, in the section *Models and Predictions* we conceptually model the relationship of these features to neurophysiological, social, and cultural processes, significantly expanding existing models of IME. We discuss influential existing models of IME, before setting out our own model, one that explicitly separates out synchronization and coordination, gives greater prominence and more detailed explanations of processes taking place over longer timescales than SMS, and considers which aspects of entrainment depend on culturally shared knowledge and how entrainment is socially significant. The paper thus contributes to the understanding of human coordination and cooperation, and to the study of cross-cultural variation in social behavior and artistic expression.

Since this work builds on past research across distinct disciplinary traditions, the first task of this paper is to set out this background through a critical review of the relevant literatures, which comprises approximately a third of the paper. Readers familiar with some or all of these areas may of course choose to navigate other routes through the material and refer back to the review section selectively. The third section of this paper includes a short overview of methods for studying IME before presenting comparative analyses of synchronization and coordination, before our model is laid out in the final section.

## Review: Interpersonal Musical Entrainment

### Background and Operational Definitions

#### INTERPERSONAL ENTRAINMENT

Studies of interpersonal entrainment were pioneered (albeit without using this term) by Condon in the 1960s, using manual annotation of films of individuals interacting socially. Condon claimed on this basis to have observed pervasive interpersonal synchrony between humans engaged in natural conversational interactions (Condon & Ogston, 1971), and also pointed to deficits in synchronization associated with certain pathologies such as autism (Condon, 1985). Advances in technology made less labor intensive methods viable, which have been exploited increasingly since the 1990s. For example, a number of studies by Schmidt, Richardson, and colleagues developed a set of ingenious methods to explore interpersonal synchrony in conversational interactions, such as asking participants to swing weighted pendulums or to sit in rocking chairs (e.g., Richardson, Marsh, Isenhowe, Goodman, & Schmidt, 2008; Schmidt et al., 2007). These interventions allowed researchers to carry out experimental research by varying factors including the weight of the pendulums and the amount of visual information available to participants, an approach that has included study of impairments in social interaction (e.g., Fitzpatrick et al., 2016).

While research on the dynamics of interpersonal entrainment (how tightly and under what conditions people synchronize with each other) has expanded over recent years, others have focused more on the possible motivations and effects of interpersonal entrainment. The historian William McNeill (1995) published a seminal account of the role of coordinated movement—dance and drill—in human history. Evolutionary anthropologists have discussed the possible importance of synchronized action and chorusing in

the emergence of music as well as its wider impact on human development (Merker, 2000; Mithen, 2005; Tomlinson, 2015). Dunbar and colleagues argue, for instance, that synchronous behaviors cause the release of neurohormones that influence social bonding—allowing humans to form affiliations more efficiently, and over larger social groups, than other primates (Launay et al., 2016).

The Durkheimian tradition in sociology and anthropology makes a parallel argument for the importance of ritualized behavior, including synchronized action, in the development of social bonds, institutions, belief systems, and moral codes (Collins, 2004; Durkheim, 1912/1995). Empirical work drawing on this tradition has explored the social efficacy of rituals with specific reference to the occurrence of interpersonal synchrony: Fischer, Callander, Reddish, and Bulbulia (2013) found that in a comparison between diverse activities, “rituals with synchronous body movements were more likely to enhance prosocial attitudes,” with sacred values mediating the effect of synchronization (p. 115).

Interpersonal coordination has been investigated in the field of experimental psychology by researchers concerned with joint action; that is, human behavior that involves multiple individuals coordinating their thoughts and movements in space and time, with the goal to communicate (Clark, 1996) or to effect a change in the environment (Sebanz, Bekkering, & Knoblich, 2006). Pioneering laboratory studies of joint action examined how co-acting individuals mentally represent each other’s role when sharing a task (Sebanz, Knoblich, & Prinz, 2005). Extensions of this work have focused on how using one’s own sensorimotor system to simulate a co-actor’s actions facilitates interpersonal coordination by allowing one individual to predict another’s upcoming actions in real time (Sebanz & Knoblich, 2009). Related theoretical work has sought to understand the relationship between shared task representations, action simulation, and basic mechanisms of entrainment in order to account for a range of intentional and unintentional forms of interpersonal coordination (Knoblich & Sebanz, 2008; Knoblich, Butterfill, & Sebanz, 2011).

In summary, the related ideas that the widely distributed capacity of human beings to coordinate actions in time have played an important part in: 1) human evolution; 2) the development of social institutions and the formation of social groups; as well as 3) the shared emotional lives of human societies, have been aired in various forms and in various academic disciplines over many decades—even in the absence of sustained engagement between those disciplines. Relating

empirical study of the dynamics of IME, such as it exists, to anthropological or sociological perspectives on interpersonal coordination and social bonding, remains a significant challenge, while virtually nothing is known about the cultural variability of IME. This paper aims to advance our theoretical models in these areas.

#### INTERPERSONAL ENTRAINMENT IN MUSIC

Music is undoubtedly a fruitful area in which to investigate interpersonal entrainment. Music performance involves the temporal coordination of movements between individuals, in ways that are often more complex or more precise than interpersonal coordination in other social situations. The fact that different individuals can each produce different traces in sound and movement means that in many cases, IME can be studied on the basis of normal music-producing behavior without the need for experimental interventions (D’Ausilio, Novembre, Fadiga, & Keller, 2015). However, interest in this topic amongst musicologists has been slow to build. In ethnomusicology a number of seminal studies have pointed to the importance of the phenomenon, without making significant breakthroughs. Lomax’s (1968) Cantometrics project, for instance, takes the “fusion” or “cohesiveness” between different parts in an ensemble as a key cross-cultural comparator: he wrote explicitly of Condon’s approach and its potential applicability to song performance. Blacking (1977, 1981), building on the Durkheimian tradition and anticipating later work on embodiment, wrote of the importance of synchronized movement in musical performance in his work on the anthropology of the body, again without developing a methodology for its study.

Empirical studies of musical synchronization were pioneered by Rasch (1979, 1988), working from recordings of Western classical trios, who devised measures for the estimation of synchronization between players (see section *Measuring Entrainment in Musical Ensembles*). Friberg and Sundström (2002) used jazz recordings to analyze the relationship between soloist and ride cymbal, while Schögler’s analysis (1999) of jazz duets explored the relationship between timing coordination and the overall flow of performances. Keil developed a theory of “participatory discrepancies.” For Keil, the small differences in timing (asynchronies) between individuals were crucial in giving life to a performance (1987/1994, 1995; related empirical studies are Alén, 1995; Butterfield, 2010; Prögler, 1995). The relationship between asynchrony and aesthetic preference was explored by Senn et al. (2016), whose findings did not support Keil’s theory, since in a set of rhythm section



TABLE 1. *Operational Definitions of Relevant Terms in Discussing Interpersonal Musical Entrainment, as Employed in This Paper*

Term	Operational definition
<b>Interonset Interval (IOI)</b>	Duration between the attack points of two successive auditory events
<b>Rhythm</b>	A sequence of auditory event durations or inter-onset intervals
<b>Beat</b>	A regular/repeated pulse that is abstracted from (and not necessarily perceptually present in) the rhythmical surface. Multiple beat levels combine to form a meter. Typically in the IOI range 250-2000 ms.
<b>Subdivision</b>	A fast regular/repeated pulse that is abstracted from the rhythmical surface, which subdivides the slower “beat.” Typically in the IOI range 100-250 ms.
<b>Meter</b>	A hierarchical structure, comprising two or more levels, into which beats and beat subdivisions are organized
<b>(Metrical) cycle</b>	A periodically repeating pattern comprising a hierarchical arrangement of beats on more than one level. Actual musical events or rhythmic patterns do not need to repeat periodically for a periodic metrical cycle to be inferred, although they may do so.
<b>Metrical position</b>	Location within the metrical hierarchy, i.e. beat or subdivision number.
<b>Tactus</b>	The beat level which is most comfortable to tap along to. Typically in the IOI range 350-700 ms.
<b>Tempo</b>	The perceived speed of the music, usually calculated as the frequency of the tactus (beats per minute).
<b>Event density</b>	The number of musical events (e.g., note onsets) occurring per unit of time.
<b>Entrainment</b>	The interaction of autonomous rhythmic (oscillatory) processes, often resulting in their synchronization.
<b>Sensorimotor synchronization (SMS)</b>	“[T]he rhythmic coordination of perception and action” (Repp, 2005). In a musical context, the process by which musicians use sensory input in order to synchronize with co-performers.
<b>Coordination</b>	Coordination can mean any process enabling medium and long-term musical processes (roughly > 2 s) to be or remain temporally aligned. This can include the cueing of transitions and the use of mutual attention and coordinated body movement to manage changes or reaffirm a shared understanding of the musical structure.
<b>Non-isochrony</b>	A regular pattern of unequal time intervals (usually at beat or subdivision levels).

recordings, those using the original timings from professional performances and those with the asynchronies eliminated were rated equally highly for “groove.” Clayton et al. (2005) integrated the ethnomusicological tradition of Blacking, Keil and others with application of empirical analysis informed by entrainment theory. Subsequent empirical studies of interpersonal entrainment in musical performance have explored the use of visual information to track coordination between musicians (Clayton, 2007b) and integrated the analysis of entrainment between musicians with ethnographic study of group dynamics in jazz ensembles (Doffman, 2013) and Cuban dance groups (Poole, 2012). Lucas, Clayton, and Leante (2011) explored the dynamics of entrainment between distinct musical groups in an Afro-Brazilian ritual tradition.

In parallel with this work, a large body of research has explored the process of sensorimotor synchronization using laboratory tasks that require individuals to move in synchrony with a perceived stimulus, such as tapping along to a periodic, auditory sequence or the beat of a piece of music (for an overview see Repp, 2005, and Repp & Su, 2013). Another common approach has been to explore coordination and joint action in musical ensembles within experimental paradigms (Glowinski

et al., 2013; Keller, 2008, 2014; Varni, Volpe, & Camurri, 2010). Until now, however, we lack a broader overview and model of IME: what it is, the mechanisms and timescale involved, what differences exist between IME in different cultures and musical genres, and how they may be accounted for.

For the purposes of this paper, IME refers to the temporal coordination between co-performers (perceptual factors in non-performers are considered in this literature review but do not form part of our model at this stage; bodily movement of musicians is considered, although full consideration of dance is deferred). IME is manifested on a number of different levels, from the phenomenon of different individuals sounding instruments (or clapping along or shifting their weight in dance) at the same time, to the alignment of metrical and phrase structures, to coordinated transitions between pieces or sections. We propose in this paper a distinction between alignment at short timescales based on a process of sensorimotor *synchronization*, and *coordination* at larger timescales, which is more dependent on conscious decision-making and negotiation. For reference, in Table 1 we define some of the key terms in musical entrainment, as a key to the discussion that follows.

## FOUNDATIONS OF MUSICAL ENTRAINMENT

*Evolution and Development.* Although the cultural value placed on music making, and the variety of temporal patterns used in music may be uniquely human, examples of entrainment to rhythmic, periodic stimuli have been reported in several non-human species.<sup>2</sup> Several of these studies have also demonstrated that data collected from animal subjects can be explained using mathematical models that have previously been applied in studies of human entrainment, including the generalized Weber's law and a coupled oscillator model (García-Garibay, Cadena-Valencia, Merchant, & de Lafuente, 2016; Rouse, Cook, Large, & Reichmuth, 2016; see the review section *Models of Synchronization and Beat Perception* for more on entrainment models). Initial findings in this research area, in particular on cockatoos and other parrots, led to the proposal that the ability to entrain to a beat is related to vocal mimicry capacity (Patel, 2006; Schachner, 2010), although the phylogenetic diversity of animals that have since shown evidence of entrained behavior may now suggest otherwise. A theoretical account that has more recently gained traction is that beat-matching and synchronization abilities share a common origin in general entrainment mechanisms and signaling behaviors that are widely distributed across the animal kingdom (Bispham, 2006; Merker, Madison, & Eckerdel, 2009; Phillips-Silver, Aktipis, & Bryant, 2010; Wilson & Cook, 2016), but vary in their accuracy and flexibility due to cross-species differences in motor control, perceptual abilities, cognitive and emotional factors such as attention, and the strength of coupling between the auditory and motor systems (Kotz, Ravignani, & Fitch, 2018; Merchant & Honing, 2014; Wilson & Cook, 2016).

In regard specifically to human musical entrainment, various prerequisite abilities are present from an early stage of development. Winkler, Háden, Ladinig, Sziller, and Honing (2009) found that newborn infants were sensitive to omissions of downbeats from isochronous rhythmic patterns and thus suggested that

musical beat perception is innate. Phillips-Silver and Trainor (2005) found that bouncing 7-month-old infants to a binary or ternary beat when listening to an ambiguous auditory rhythm influences subsequent preferences towards a version of the rhythm containing accentuations that match the metrical interpretation to which the infant has been bounced. This suggests that the ability to retain metrical information and, specifically, to integrate motor and/or vestibular information with auditory rhythms is present from an early developmental stage. Zentner and Eerola (2010) found that 5- to 24-month-old infants demonstrated more spontaneous rhythmic movement to music and musical rhythms than speech, with evidence of tempo sensitivity (movements became faster with increasing tempo) although the absolute periods of the music and movements did not reliably match. It appears the ability to overtly synchronize motor responses to a musical beat requires a more refined degree of motor control that emerges around age 2 (Eerola, Luck, & Toivainen, 2006; Kirschner & Tomasello 2009). In general, synchronization accuracy and the range of musical tempi to which one can accurately synchronize have both been shown to increase throughout childhood (Drake, Jones, & Baruch, 2000; McAuley, Jones, Holub, Johnston, & Miller, 2006).

## SENSORIMOTOR SYNCHRONIZATION (SMS)

Basic psychological processes that underpin musical entrainment have been interrogated in laboratory studies of SMS. The processes of interest allow an individual to perceive the rhythm of an external sequence of events, to anticipate the timing of upcoming events based on this rhythm, to produce rhythmic movement, and to align the timing of these produced movements with events in the external sequence. SMS studies have generally sought to understand how these processes combine to determine the accuracy and precision of SMS in tasks that require an individual to synchronize simple movements (finger taps, drum strikes, or limb oscillations) with repetitive events in auditory, visual, or multimodal (e.g., audio-visual) pacing sequences (Repp, 2005; Repp & Su, 2013). Performance on such tasks can be quantified by measuring the phase relationship (or temporal alignment) between each movement and its corresponding pacing event, as well as the degree to which the period (or interval) of time between successive movements matches the intervals demarcated by pacing events. SMS *accuracy* is high to the extent that errors in phase (asynchronies) and period (tempo) are low, while *precision* is high to the extent that asynchronies and tempo remain constant over time.

<sup>2</sup> These include evidence of quite reliable and flexible entrainment in a sea lion (Cook, Rouse, Wilson, & Reichmuth, 2013; Rouse et al., 2016), entrainment to complex musical stimuli in sulphur-crested cockatoos (Patel, Iversen, Bregman, & Schulz, 2009) and African grey parrots (Schachner, Brady, Pepperberg, & Hauser, 2009), synchronization to simple isochronous stimuli in budgerigars (Hasegawa, Okanoya, Hasegawa, & Seki, 2011), a bonobo (Large & Gray, 2015), a chimpanzee (Hattori, Tomonaga, & Matsuzawa, 2013), and rhesus monkeys (Zarco, Merchant, Prado, & Mendez, 2009), spontaneous synchronization of button-pressing movements in pairs of macaques (Nagasaka, Chao, Hasegawa, Notoya, & Fujii, 2013), and preliminary evidence for entrainment in horses (Bregman, Iversen, Lichman, Reinhart, & Patel, 2013).

Laboratory studies have identified several factors that affect SMS performance, including the modality of the external pacing sequence, tempo (musical speed), and an individual's musical expertise. SMS is typically more accurate and precise with auditory than visual stimuli for sequences of discrete events, but this does not necessarily hold for continuous stimuli (Hove, Fairhurst, Kotz, & Keller, 2013; Repp, 2003). With regard to tempo, accurate and precise SMS with auditory pacing sequences is possible at IOIs in the 100–1800 ms range (600–33 BPM), with synchronization at fast extremes requiring biomechanical constraints on repetitive movement to be overcome by using alternating fingers, arms, or feet (Repp, 2006). Individual pacing events are difficult to distinguish at faster tempi and the timing of events is difficult to predict at slower tempi falling outside this range. Trained musicians may, however, benefit from relatively fine temporal acuity at the fast end and from compensatory task strategies such as mental subdivision at extremely slow tempi. Music training is generally associated with superior SMS performance. Highly trained musicians are able to limit the variability of asynchronies to about 2% of the pacing IOI, while variability in untrained individuals is typically at least twice as large as this (Repp, 2005).

The sources of individual differences in SMS performance have been investigated in studies exploring models that correct for phase and period errors, as well as in brain imaging studies examining the neural correlates of SMS and work on SMS in naturalistic musical contexts. Modeling studies have sought to determine how differences in estimates of parameters representing basic SMS mechanisms account for individuals' behavioral SMS performance (Mills et al., 2019; Palmer, Lidji, & Peretz, 2014; van der Steen, Jacoby, Fairhurst, & Keller, 2015). Results of this work suggest that variance in SMS skill is explained by combined independent contributions of reactive error correction and predictive tempo-change extrapolation processes. Brain imaging studies suggest that individual differences in SMS skill may reflect the efficiency of neural responses at early levels of auditory processing, such as brainstem responses (Tierney & Kraus, 2013), and the degree of connectivity between higher-level auditory and motor cortical regions (Chen, Penhune, & Zatorre, 2006; Halwani et al., 2011). Work on SMS in musical contexts has addressed individuals' abilities to keep in time with tempo changes in expressive musical performances or simple tone sequences containing gradual tempo changes resembling those found in such performances (Pecenka & Keller, 2009a, 2011; Rankin, Large, & Fink, 2009; Repp, 2002; Schulze, Cordes, & Vorberg, 2005). The ability to predict on-

going tempo changes is reflected in the degree to which intertap intervals match or lag behind prior interonset intervals. These studies indicate that the tendency to predict tempo changes (reflected in matching more than lagging) is positively correlated with musical experience, and these individual differences are related to the accuracy and precision of synchronization with regular pacing sequences and sounds produced by another individual in tapping tasks (Pecenka & Keller, 2011). Temporal anticipation requires relatively high-level cognitive processes. Tempo change prediction is reduced with increased attention load (Pecenka, Engel, & Keller, 2013) and individual differences in prediction are positively correlated with working memory capacity (Colley, Keller, & Halpern, 2018) and auditory imagery ability (Pecenka & Keller, 2009a, 2009b). While most SMS research focuses on isochronous sequences, Repp, London, and Keller (2008) explored non-isochronous sequences, finding larger phase corrections following longer intervals (600 ms) than after shorter (400 ms).

#### NEUROSCIENCE AND PHYSIOLOGY OF MUSICAL ENTRAINMENT

Research in this domain has explored the theory that a sense of pulse and meter arise as a result of neural oscillations resonating to the periodicities of a regular auditory stimulus. A range of methods has been employed, with differing strengths and limitations, to address entrainment-related phenomena at the level of the brain (Doelling, Assaneo, Bevilacqua, Pesaran, & Poeppel, 2019; Lenc, Keller, Varlet, Nozaradan, 2018a, 2018b, 2019; Novembre & Iannetti, 2018; Rajendran & Schnupp, 2019). Studies using electroencephalography (EEG) and magnetoencephalography (MEG) have shown that exposure to a periodic tone sequence or music with an isochronous beat can elicit neural activity consistent with entrainment at matching and harmonically related frequencies to the stimulus pulse (Fujioka, Trainor, Large, & Ross, 2009, 2012; Nozaradan, Peretz, Missal, & Mouraux, 2011; Snyder & Large, 2004, 2005; Tierney & Kraus, 2015). In addition, neural oscillations and modulations of power in specific frequency bands appear to play different roles in terms of entrainment to a musical stimulus; in particular, amplitude modulations in the power of beta oscillations (15–30 Hz) appear to be more exogenously driven by the perceived stimulus, while periodic gamma activity (> 30 Hz) persist even when a tone is omitted, suggesting a role in endogenous anticipatory processes (Fujioka et al., 2009; Large & Snyder, 2009; Zanto, Large, Fuch, & Kelso, 2005). Another common methodological approach, known as “frequency-tagging,” measures the correspondence between the frequency of a periodic auditory

stimulus and the frequency of neural oscillations by analyzing steady-state evoked potentials (SSEPs) from the neural signal (see Nozaradan, 2014, for an overview). The frequency tagging approach has been particularly effective for exploring the range of conditions under which neuronal activity suggestive of entrainment occurs, due in part to the high signal-to-noise ratio afforded by this method, and has also revealed insights on the neural dynamics of audiovisual (Nozaradan, Peretz, & Mouraux, 2012a) and sensorimotor (Nozaradan, Zerouali, Peretz, & Mouraux, 2015) integration. Approaches based on measuring transient event-related potentials (ERPs) from the neural signal have also been successfully employed in several studies, in particular providing insights on predictive processes and temporal expectancy (Stupacher, Witte, Hove, & Wood, 2016; Zanto, Snyder, & Large, 2006). The modulation of neural responses via a musical beat has been demonstrated not only at the level of cortical activity, but also at the subcortical level of the auditory brainstem (Nozaradan, Schönwiesner, Caron-Desrochers, & Lehmann, 2016; Tierney & Kraus, 2013, 2014).

In addition to the correspondence between the perceptual pulse rate and neural oscillation rate, it has been found that imagined metrical interpretations that are not physically present in the acoustic signal can elicit neuronal responses at the level of the imagined meter. This has been demonstrated by comparing EEG data from trials in which participants imagined binary vs. ternary metrical interpretations for the same auditory stimulus, as well as using paradigms in which different patterns of subjective accentuation were imagined using the same, metrically ambiguous rhythmic pattern (Brochard, Abecasis, Potter, Ragot, & Drake, 2003; Iversen, Repp, & Patel, 2009; Nozaradan et al., 2011). Beat-related neuronal responses also still occur when the acoustic energy of the signal is not predominant at beat onsets, as is the case for syncopated rhythms (Nozaradan, Peretz, & Mouraux, 2012b; Tal et al., 2017). Thus, neuronal entrainment to music appears to be influenced not only by bottom-up properties of the auditory stimulus, but also by top-down interpretations or abstractions of the beat that may provide additional or even conflicting information to the perceived stimulus.

In terms of interpersonal entrainment, research using dual-EEG recording has shown that increased coordination between pairs of participants, as measured using hand/finger movement tasks (including tapping) requiring varying levels of spontaneous and planned coordination, is associated with suppression of activity in the alpha frequency band (neural oscillations of approximately 7–13 Hz) (Dumas, Martinerie, Soussignan, &

Nadel, 2012; Konvalinka et al., 2014; Naeem, Prasad, Watson, & Kelso, 2012; Tognoli, Lagarde, DeGuzman, & Kelso, 2007). The topography of patterns of activation observed in these studies is broadly consistent with the involvement of sources in the putative human mirror neuron system (see Rizzolatti & Sinigaglia, 2010). Novembre, Sammler, and Keller (2016) investigated the interaction between synchronization and higher-level knowledge structures (co-representation of actions and goals) in a musical joint action paradigm with piano duos by manipulating each of these two variables. In accordance with previous research, they found suppression of alpha-band activity over right centro-parietal regions in the more synchronous performance condition and alpha enhancement in the less synchronous condition, with this difference in alpha activation levels being amplified in the condition that favored co-representation (pianists were familiar with one another's parts) as compared to the condition that did not allow for co-representation (pianists were unfamiliar with one another's parts).

Some effects of expertise and individual differences between participants on neural entrainment have also been revealed. For instance, Doelling and Poeppel (2015) found that musicians displayed more precise neuronal entrainment to music, and entrainment over a wider frequency range than nonmusicians. Tierney and Kraus (2014) showed that participants who were less variable in tapping during a sensorimotor synchronization task also displayed decreased variability of the auditory brainstem response in terms of phase locking to a periodic auditory stimulus. Other work by Tierney and Kraus (2013) has revealed correlations between both SMS ability and years of music training with cortical sensitivity to a musical beat. Research using a frequency-tagging approach has shown that the amplitude of SSEPs was positively correlated with SMS ability, while the strength of endogenous neural activity related to a non-perceptually present beat was positively correlated with temporal prediction abilities (Nozaradan, Peretz, & Keller, 2016). Finally, evidence of neuronal entrainment has also been found in infants as young as 7 months of age, although such responses appear to be modulated by both the musical experience of the infant and its parents (Cirelli, Spinelli, Nozaradan, & Trainor, 2016).

Going beyond questions of entrainment at the level of populations of neurons, the relationship between other periodic, physiological signals (e.g., heart rate, respiration rate) and musical pulse rate has been a topic of empirical interest for decades (e.g., Diserens, 1926), although there are still many open questions in this



area. Some studies have reported the spontaneous adaptation of blood pressure, heart rate, and respiration rate toward the tempo of musical stimuli (Bernardi et al., 2009; Etzel, Johnsen, Dickerson, Tranel, & Adolphs, 2006; Gomez & Danuser, 2007; Haas, Distenfeld, & Axen, 1986), although the exact mathematical relationship between physiological rhythm rates and musical tempo is not straightforward, most likely due to biological constraints on the possible range of periodicities in autonomic functions. Ongoing research in this area has key implications for understanding emotional responses to music (Juslin, 2013), the development of music therapeutic interventions that aim to regulate physiology (e.g., Ellis & Thayer, 2010; Thaut, McIntosh, & Hoemberg, 2015), and the investigation of the effects of music on exercise and sporting performance (Brooks & Brooks, 2010; Karageorghis & Terry, 2008).

#### AUDITORY PERCEPTUAL FACTORS IN IME

Perceptual factors may influence IME in performance in a number of ways. First, the perceptual onset—or p-center—is the subjective moment of occurrence of an event, in the case of music, a percussive sound, pitched tone, or chord onset. P-centers of musical sounds are determined by acoustic rise time, duration, pitch, and timbral qualities (Danielsen et al., 2019; London et al., 2019; Rasch, 1979; Vos & Rasch, 1981). Instrumental sounds therefore have different p-centers to the extent that the method of sound production is associated with differences in these physical features. P-centers are closer to physical onsets in percussive sounds or plucked string tones with abrupt rise times than in bowed string or breathy wind tones with gradual rise times. The effects of differences in p-center on the dynamics of entrainment have been examined in SMS studies of paced finger tapping. For example, longer rise times of individual tones and the temporal displacement of multiple tone onsets in chords have been found to draw taps later in time relative to physical onset of the tone or chord (Hove, Keller, & Krumhansl, 2007; Vos, Mates, & van Kruysbergen, 1995). The task for ensemble musicians is to align sounds based on p-centers rather than physical onsets (Rasch, 1988).

Measurements in synchronization precision and accuracy may be considered in relation to thresholds for judging the temporal order of two sounds. These are about 20 ms for steady-state tones (Hirsh, 1959) and 30 ms for musical (piano) tones (Goebel & Parncutt, 2001; see also Butterfield, 2010). The latter value is close to the measured SD of asynchronies that have been observed in Western classical ensemble performance, where onset time differences often vary with the range

of around 0–30 ms and rarely exceed 50 ms (Keller & Appel, 2010; Keller, Knoblich, & Repp, 2007; Rasch, 1988; Shaffer, 1984). This correspondence suggests that musicians' abilities to synchronize with one another thus match listeners' abilities to perceive the temporal relations between different instrumental parts in music.

Studies on regularity in production of temporal sequences and perceptual thresholds for detecting variability in performance tempo provide indications of the size that ensemble asynchronies would need to be in order to be detected by a listener. The limits of temporal regularity in production are often investigated in simple laboratory tasks requiring self-paced finger tapping. When the goal is to produce a regular sequence at tempi within the comfortable range of 100–120 BPM (500–600 ms IOIs), the standard deviation of intertap intervals is typically around 15–30 ms, that is,  $\approx 3\text{--}6\%$  of the interbeat interval (Madison & Merker, 2002, 2004). The proportional degree of variability in production remains constant across a range of rates relevant to the experience of beat and metric beat subdivisions (from  $\sim 150$  ms to  $\sim 1500$  ms), leading to an absolute increase in variability with decreasing rate (Vorberg & Wing, 1996; Wing & Kristofferson, 1973a, 1973b). Variability in production corresponds with perceptual thresholds for detecting temporal irregularity, which are around 3% or lower for musicians and 4% or higher for nonmusicians (Friberg & Sundberg, 1995; Madison & Merker, 2002, 2004). Listeners therefore do not consciously detect irregularities that are smaller than those that characterize production, suggesting a close match between limits in perception and action. The convergence of perceptual and production limits on values around 3% of the IOI or greater suggests that smaller asynchronies are uninformative regarding ensemble performers' intentions and may go unnoticed.

In addition, perceptual studies of expressive performance have shown that the detection of timing deviations depends on their position within the musical structure. Detection is poorer at the end of melodic-rhythmic groups and phrases than at the beginning or middle of such structural units (Repp, 1999). This link between musical structure and temporal acuity suggests that the perception of asynchrony may also vary as a function of position in musical structure. Specifically, acuity in asynchrony detection may decrease at points approaching significant melodic-rhythmic and phrase boundaries, where in Western art music there is typically a slowing down of local tempo. In accordance with Weber's psychophysical law (which states that perceptual sensitivity to a change in some property of a stimulus is inversely proportional to the physical dimensions

of that property of the initial stimulus), asynchronies may need to occupy a greater absolute amount of the interbeat interval at slow than at fast rates in order to be noticeable.

#### VISUAL INFORMATION AND CROSS-MODAL EFFECTS IN IME

Visual information about co-performers' body movements can also facilitate interpersonal coordination. In general, auditory information is considered to be superior for communicating fine temporal structure (e.g., expressive microtiming) and discrete onset timing, while visual information is optimally suited for communicating larger-scale spatial-temporal structures and the continuous dynamics of events as they unfold (MacRitchie, Varlet, & Keller, 2017). Although the visual system has lower temporal resolution than audition (Holcombe, 2009), it excels at processing spatiotemporal information, as evidenced by the high precision with which moving stimuli can be intercepted (Bootsma & van Wieringen, 1990). This is especially the case for biological motion (Aymoz & Viviani, 2004). In music, continuous spatiotemporal trajectories allow the future course of performers' movements to be predicted (Wöllner & Canal-Bruland, 2010), which assists ensemble co-performers to coordinate their sounds (Glowinski et al., 2013; Kawase, 2014; King & Ginsborg, 2011) and may also be used as a cue by audience members in evaluating performance quality.

Musicians make use of two broad categories of movement when performing. Sound-producing movements (e.g., key presses, bow strokes) occur over the same timescales as auditory onsets. Sound-facilitating, or *ancillary*, movements (e.g., head nods, body sway) are movements that do not play a direct role in the production of sound and typically occur over longer timescales than sound-producing movements. Ancillary movements may thus play a role in coordinating both temporal and expressive intentions of performers over slower timescales (e.g., sections or phrases rather than note-to-note onsets; Dahl & Friberg, 2007; Davidson, 1993; Ginsborg & King, 2009; Teixeira, Loureiro, Wanderley, & Yehia, 2015; Vines, Krumhansl, Wanderley, & Levitin, 2006; Wanderley, Vines, Middleton, McKay, & Hatch, 2005; Williamon & Davidson, 2002). In addition, Bishop and Goebel (2017) have found that kinematic features of communicative head gestures in piano duos are predictive of note-level synchronization, suggesting a link between ancillary movement and accuracy of sound-producing movements.

In terms of listeners' usage of visual cues for assessing aspects of performance, Moran, Hadley, Bader, and Keller (2015) have demonstrated different sensitivities in

relation to musical style. Specifically, listeners were able to discriminate between real and fake pairings of improvising duos on the basis of visual displays of performers' body movements in free improvisation, but did not perform above chance for standard jazz performances. The apparent advantage for free improvisations may have been due to these performances being more "conversation-like" than standard jazz in terms of interpersonal coordination dynamics. In addition, there was a positive correlation between auditory rhythm perception skills and the ability to discriminate real from fake visual cues in the standard jazz condition (but not for free improvisation), indicating some relationship between auditory temporal acuity and visual temporal acuity.

In an investigation of the cross-modal aspects of ensemble synchronization, Arrighi, Alais, and Burr (2006) asked participants to judge the synchrony of video and audio streams of displays of conga drumming. The results indicated that the auditory stream needed to be delayed in order for onsets conveyed by sight and sound to be perceived as synchronous, presumably due to the relatively sluggish processing of visual information. The audio-visual temporal integration window—i.e., the range where synchrony perception tolerates auditory delays—decreased with increasing tempo and was around 200 ms for 1 Hz (60 BPM) drumming movements and 100 ms for 4 Hz (240 BPM) movements. This may be an instance of Weber's law holding cross-modally.

Aspects of audiovisual perception also vary in relation to musical experience and task type. Petrini and colleagues (Petrini, Russell, & Pollick, 2009) found an expertise effect when comparing drummers to musical novices in the detection of asynchronies between visual and auditory components of audiovisual displays of solo drum strikes, in particular in regard to the temporal integration window (the range of values where asynchrony goes unnoticed). Building on this work, Love, Petrini, Cheng, and Pollick (2013) investigated the relationship between synchrony judgments and temporal order judgments in audiovisual stimuli. No correlation was found across the two tasks, and the authors thus concluded that synchrony and temporal order judgments are underpinned by different perceptual mechanisms.

#### IME AND LISTENER PREFERENCES

It is not clear to what extent differences in ensemble synchrony across different musical traditions may be aesthetically motivated, but evidence does exist of listener preferences. Such preferences might arise because the degree of synchrony and coordination exhibited

within the ensemble influences the strength of coupling experienced by an observer or listener, thereby shaping his or her affective responses to the music (Labbé & Grandjean, 2014; Trost, Labbé, & Grandjean, 2017). Potential mediating factors in this process may include the observer's musical experience (Novembre & Keller, 2014) and empathy (i.e., ability to understand others' thoughts and feelings; Babiloni et al., 2012).

Preferences for particular patterns of interpersonal coordination are not specific to interaction in musical contexts. For instance, Miles, Nind, and Macrae (2009) found that observers' ratings of rapport between pairs of individuals walking together were highest for in-phase (relative phase = 0 degrees) and anti-phase (180 degrees) coordination. When considered in light of a large body of research showing that in-phase and anti-phase relations are the most stable amongst coordination modes (Schmidt & Richardson, 2008), these findings suggest that observers prefer coordination modes that are relatively comfortable to produce.

In a study of musical entrainment, Labbé, Glowinski, and Grandjean (2017) examined the effects of ensemble and solo performance of Schubert's *String Quartet No. 14 in D minor* ("Death and the Maiden") on the affective experiences of listeners. Ratings of motor entrainment (e.g., "to what extent did you feel like moving?") and visceral entrainment (e.g., "to what extent did you feel your own bodily rhythms change?") were higher for ensemble than solo performances. The motor entrainment effect may be due to greater event density making one want to move (see also Senn, Kilchenmann, Bechtold, & Hoels, 2018), while the visceral effect may be due to the urge to move increasing arousal. In addition, wanting to move (motor entrainment) was predictive of positive emotions such as power and wonder, while having a sense of one's own bodily rhythms changing (visceral entrainment) was predictive of both positive and negative emotions.

The concept of motor entrainment is closely aligned with the notion of musical "groove"—a quality of music that induces the pleasurable urge to move in listeners. In a study of experiences related to groove, Hurley, Martens, and Janata (2014) investigated spontaneous sensorimotor coupling with multipart music (rock, jazz, funk, bluegrass, hip-hop, reggae, new age, and electronic dance music) comprising 1–4 instruments. Continuous ratings of groove (i.e., "the aspect of music that compels the body to move") while the music was playing were higher for music with multiple instruments than for solo music, and, in the multipart textures, groove ratings increased with each staggered instrument entry. Post-excerpt ratings also indicated a higher urge to

move, enjoyment, and wanting the music to continue when there were multiple instruments and their entries were staggered. Thus, motor entrainment related to groove was positively associated with listener engagement and preference.

Preferences can be influenced by the overall degree of entrainment (e.g., as reflected in the average magnitude of asynchronies), as well as by the dynamics of mutual influence within an ensemble. D'Ausilio et al. (2012) compared aesthetic preferences of listeners against objective measures of leader-follower relations between a conductor and violinists in orchestral performances. Results revealed differences in preferences for patterns of conductor and co-performer influence across musical excerpts. For one excerpt, perceived quality was high when the conductor strongly influenced members of the violin section but mutual influence between violinists was weak, while, for another excerpt, perceived quality was high when the influence of the conductor was weak but mutual influence between violinists was strong. These findings may reflect differences in conductor experience and leadership style. A strong leader effectively guides performers but, when faced with a weaker leader, co-performers are forced to rely on one another (see later discussion of leadership in the section *Social Differentiation: Role, Leadership, and Out-groups*).

A further question concerns the relationship between sensitivity to, and preferences for, different degrees of IME. In a relevant study, Engel and colleagues (Engel, Hoefle, Monteiro, Bramati, et al., 2014; Engel, Hoefle, Monteiro, Moll, & Keller, 2014) investigated the effects of synchrony between percussion instruments on experienced pleasure and the desire to move or dance while listening to Brazilian samba music. Results suggested that listeners were sensitive to each step in the synchrony manipulation that was introduced by the experimenters, although there were considerable individual differences in judgments that related to rhythm and time perception skills. Furthermore, stimuli that had higher synchrony between instruments were perceived as more pleasurable and evoked a greater desire to move or dance. The general correspondence between judgments of synchrony, pleasure, and the desire to move suggest that there is a close relationship between direct measures probing perceived entrainment and indirect measures probing the listener's experience of their own entrainment to music.

#### MODELS OF SYNCHRONIZATION AND BEAT PERCEPTION

One of the aims of this paper is to develop a model of IME. Of the various aspects of the phenomenon touched on so far, that which has been subject to the

most extensive modeling is sensorimotor synchronization. These models generally aim to explain the mechanisms underlying SMS in terms of either: 1) linear, event-based, and information processing accounts, or 2) nonlinear, emergent timing approaches based on oscillator models and dynamical systems theory. The following paragraphs describe some key assumptions and examples of each of these two classes of models.

Event-based, information processing models are often used to explain the timing mechanisms underlying series of discrete events, such as a sequence of finger taps or acoustical onsets produced by a musical instrument. Such approaches have evolved from interval timing models, which have a long historical precedence in time perception research. Scalar expectancy theory (SET; Gibbon, 1977; Gibbon & Church, 1990; see also Treisman, 1963) is a prominent model that has been used to explain a host of time-related animal behaviors, as well as many aspects of human performance on duration estimation and production tasks (Buhusi & Meck, 2005; Malapani & Fairhurst, 2002). SET implicates a pacemaker-accumulator mechanism (“internal clock”) that generates and stores a series of pulses over a certain time period for comparison to a reference memory in making time-related decisions. The role of attention has been further specified in subsequent adaptations of internal clock models (e.g., Zakay & Block, 1996), as attention allocation has been shown to have a substantial influence on the accuracy of timing tasks (Brown, 1997). In addition, an internal clock model that specifically aims to explain the processing of auditory sequences has been proposed by Povel (1981) and Povel and Essens (1985), which takes into account the hierarchical, beat-based organization of musical rhythms.

The work of Wing and Kristofferson (1973a) integrates concepts from SET within an event-based model that aims to explain discrete, repetitive motor responses, such as a series of finger taps. Specifically, the Wing and Kristofferson model also assumes an internal clock/timekeeper that is affected by attention, but posits two sources of variability in the produced motor responses—central timekeeper variance and peripheral motor variance. These two sources of variability are presumed to be independent of one another, such that separate estimates of timekeeper and motor variability can be attained for a produced motor sequence. Subsequent extensions of the Wing and Kristofferson model to SMS implicate a linear autoregressive phase error correction process, which essentially posits that the duration of the current motor response (e.g., tap) is directly influenced by the asynchrony of the previous response (or two) (Pressing, 1998; Vorberg & Schulze,

2002; Vorberg & Wing, 1996). Jacoby and colleagues (Jacoby, Keller, Repp, Ahissar, & Tishby, 2015; Jacoby, Tishby, Repp, Ahissar, & Keller, 2015) have demonstrated that adding an additional assumption (based on existing behavioral data) that the motor variance is less than the timekeeper variance can substantially reduce parameter estimation error and bias.

Several studies have also revealed the need for a distinction between phase error correction, which is a generally automatic process, and period error correction, which is under conscious control and requires attention (Mates, 1994a, 1994b; Repp & Keller, 2004; Schulze et al., 2005). When synchronizing with an isochronous sequence, only phase correction is needed, whereas synchronization to a stimulus that changes tempo (i.e., most scenarios involving music making between humans) also requires period correction of the internal timekeeper (Repp & Keller, 2004). Complicating this picture, the existence of an intermittent phase resetting process alongside continuous phase correction has also been suggested (Repp, 2001; Keller & Repp, 2005; Rimele, Morillon, Poeppel, & Arnal, 2018). Motivated by behavioral findings that humans predict on-going tempo changes (such as expressive timing variations), error correction processes that facilitate reactive temporal adaptation have been supplemented by anticipatory processes in the ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization proposed by van der Steen and Keller (2013; see also van der Steen et al., 2015). More broadly, one of the precursors of ADAM, described in Phillips-Silver & Keller (2012), draws extensively on theoretical approaches to joint action (Knoblich et al., 2011; Sebanz et al., 2006). In this view the mechanisms by which ensemble members mutually entrain include not only adaptive timing but also prioritized integrative attending (i.e., attending to one’s own actions, those of others, and the integrated sound at the same time) and anticipatory imagery, which helps to predict the future sounds of co-performers (Keller, 2008; Keller & Appel, 2010). Related research outside the music domain suggests that joint action more generally is supported by a mixture of high-level cognitive processes and basic sensorimotor mechanisms that can be strategically modulated through behavioral modifications (e.g., increased temporal regularity and spatial extent of movements) that smoothen interpersonal coordination by making actions clearly perceivable and predictable (Vesper et al., 2017; Vesper, Butterfill, Knoblich, & Sebanz, 2010).

A complementary body of work considers entrainment in terms of nonlinear oscillatory processes and conceptualizes entrainment as a continuous process,



rather than a sequence of discrete events. Such models are based on dynamical systems theory, as developed within mathematics and physics for modeling complex and nonlinear systems (for an overview see Guckenheimer & Holmes, 1983, or Hirsch, Smale, & Devaney, 2004), which has also been highly influential within such diverse fields as biology, economics, and psychology (e.g., May, 1976; Vallacher & Nowak, 1994). One prominent approach to modeling musical entrainment within this tradition is dynamic attending theory (DAT), as proposed and refined by Jones and colleagues (Jones, 1976, 2019; Jones & Boltz, 1989; Large & Jones, 1999). DAT implicates a set of internal (neural), self-sustaining oscillations or “attending rhythms” that can entrain to external events and direct attentional energy to expected points in time. The entrainment of these internal oscillations to a periodic auditory sequence facilitates the processing of sounds presented in phase with the sequence (Jones, Moynihan, MacKenzie, & Puente, 2002, though see also Bauer, Jaeger, Thorne, Bendixen, & Debener, 2015, for some conflicting evidence). Subsequent models have specified how temporal expectations become stronger and more focused with more iterations of an auditory stimulus and have proposed systems comprising multiple, nested oscillators that track different periodicities in line with the hierarchical structure of music (Large, 2008; Large & Jones, 1999; Large & Kolen, 1994; Large & Palmer, 2002; McAuley & Kidd, 1995). van Noorden and Moelants (1999) have also proposed a model that implicates an oscillator with a natural resonance of approximately 2 Hz, to account for human preferences for musical tempo rates around 2 Hz and the limited frequency range over which tempo perception/production can occur. Finally, neural resonance theory builds on initial ideas from DAT and posits spontaneous oscillations of neural populations at an endogenous periodicity in the absence of a perceived stimulus, which can become coupled (entrained) to an external stimulus over a broad range of phase relationships, while higher order resonances in the oscillators can give rise to the perception of metrical accents (Large, 2008; Large & Snyder, 2009). Further development of neural resonance theory has demonstrated that oscillatory interactions between auditory and motor brain networks can explain the perception of a musical pulse even in the absence of energy in the acoustic signal (e.g., as is often the case with syncopated rhythms, where acoustical onsets do not coincide with the metrical pulse; Large, Herrera, & Velasco, 2015; Velasco & Large, 2011).

Both linear and nonlinear models have proved successful in modeling aspects of SMS, in the case of the

former especially so when an anticipatory module capable of tracking tempo changes is added (as in ADAM). Iversen and Balasubramaniam (2016) argue that a full account of SMS may require both neural resonance and interval timing mechanisms: “the existence range of SMS closely corresponds to the temporal interval range over which both systems are active in interval timing” (2016, p.177). Their model of “rich BPS” (Beat Perception and Synchronization) offers the advantage of accounting for the rich and hierarchical temporal patterns that are operative in musical production and perception. An explicit role not only for neural resonance but also for memory and internal representation, and connections in both directions between sensory and motor processing, allows us to consider the ways in which synchronization may vary between musical genres and cultures (see Section 4). Along similar lines, a number of studies theorise the role of top-down processes, or “Active Sensing,” within a nonlinear model of temporal perception (Morillon & Schroeder, 2015; Patel & Iversen, 2014; Schroeder, Wilson, Radman, Scharfman, Lakatos, 2010). Rimmele et al. (2018), for instance, recast DAT and neural resonance theory to account for the influence of top-down factors on entrainment: in this context they suggest that “the brain can arguably exploit any available source of top-down priors” (p. 876). These priors may include both efferent motor signals and symbolic cues based on working or long-term memory.

Much remains to be achieved, nonetheless, in terms of accounting for synchronization to a wide variety of metrical structures, including those with stable non-isochronous patterns (Polak, London, & Jacoby, 2016), existing in specific cultural settings (Jacoby & McDermott, 2017; Patel, Iversen, & Ohgushi, 2006; Sadakata, Ohgushi, & Desain, 2004). We also suggest below that more work is needed to model long-term anticipatory processes that may involve detailed expectations regarding transitions, in terms of not only tempo but metrical pattern, texture, rhythmic patterning and other factors that rely on expert knowledge. The model presented in the *Models and Predictions* section of this paper is intended as a step in this direction. We will also expand our model beyond the narrow frame of synchronization mechanisms, both with reference to the joint action model and by locating the kinds of knowledge that support forms of attending and anticipatory imagery that may be specific to the genre of music being played.

#### SOCIAL AND CULTURAL DIMENSIONS OF MUSICAL ENTRAINMENT

As noted earlier, interpersonal entrainment has been associated with a variety of social factors. These range

from speculation about the role of interpersonal entrainment in human evolution and the development of societies through to current psychological research on the capacity for interpersonal entrainment to stimulate prosocial behavior and entitativity or “groupiness.” A long history of speculative evolutionary and social theory posits a key role in human sociability for entrained behaviors involving coordinated movement and sound production—in other words, music and dance. Some of these accounts stress the affective dimension of interpersonal entrainment, its role in allowing individuals to share emotional states and the role of affect in strengthening social bonds; hence these issues are covered together here. The following sections cover both social effects – i.e., the effects of IME on the formation and perception of social bonds and distinctions – and cultural factors – i.e., the ways in which IME may be practised differently in varying social settings and thus index cultural difference. The two terms are distinct, although related insofar as social bonds enable groups of people to establish distinctive, “cultural” practices.

#### PROSOCIAL BEHAVIOR AND AFFECT

A variety of studies have revealed that synchronized movement, both musical and otherwise, can affect attitudes and cooperative behaviors toward one’s co-actors. For instance, Hove and Risen (2009) found that interpersonal synchrony in a joint finger tapping task was related to greater affiliation ratings of one’s tapping partner. In a task in which participant dyads moved together in rocking chairs, Demos, Chaffin, Begosh, Daniels, and Marsh (2012) found that affiliation ratings increased when participants were more synchronized to a piece of background music, despite the fact that the music actually reduced coordination between the dyads in comparison to a silent condition. Good, Choma, and Russo (2017) demonstrated that moving in time to a musical beat with others can also influence social categorisation and cooperation across intergroup boundaries. In addition, synchronization has been found to be related to increased ratings of trust in one’s partner in a paradigm in which participants were informed they were tapping with another person but actually tapped with a virtual, computer-generated sequence (Launay, Dean, & Bailes, 2013). Such prosocial effects of moving together through music also appear to be present from a young age; for instance, Kirschner and Tomasello (2010) found joint music making can increase cooperative behavior in 4-year-old children and Cirelli et al. (2014) reported increased helping behavior toward an experimenter when 14-month-old

infants were bounced to music in synchrony with the experimenter in comparison to an out-of-sync bouncing condition. Reddish, Fischer, and Bulbulia (2013) found that group cooperation was highest when both synchrony and shared intentionality were combined in experimental tasks; in our terms, coordination depends on shared intentionality and thus we might hypothesize that both components of IME are necessary for some of music’s social effects.

Prosocial effects of synchronized movement have also been observed in relation to dance (Reddish et al., 2013; Tarr, Launay, Cohen, & Dunbar, 2015; Vicary, Sperling, Zimmermann, Richardson, & Orgs, 2017), as well as periodic activities not directly related to music, such as rowing (Cohen et al., 2010) and even walking (Miles et al., 2009; Wiltermuth & Heath, 2009). As such, it has been posited that human ability to synchronize with one another may have served a broader evolutionary function in terms of enabling social bond formation across large groups (Launay et al., 2016), facilitated by a sense of self-other merging between co-actors and endorphin release as a result of synchronized exertive movements (Tarr, Launay, & Dunbar, 2014).

The role of neurotransmitters connects the effects of synchronized movement to “*affective* entrainment”—the sharing of affective states between individuals in joint music making (e.g., Phillips-Silver & Keller, 2012). Affective entrainment may then be responsible for the formation of social bonds and prosocial behaviors that can be elicited by musical joint action, and hence implicated in the emergence of larger and more complex social groups than would otherwise be possible. However, Mogan, Fischer, and Bulbulia (2017) conducted a meta-analysis of 42 such studies of the relationship between interpersonal synchrony and four dimensions of response: prosocial behavior, perceived social bonding, social cognition, and positive affect. Their analysis suggests that phenomena that are facilitated by small-group interactions—such as exact behavioral matching—may be linked to prosocial behavior, “possibly through self-other blurring and increased attention” (Mogan et al., 2017, p. 19). Positive affect, on the other hand, increases significantly with group size and may therefore depend on a distinct mechanism.

The role of entrainment in inducing affective responses to music has been discussed more generally in music psychology: for instance, Juslin (2013) includes this factor in his BRECVEMA (Brain Stem Reflex, Rhythmic Entrainment, Evaluative Conditioning, Contagion, Visual Imagery, Episodic Memory, Musical Expectancy, and Aesthetic Judgment) model of emotion in music. Trost et al. (2017) review research in this field,

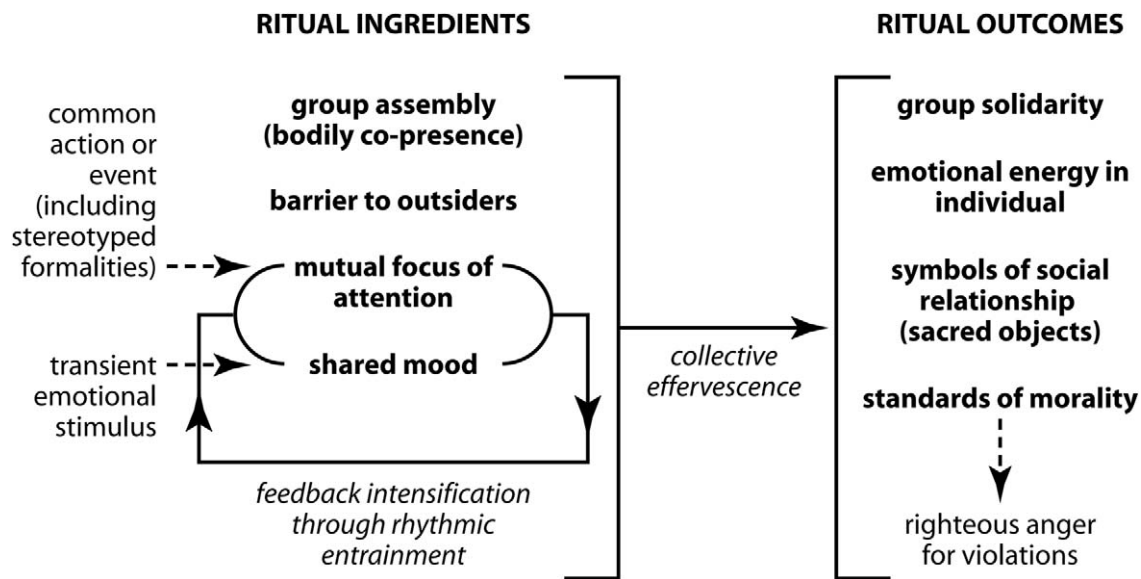


FIG. 1. Randall Collins' Interaction Ritual Chains model. Figure republished with permission of Princeton University Press from *Interaction Ritual Chains*, Randall Collins, © 2004; permission conveyed through Copyright Clearance Center, Inc.

noting the importance of distinguishing the different levels at which entrainment occurs (“neural, perceptual, autonomic physiological, motor, and social”), concluding that apart from the neural level, “all other forms of entrainment have been described as involving a kind of affective experience” (p. 106). It is not yet clear what mechanisms link entrainment and emotional or affective experiences, however, and the experimental evidence is unclear—as Mogan et al. (2017) point out, the strongest evidence comes from ethnographic observation.

#### THE SOCIOLOGY OF MUSIC AND RITUAL

For Durkheim, in his classic work *The elementary forms of the religious life* (1912/1995), face to face interactions between groups of people, in which individuals coordinate their actions towards a common focus of attention and may experience a surge in energy and positive affect he termed “collective effervescence,” was crucial to the development of social groups, belief systems, and shared symbols. These ideas can be traced in research exploring the place of bodily coordination in music performance (Clayton, Sager, & Will, 2005). Collins (2004) combines elements of Durkheim’s theory with Goffman’s (1959) equally influential microsociological approach in a theory he terms “Interaction ritual chains,” a key element of which is rhythmic entrainment. Collins’s model is encapsulated in Figure 1: a ritual event requires a group of participants, distinguished from outsiders, who agree on a mutual focus of attention and come to share an

emotional state. A positive feedback loop of attentional focus and emotional intensification is facilitated by rhythmic entrainment. The resulting “collective effervescence” leads to the kind of short and long-term (through repetition) outcomes of which Durkheim wrote: the development of group identification and solidarity, moral codes, and symbols of social relationship and affiliation. Collins’ updated model has not been tested extensively on musical case studies (see Heider & Warner, 2010, for a case study on shape note singing). Importantly, however, it points to the significance of a longitudinal perspective, and the role of learning and transmission in enabling the repetition of specific patterns of IME which form an essential part of cultural expressions.

#### SOCIAL DIFFERENTIATION: ROLE, LEADERSHIP, AND OUT-GROUPS

The social significance of IME is not limited to its effectiveness in promoting social bonding. Apart from the short and long-term formation of social groups we should consider that groups tend to be defined in contrast to out-groups. The out-group is often not present at the time of musical and ritual performance, but Lucas, Clayton, and Leante (2011) illustrate the case of groups refusing to entrain with out-groups in a ritual context. Moreover, within the group there almost always exists some form of differentiation of role and/or status. Different instrument types or voice ranges may be conceived as complementary, but are often conceived as

organized hierarchically, with one or more individuals assuming a leadership role: this is an important concept in the social aspects of IME.

The question of ensemble leadership has been empirically explored using various operationalizations of the concept, including by investigating the temporal relationship between sound or movement onsets from co-performers (e.g., the person who tends to move or produce their sounds first is assumed to be the leader), by contrasting roles that are assumed to naturally vary in terms of their leadership properties (e.g., first vs. second violinist of a string quartet), or by assigning leader/follower roles to performers within an experimental paradigm. In the auditory domain, several studies of Western classical music have reported the presence of “melody lead,” i.e., the tendency for the melody line to be played slightly ahead of the accompaniment (Keller & Appel, 2010; Palmer, 1989; Rasch, 1988), although it remains to be investigated whether such a feature generalizes to other styles of music (see the section *Measuring Entrainment in Musical Ensembles*). In an experiment in which leadership roles were explicitly assigned within piano duos, Goebel and Palmer (2009) found that leaders played slightly ahead of followers on average, but also revealed that both performers tended to mutually adjust their timing to one another (despite the leader/follower designation). Wing, Endo, Bradbury, and Vorberg (2014) found that in addition to playing slightly ahead of her co-performers, the leader of a string quartet (first violinist) applied less temporal error correction in her playing than the other ensemble members, suggesting that the rest of the ensemble adjusted their timing to her part. However, this error correction result was not replicated in a second string quartet within the same study, in which all members of the quartet demonstrated more equal levels of error correction, indicating differences in terms of leadership styles between the two quartets. In terms of investigating such individual differences in musical leadership styles, one study by Fairhurst, Janata, and Keller (2014) investigated leadership tendencies by asking participants to tap in synchrony with an auditory stimulus that varied in the degree to which it temporally adapted to the participant. They found different behavioral patterns between subgroups of participants (“leaders” felt the task of synchronizing was easier when they felt more in control and vice versa for “followers”), which was reflected in differences in brain activity in areas involved in self-initiated action.

Another body of research has focused on the development of innovative methods for studying ensemble leadership using movement data from performers. For

instance, Varni et al. (2010) developed a computational model to compute a leadership index in violin duos and string quartets in real-time, starting from cues in head movements. D’Ausilio et al. (2012) used kinematic data from violinists’ bows and conductors’ batons to investigate orchestral leadership patterns in terms of the influence of a conductor on the violinists and the violinists on each other (listener preferences were also investigated; see the section *IME and Listener Preferences*). Glowinski et al. (2012) used a similar method to investigate leadership in string quartets via head positional data, and found differences between the ensemble players (e.g., the first violinist led more frequently) and differences between sections of the musical piece (e.g., some sections comprised one clear leader whereas others had more distributed leadership, which roughly corresponded to the musical structure). One area that remains to be explored further is how the kinematic properties that are used to index leadership in these studies relate to note-to-note synchronization in the auditory domain. This question has been partially explored, for instance, by Bishop and Goebel (2017), who found that piano duos exhibited greater synchronization in terms of note onsets when the leader (as assigned by the experimenter) used head cue gestures that were smoother, greater in magnitude, and less prototypical.

Finally, music listening experiments have revealed certain perceptual biases in the assessment of ensemble leadership. Uhlig, Fairhurst, and Keller (2013) asked pianists to judge leader-follower relations in different versions of a piano duet consisting of a melody and accompaniment part. The original performance contained natural, local variations in leader-follower phase relations, but no systematic phase difference at the global level (i.e., the median asynchrony was 0 ms). In addition to this natural version, synthesized versions of the performance were created, which introduced a global melody or accompaniment lead (28 ms on average). As expected, the natural version without global leader-follower relations attracted intermediate ratings. However, for the synthetic versions, listeners were able to detect when the melody led the accompaniment but not when the accompaniment led, suggesting a perceptual asymmetry whereby listeners are especially sensitive to melody lead. Ragert, Fairhurst, and Keller (2014) extended this work by comparing leader-follower judgements in the context of the natural performance (containing local but not global leader-follower relations) and a synthetic rendition of the duet without asynchronies (i.e., no leader-follower relations even at the local level). Results indicated that listeners were biased towards hearing the melody as leading in the synthetic



performance (without asynchronies) when attention was focused on the melody. This bias did not occur when natural fluctuations in asynchrony were present (consistent with Uhlig et al., 2013) and no bias towards hearing the accompaniment as leading was found when attention was focused on the accompaniment. Taken together, the results of these studies suggest that listeners are not only sensitive to melody lead, but they may be biased towards perceiving it when it is in fact absent under certain circumstances.

#### THE ROLE OF SOCIAL AND CULTURAL FACTORS IN A MODEL OF IME

A comprehensive model of IME needs to take account of the great diversity of performance configurations that may be observed in musical performances across the world. In order to consider either the social efficacy or the cultural variability of IME, we need to consider different aspects of the performance situation. In practical terms, when characterizing the performance space of music ensembles as sites for interpersonal entrainment, we may take into account the following factors:

- Group size: The number of participants in an event could range from two to thousands. The definition of “participant” is broader than that of “musician,” including anyone whose actions in the performance space can influence others. Increasing group size brings new possibilities and challenges in interaction, potentially changing group dynamics and the generation of positive affect (Mogan et al., 2017; Moreland, Levine, & Wingert, 1996).
- Spatial organization: How is the group distributed in space? Is it set apart from, or higher than, other co-present individuals (such as an audience)? Who faces who, and who can communicate with whom? The orientation and attention of members towards each other, distances between people and lines of sight, hearing, and touch through which information may be exchanged can affect group dynamics; there may also be effects of acoustic delays due to distances between performers.
- Subgroups: Participants may be organized into subgroups (in large gatherings, groups may be identified as “the choir,” “the audience”; a choir may be divided into vocal ranges, orchestras into sections, etc.). Ways of differentiating participants into groups are diverse (e.g., in some cases rather than “musicians”/“dancers,” a more pertinent distinction to participants might be “ritualists”/ “lay persons”).
- Role: Participants may be differentiated by role (conductor and musicians, soloist and accompanist, etc.). This varies cross-culturally and by genre: “leader” may not be recognized in a highly egalitarian society, for example; the extent to which “listeners” are considered to be active participants in the performance may also vary.
- Leadership structure: Does a single individual have overall responsibility for the performance, is the leadership distributed, or can it be contested or change over time? Is there a distinction between musical and social “leadership”? What is at stake socially in the leading of the ensemble?
- Participation: This factor includes the relative closed-ness of the group (how easily can individuals enter or leave, or switch roles within the group). Who is “part of” the event, and who is a non-participant who happens to share the physical space (e.g., to be within sight and hearing of it)? Are members selected by skill, social class, or gender?
- Technology: The types of sound-producing instruments available and the role of technological mediation (e.g., amplification and sound monitoring) may also impact on IME.
- Knowledge: The organization of activities in terms of shared representations, structures, processes, and goals.

Some of these factors have a clear impact on the kind of interpersonal entrainment that can be manifested. For instance, more complex patterns of entrainment are possible with 50 musicians than with 2, and the style of interaction exhibited by a quartet of musicians each with independent parts will be different than that exhibited by a choir singing in unison and guided by a conductor. There are also more subtle distinctions that may nonetheless have a significant effect on ensemble coordination. Does a quartet of nominally equal musicians entrain differently from a quartet with a clearly denoted leader, for example? These questions are impossible to answer without further empirical study.

Rather than attempt to identify archetypal performance situations, we propose rather to regard the performance “space” as flexible and reconfigurable (and its boundaries in many cases permeable); and to consider any of the factors above as potentially significant for IME in any given event. Taking all of this into account, we focus analytically in the section *Measuring Entrainment in Musical Ensembles* on the individuals who can be observed to be interacting in its musical aspects: those singing or playing instruments, whose movements may be coordinated with the music and

each other. We aim to model the observable interactions between individuals in different modalities, particularly sound and vision. Understanding of the specific cultural frameworks within which a performance takes place is important, even indispensable, in interpreting these interactions, but this should not constrain the investigation unduly (e.g., in Western concert music, consideration of audience members as participants is not discounted a priori simply because they are not considered to be part of the performing group). Understanding IME involves not only quantitative analysis of observable aspects of entrainment, but a qualitative understanding of as many of these factors as possible, which in many cases is accessible through ethnographic enquiry.

All of the factors above can be regarded as **social** in the sense that they define who interacts with who, and how. Particular social groups decide who should gather, in what numbers and locations, and how their music-making should be organized. Instruments may be dictated by either the material available locally, or the resources to purchase items manufactured elsewhere. The resources available to the group, and their distribution, influence the likelihood that the group is organized hierarchically and of the performers being specialized (since this requires their having the resources and time required to develop musical expertise above the level possible for someone occupied full-time in employment or subsistence activities). In the broadest sense, IME is also **cultural** to the extent that interactions between participants are observably different in different situations, whether that is due to instrument choices, gender roles, or the organization of the means of economic production.

It is nonetheless important to flag up the last, but not the least important, item on the list of factors affecting entrainment above: Knowledge. For example, anticipatory processes in the ADAM model are built on the idea of action simulation, implicating internal models of the relationship between motor commands and their effects (van der Steen & Keller, 2013). This alone does not strongly imply a cultural dimension, even if the specific kinds of internal models people develop are no doubt related to the kinds of objects with which they interact in their environments. “Rich BPS” incorporates internal representations of complex hierarchical timing patterns and their influence on perception, while Rimmele et al.’s (2018) approach suggests that symbolic factors encoded in memory can influence entrainment, referring to phase resetting in particular. Consideration of how such representations are shaped by broader cultural aspects, such as those enumerated above, would link the neural

level of IME to the sociocultural. Understanding what specific representations may be pertinent in the context of IME requires not only bottom-up inference from timing patterns, but also musicological and ethnographic knowledge that can illuminate the ways in which they are represented and learned. Such shared representations are likely to include short-term patterns such as typical meters and how they can be inferred from acoustic signals, but also longer term processes: it is necessary to think from a musicological perspective about the complex plans groups of people develop regarding their musical performances, and the ways in which different parts need to coordinate in order to achieve an appropriate result. This is linked to longer-term expectations that can be deliberately planned for and managed, and these plans are very much specific to the cultural environment, indeed often to a particular item of repertory recognized by the culture. IME thus involves participants being able to plan and anticipate, and this involves the modeling or representation of musical performance: not only what has been played and shapes expectations through internal representation of temporal structure, but what might be or what ought to be produced in the foreseeable future. This involves representations of knowledge, which may be externalized and given physical form (such as scores), may be explicitly known and represented by oral notations, or may be elements of practice so well-drilled that they do not need to be written down or explicitly taught. Such patterns can be described as culturally shared and acquired knowledge, and where we find cultural difference in music we are likely to find such representations. What is not clear at this stage is how, and to what extent, these representations shape IME in practice.

Another possible area of cultural variability is: to what extent do groups consciously manipulate the tightness of their mutual entrainment? How precisely synchronized is *good enough* in a given context, and how much is *optimal*? This is likely to be benchmarked according to a number of factors, including what is considered achievable—we don’t expect a primary school band to be as tightly entrained as a professional orchestra because we know it is not possible to achieve this, and similarly many participatory genres may accept variability in synchrony because the only alternative would be to be more selective about who can take part. In such situations there will also be a trade-off with the level of complexity of the musical material, however: is a higher level of synchrony with a simpler musical texture preferred to a lower level of synchrony playing more challenging music? This is another way in which a society’s values may impinge on musical entrainment.

In summary, existing models of SMS provide clues as to how IME is on the one hand a shared human capacity, while on the other accessible to manipulation and perhaps cultural variability (via conscious period correction, phase resetting, and complex internal representation of temporal hierarchies). However, in order to robustly connect neurophysiological models to an understanding of sociocultural factors, new research is required. This should involve both empirical exploration of cultural variability in IME (exemplified in the section *Measuring Entrainment in Musical Ensembles*), and conceptual modeling of the interactions between these levels (see section *Models and Predictions*). This endeavor requires both theoretical models drawing on sociological theory, and ethnomusicological research that can both interpret diverse musical structures and contextualize these structures in social and ritual practice.

#### SUMMARY

Not only is entrainment a fundamental part of music making, but the ability to entrain to music is a fundamental part of the human experience. Mechanisms that underlie musical entrainment ability are present from an early stage in the human lifespan, and the unique combination of human perceptual skills, cognitive abilities, and refined audiomotor integration capacity have allowed for the development of a sophisticated system for the coordination and exchange of auditory and related (visual, haptic, etc.) information between performers that we know as music.

Music elicits spontaneous motor and physiological responses in its performers and listeners, including entrainment of neural populations and changes in heart and respiration rates, which can influence cognitive evaluations of the experience that manifest as emotional responses or social affiliation between co-performers. The expression of our capacity for entrainment through music is likely to have played a significant role in the development of our species, particularly in facilitating more complex forms of social organization and collective identities. The ways in which this process motivates musical behavior and underlies processes of social affiliation and cultural expression in the present is of profound interest to anthropology and sociology.

The discussion in the section *Foundations of Musical Entrainment* moved from a survey of the evolutionary, developmental, and neurological underpinnings of IME to consider the state of the art in terms of models, both linear and nonlinear. We emphasized in particular models that include top-down processing: either an element of anticipation and planning (ADAM) or internal

representation of complex patterns contributing to top-down perception (“active sensing”). Building on these models, more needs to be done to integrate both longer-term processes and the role of culturally specific knowledge representations. The most promising biological explanations for social effects such as group bonding refer to the role of neurotransmitters such as oxytocin and endorphins, which are linked to both physical movement and the detection of human agency (Launay et al., 2016). Although much remains to be explained, the connection between entrainment, affect, and social bonding appears to be consistent with long-standing sociological speculation (Collins, 2004).

The section *Social and Cultural Dimensions of Musical Entrainment* explored issues concerning the social and the cultural aspects of IME, with a view to developing a model of IME in these respects. We briefly surveyed findings in psychology regarding the effects of IME on prosocial behavior and groupiness, and sociological and anthropological theories linking IME to the social functions of ritual. Studies of ensemble leadership, being a key way in which members are differentiated and hierarchies expressed, highlight a major factor in social *differentiation*. In the absence of significant literature linking complex social and cultural factors to IME, we concluded this section by discussing a few factors that may be taken into account in future models. We flagged up in particular, the importance of culturally shared knowledge representations that allow participants to plan and anticipate.

In the next section we present analyses of the two components of IME distinguished in Section 1: synchronization (at relatively short timescales), and coordination (at longer timescales, based on shared understanding of metrical and formal structures and processes). Ragert et al. (2013) suggested in a study of piano duos that these two ways of coordinating ensembles may be dissociated, in the sense that at short timescales synchronization accuracy is dependent on a musician’s ability to make predictions based on a co-performer’s playing style, while for longer-scale coordination, familiarity with the structure of the piece alone aids accuracy. As argued in the Introduction of this paper, since they operate at different timescales and favor different sensory modalities, it is logical to consider them as distinct components of IME (although we can also explore ways in which they may overlap or influence each other; for example, as noted in the section *Models of Synchronization and Beat Perception*, the period correction aspect of synchronization incorporates conscious control). These two processes may be present to different degrees in different situations. In

music that can be characterized to a large degree in terms of the relationship between percussive event onsets of different timbral qualities, as in many drum ensembles, the coordination element includes the necessity of aligning metrical structures (so that ostinato patterns interlock correctly, for example), and a mechanism for ensuring any section transitions and tempo changes are coordinated. At the other extreme, music comprising only sounds with slow attacks in which clear onset times cannot be determined, such as some bowed string sounds, can hardly be said to use a synchronization mechanism based on onset identification: and yet such ensembles are generally coordinated in relation to agreed procedures or formal principles governing the ways in which parts relate to each other temporally. In many cases different instrument types combine, so for example a legato vocal or string part is coordinated with a percussive drum pattern: the former may align with the latter taking into account the synchronization of endogenous neural oscillations to the drum part, but this synchronization will not be unambiguously evidenced in their own part if event onsets are not clearly defined.

The analyses in the following section draw on diverse corpora, including North Indian raga, Malian jembe and Uruguayan candombe (both Afrogenic drum ensemble traditions), Cuban son and salsa, stambeli ritual music from Tunisia, European string quartets and “Improvising duos” (standard jazz and free improvisation).<sup>3</sup> All are small group traditions (ensemble size 2–7). We concentrate on interactions between musicians, rather than taking into account audiences or dancers, and focus our comparative analysis on aspects that are readily quantifiable from recordings of natural performances, such as onset times and gross body movements.

### Measuring Entrainment in Musical Ensembles

As noted above, although a significant amount of empirical research has been carried out on synchronization, relatively little has been done on longer-term coordination, and in either case almost none comparing different musical genres cross-culturally. In order to further develop models of IME that take account of cultural variability as well as including different timescales and processes of temporal alignment, in this section we present sample cross-cultural analyses. We begin with an overview of methods for measuring

entrainment in musical ensembles before presenting two different approaches to analysis. In the first we discuss the differences in synchronization parameters between genres and their possible causes; in the second, we present a comparative analysis of the coordination of ancillary body movement between performers and its relationship to musical structure. These approaches do not, of course, exhaust all possible ways of analyzing IME cross-culturally, but are used as exemplars of a comparative approach based on empirical analysis of performance data, and offer an indication of ways in which such comparison may prove fruitful.

#### MEASURING IME

Many approaches are possible for analyzing IME, depending on the types of data that are available and the timescales on which one focuses. In this section we give a brief overview of common approaches, before implementing a subset of these methods in subsequent sections using a collection of audiovisual data from a diverse range of musical cultures.

#### DATA TYPES

Studies of IME make use of various sources of data in order to examine temporal relationships between two or more co-performers in terms of the sounds they produce, movements they make (either to produce these sounds or to communicate with one another), or biological signals. Several key examples of such data are shown in Table 2. The analysis method that is subsequently applied to these data is dependent on the data type (e.g., whether the data represent discrete events/time points or continuous trajectories over time)<sup>4</sup> and research question (e.g., whether one intends to examine entrainment over long or short timescales, address leader/follower relations, etc.).

#### DATA ANALYSIS

In this section we briefly describe several key analysis methods that are commonly applied in research on IME and related work in which the aim is to examine contingencies between multiple time-dependent events or time series.

*Discrete Data Analysis.* Multiple series of discrete events such as instrumental onsets or tapping data are

<sup>3</sup> This analysis, and the collection of diverse corpora, was made possible by the “Interpersonal entrainment in music performance” project (IEMP).

<sup>4</sup> In reality, no data recorded using the methods described here are truly continuous and depend on the sampling rate of the data collection method. In addition, continuous signals may sometimes be converted to discrete data types by isolating particular events within the signal (e.g., identifying peaks of velocity changes of particular movement types from a continuous movement trajectory).



TABLE 2. *Examples of Data Commonly Collected in Studies of Interpersonal Musical Entrainment*

Data	Example data recording method(s)	Typical modality	Data type
<b>Instrumental/vocal note/event onsets</b>	Audio recordings, MIDI recordings	Auditory	Discrete
<b>Instrumental/vocal envelopes</b>	Audio recordings	Auditory	Continuous
<b>Movement trajectories</b>	Motion capture, accelerometers, video recordings, force-plate data	Spatial	Continuous
<b>Movement classes (e.g. head nod, specific hand gesture or signal)</b>	Motion capture, accelerometers, video recordings	Spatial	Discrete
<b>Brain activity</b>	Electroencephalography, magnetoencephalography	Physiological	Continuous
<b>Autonomic nervous system activity</b>	Heart rate monitor, electrocardiography (ECG), respiration monitor	Physiological	Continuous

typically analyzed in terms of the temporal asynchronies between event timings (often given in milliseconds, e.g., 0 ms asynchrony = perfect alignment). Asynchronies between events can be analyzed to give a measure of the precision of synchronization (the *SD* of the asynchronies, termed “Asynchronization” by Rasch, 1988; multiple Asynchronization measures can be combined as a Group asynchronization measure). Asynchrony data can also give a measure of two instruments’ relative position (sometimes described as “accuracy,” see the section *Sensorimotor Synchronization*), i.e., whether one part tends to play ahead or behind another (cf. “melody lead”). This approach is outlined in more detail in the section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis*.

Alternatively, circular statistics are a common, non-linear method for dealing with periodic series of discrete events (Fisher, 1993; Mardia & Jupp, 2009). The main measures of pairwise relationships are analogous to those of asynchrony analysis: the precision of synchronization (given by the mean vector length) and the accuracy or mean relative position (given by the vector angle). Phase analysis has been recommended as a primary method for entrainment analysis (e.g., Clayton et al., 2005), partly because exploring the stabilization of phase is more consistent with a dynamical systems model of entrainment than calculating the variability of asynchronies. However, using relative phase data for comparison between multiple genres is problematic: the phase calculations depend on the specification of a reference period, and different kinds of music have different types of metrical structure that make it impossible to establish a single objective definition of this period. Different reference periods produce significantly different mean phase angles and vector lengths, and therefore it is very difficult to be sure one is comparing like with like. The section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis*,

therefore, employs asynchrony analysis. In other circumstances, particularly where comparison between genres is not necessary or where patterns interlock and there are therefore few “asynchronies” to calculate, this method may be preferred.

Another method that can be applied to event-based data is event synchronization (ES; Quiroga, Kreuz, & Grassberger, 2002), in which the degree of synchronization is calculated based on the number of quasi-simultaneous appearances of events (within a specified time window) and the delay between the two event series is calculated based on the number of events in one signal that precede the other. Although the ES method does not provide the same precise phase information as circular statistical analysis, the ease with which this method can be applied makes it particularly suitable for online implementations.

*Continuous Data Analysis.* The similarity between two time series can be examined via cross-correlation, which allows one to determine the time lag at which the two series are optimally similar. Cross-correlation assumes stationarity of the data (i.e., the mean and variance are constant over time) and may not be optimal for analysis in which the individual time series are autocorrelated (i.e., do not comprise independent values; Dean & Dunsmuir, 2016). As music often affords a periodic temporal structure, another relevant technique is cross-wavelet transform (CWT) analysis,<sup>5</sup> which examines co-occurrences of periodic behaviors within two time series across multiple frequency bands. CWT is calculated as the pointwise multiplication of the wavelet transform (WT) of two individual time series (Grinsted, Moore, & Jevrejeva, 2004; Torrence

<sup>5</sup> In some fields XWT is preferred as an abbreviation for Cross-Wavelet Transform (and CWT is used for Continuous Wavelet Transform). Our usage is consistent with that of Issartel, Bardainne, Gaillot, and Marin’s (2015) study of human interaction.

& Compo, 1998; see also Issartel, Bardainne, Gaillot, & Marin, 2015, for an application within psychological research). In our case, CWT analysis enables the detection of shared periodic movements of pairs of performers across different frequencies and time, which allows us to examine movement coordination across multiple timescales, from fast head nods to slow body rotations. This technique has been previously applied in Eerola et al. (2018) to the “Improvising Duos” corpus that we also make use of in this paper; that work demonstrated that a measure of CWT Energy of performers’ ancillary movements across a broad frequency range (0.3 to 2.0 Hz) served as a significant predictor of “bouts of interaction” (i.e., periods of visually apparent communication between co-performers), as labeled by expert musicians. CWT analysis has also been applied to describe different patterns of limb and head coordination within piano duos (Walton, Richardson, Langland-Hassan, & Chemero, 2015). This approach is explained in more detail, with examples, in the summary at the end of this section.

Another analysis method that has often been applied in research on interpersonal coordination in conversation and everyday behavior (e.g., Shockley, 2005; Shockley, Richardson, & Dale, 2009) is cross-recurrence quantification analysis (CRQA). CRQA is a nonlinear method that does not assume data stationarity or periodic behavior, can deal with pairs of time series of different lengths, and can also be applied to categorical/discrete data. This method can be used to determine the presence and duration of overlap between the dynamics of two different time series by quantifying the regularity, predictability, and stability of two concurrent behavioral performances in reconstructed state space. In addition, CRQA can be used to quantify the lag at which one individual’s behavior maximally matches another’s, in order to identify whether there is a leader–follower type of relationship. CRQA can, for example, be applied to measure if two players exhibit similar patterns of behavior during a music performance. One example of an application of CRQA in music research (although not specifically focused on interpersonal entrainment) has been to use this method for automatic identification of cover songs (Serrà, Serra, & Andrzejak, 2009); a similar technique has been used to compute phase synchronization in a violin duo (Varni et al., 2010).

Finally, Granger causality is a method that uses past values of one time series to predict future values of another time series. Granger causality has been used, for instance, to investigate leadership in both string quartets (Chang, Livingstone, Bosnyak, & Trainor, 2017; Glowinski et al., 2012) and orchestras (D’Ausilio

et al., 2012) using movement data from the musicians and conductors.

One caveat that should be mentioned here is that all of the aforementioned methods rely on pairwise comparisons of time series, whereas entrainment in music performance often takes place at the level of larger groups than duos (Rasch’s Group asynchronization is unusual in this regard, although it too is based on pairwise calculations). Thus, appropriate corrections for multiple comparisons may need to be applied when performing analysis on larger groups. Some research on interpersonal coordination more generally has also explored the use of cluster phase analysis, a method based on the Kuramoto order parameter (Kuramoto, 1989), for quantifying synchronization of larger groups of participants (e.g., six participants; Richardson, Garcia, Frank, Gergor, & Marsh, 2012).

*Application of Data Analysis Techniques.* As noted above, of the various options available for analyzing IME in performance, this paper focuses on just two: (a) the exploration of synchronization using asynchronies between discrete auditory event onsets, and (b) the analysis of coordination between continuous body movement using CWT analysis. The main priorities are to demonstrate analysis of the different components of IME and to present cross-cultural analyses, in order that these comparisons may help to shape the development of a model of IME in the forthcoming *Models and Predictions* section. In order to cover the full spectrum of IME in practice, other methods will need to be deployed, either in order to address different data sources such as EEG, other physiological measures or other aspects of motion, or because the approach taken in the next section—where onset timing differences are analysed in relation to a known metrical structure—is not appropriate. This would be the case where acoustic envelopes do not lend themselves to the measurement of onset times, for example (many vocal sounds or those of bowed instruments emerge gradually without any percussive “onset”), or where the temporal structure is unpredictable, for instance because there is no regular meter.

### Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis

#### PRINCIPLES

In this section, we explore synchronization in different genres by calculating time differences between the onsets of nominally simultaneous auditory events. In the simplest model of SMS between two individuals, they are understood to both perform the same simple sequence of actions, as illustrated by sequences A and B in Figure 2.

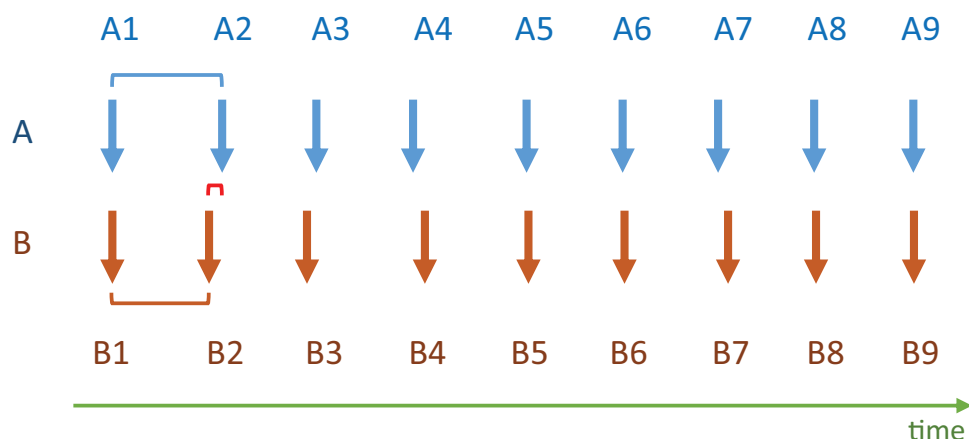


FIG. 2. Events produced by two individuals have onset times represented by the arrows in rows A and B. The blue and brown brackets indicate the interonset intervals between the first pair in each sequence; the red bracket indicates the asynchrony between onsets A2 and B2.

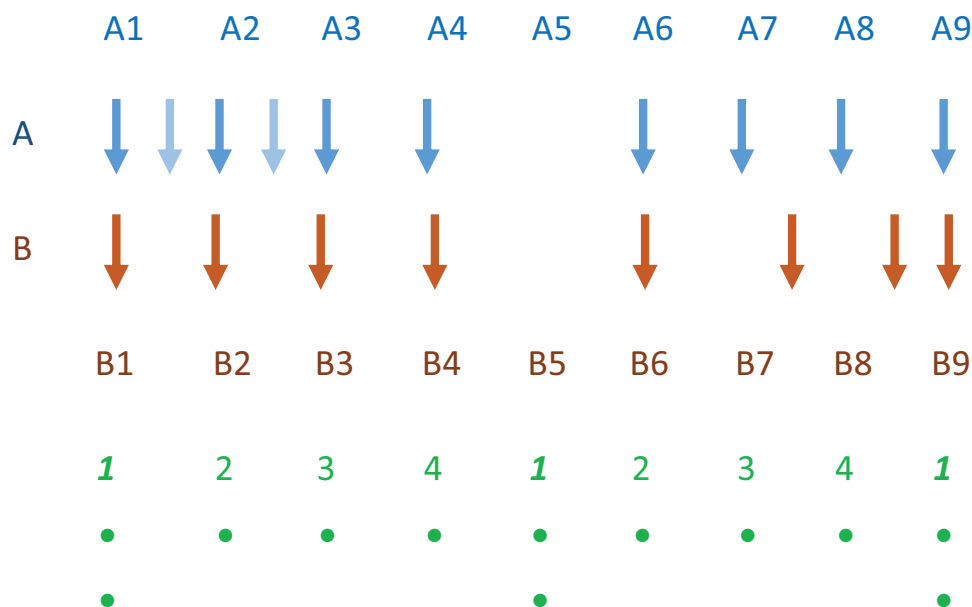


FIG. 3. Events produced by two individuals have onset times represented by the arrows in rows A and B. They are interpreted with respect to a four-beat metrical pattern (green dots).

In this case, events in the two sequences occur at approximately the same point in time: each individual adjusts to the other in order that they stay in time. Thus, B3 falls early with respect to A3; both correct this difference, as a result of which B4 falls a little later than A4.

This case of two musicians performing periodic patterns in phase with each other is a particularly simple one, of course. Almost any real-life musical example will be more complex than this. As illustrated in Figure 3: (a)

the rhythms (patterns of time intervals) may be varied, and (b) in many cases the sequences will be heard in relation to the meter (indicated by beat numbers and a hierarchical grid of pulses in green). Metrical positions are not necessarily marked by events (e.g., A5/B5), while points in between the beats may be articulated (e.g., extra event onsets are interpolated between A1 and A2, A2 and A3). In this hypothetical example, B articulates the “off-beat” in between A7 and A8, and A8 and A9. In this

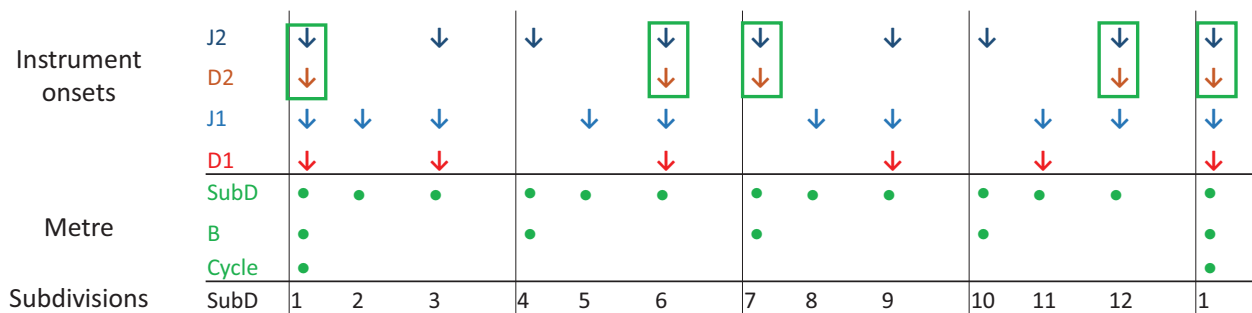


FIG. 4. The basic structure of the Malian jembe piece Manjanin. Top layer: instrument onset points (after Polak et al., 2016). D = dundun, J - jembe (types of drum). Different sounds (timbres) are not represented. Uneven horizontal spacing represents non-isochronous timing profile. Middle layer: metrical structure (B = beat, SubD = subdivision). Bottom layer: Numbering of metrical positions in the cycle by subdivision (SubD). Green boxes: points of coincidence between J2 and D2.

case, if we wish to explore SMS we cannot assume that the sets of onsets A and B mark nominally equidistant pulses; instead, we must first establish which onsets should be compared with which. Establishing the metrical structure allows us to do so, and also means that we can assess synchronization in relation to the meter (e.g., ask whether synchronization is more precise, on average, on beat 1 than on beat 2, etc.).

By meter we refer to a temporal hierarchy of cycles, beats, and subdivisions that can be inferred from the sound and act as a reference for listeners (see Table 1 above). The analyses that follow assume that in this respect meter is a general phenomenon—culture-specific in its detail but sharing some features across many cultures—and therefore that the temporal organization of North Indian raga, Malian jembe, European string quartet music, and other groups we discuss below is to this extent comparable. It is worth pointing out that metrical systems vary between the different traditions; for example in the usage of non-isochronous time intervals (see below). These traditions also vary in the ways in which these metrical structures are explicitly theorized, named, and represented—for example, North Indian musicians explicitly number the beats in their meters, while Malian jembe musicians do not—although in each case it is clear that rhythmic and metrical patterns are recognized. It is not the case that every musical performance must manifest such a structure—very many do not (Clayton, 1996). Nonetheless, in each of these six genres it is possible to identify metrical hierarchies by observing periodicities in the acoustic signal, observing behavior related to the performance, and exploring theoretical concepts expressed within the different music cultures.

Figure 4 illustrates a specific musical example, based on one of the corpora analyzed below, in a similar

graphical form to the hypothetical example in Figure 3. This figure illustrates some of the possible points at which onset times can be compared (and asynchronies calculated): green boxes highlight all the points at which the instrument parts Jembe 2 and Dundun 2 coincide (there are two instruments of each type in the ensemble). In the next section we present some analyses of asynchrony data for this and other musical repertoires, investigated by corpus, meter, instrument pair, beat position, subdivision position, and other factors.

An interesting feature of this piece is the use of a consistent non-isochronous subdivision pattern, indicated in the diagram by close spacing between the SubD 1 & 2, and larger spacing between 3 & 4. Although not all musical examples show such clear patterning, the synchronization process cannot be assumed to depend on an isochronous pulse. In these cases, we cannot model synchronization simply in terms of two isochronous pulse streams, but recognize and account for the fact that each musician must internalize the non-isochronous pattern. Nonetheless, we can still calculate asynchronies at each position and test whether the synchronization is affected in any way by the non-isochronous subdivision.

Using event onset timing information, then, we can investigate synchronization between different instruments in a musical ensemble, and this method is robust enough to cope with variability in rhythmic patterns, metrical structures and non-isochronous timing patterns. Although in cases with roughly simultaneous onsets such as these, asynchrony analysis provides a robust method affording comparison between examples, in cases where onset timings in different instrumental parts predominantly interlock rather than align, relative phase analysis is the only practical approach. The approach taken in this paper, then, is to focus first



TABLE 3. *Corpora Analyzed*

Genre	Abbr.	Origin	N	Instrumentation	Size of corpus	Dur. (min)	Researcher
<b>North Indian Raga</b>	NIR	North India	2	Sitar, sarod or guitar + tabla (tanpura drone not analyzed)	3 pairs playing 8 pieces, Mean duration = 683s ( $SD = 339$ )	91	M. Clayton, L. Leante
<b>Uruguayan Candombe</b>	UC	Uruguay	3-4	Chico, piano and repique drums	12 takes, $M = 175.5$ s ( $SD = 30.9$ )	35	L. Jure, M. Rocamora
<b>Malian Jembe</b>	MJ	Mali	2-4	Jembe and dundun drums	15 takes of 3 pieces, $M = 202$ s ( $SD = 69.1$ )	51	R. Polak
<b>Cuban Son and Salsa</b>	CSS	Cuba	7	Bass, Spanish guitar, tres, clave, bongos and other percussion, trumpet, vocals	5 songs, $M = 398$ s ( $SD = 45.5$ )	33	A. Poole
<b>Tunisian Stambeli</b>	TS	Tunisia	$\geq 4$ , 2 parts analyzed	Gumbri (lute) + shqashiq (cymbals), vocals	4 tracks comprising 8 pieces, $M = 259.8$ s ( $SD = 105.2$ )	35	R. Jankowsky
<b>European String Quartet</b>	ESQ	Europe	4	Violin x 2, viola, cello	2 takes each of 2 movements, extracts	6	M. Clayton, T. Eerola, K. Jakubowski

Note: Abbr. = corpus abbreviation; N = number of performers; Dur. = total duration.

on event onset-based methods (particularly asynchrony calculations).

#### MATERIALS

The analysis in this section employs a set of six audio-visual collections published on Open Science Framework and described in Clayton et al. (in press). The first corpus of audio recordings and derived onset timing data comprises three recordings of **North Indian Raga** (abbreviated NIR) performances on plucked string instruments with tabla drum accompaniment.<sup>6</sup> This is divided into eight distinct sections of duo performance (string instrument plus tabla). This set of recordings forms a subset of the IEMP “North Indian Raga” corpus (Clayton, Leante, & Tarsitani, 2018).

The second corpus comprises recordings of **Uruguayan Candombe** (UC) drum ensemble music (Jure, Rocamora, Tarsitani & Clayton, 2019). The corpus comprises 12 takes recorded by nine trios and three quartets. Three drums are used, named chico, piano, and repique (quartets include two repiques).

The next corpus comprises recordings of **Malian Jembe** (MJ) drum ensemble music (Polak, Tarsitani, & Clayton, 2018). The corpus comprises 15 takes: three duos, eight trios, and four quartets. Two different kinds of drum are used, named jembe and dundun (up to two of each).

The fourth corpus comprises **Cuban Son and Salsa** (CSS) music performed by the group Asere (Poole,

Tarsitani, & Clayton, 2019). The seven-piece group plays five songs on a variety of instruments: we extracted onsets for the bass guitar, Spanish guitar, tres, clave, bongo, cowbell, cajon, conga, and trumpet (but not for a few other instruments such as shakers, scraper, and crash cymbal, or the vocal parts).

The next corpus comprises recordings of **Tunisian Stambeli** (TSS) ritual music (Jankowsky, Tarsitani, & Clayton, 2019). The instruments are the gumbri (plucked lute), and three sets of shqashiq cymbals played in unison; vocal and drum parts (used in some but not all tracks) are not studied here. Four recordings are divided into eight sections.

The final corpus comprises recordings of parts of two takes each of two **European String Quartet** (ESQ) movements (Haydn Op. 76, No. 5, First Movement and Beethoven Op. 59, No. 2, Third Movement; Clayton & Tarsitani, 2019). In this case onset data were collected only from sections with predominantly staccato articulation; remaining sections require different onset detection techniques (e.g., identifying new events when the pitch changes), which makes it difficult to compare the statistics with other corpora, and for this reason are not included here.

Our intention in this section is to present summary data on the diversity of synchronization in IME, based on the measurement of asynchronies between event onset times across these varied genres. By doing so we will highlight some of the factors that appear to influence the entrainment process across multiple genres.

Event onset data were prepared by the following steps:

1. Automated extraction of event onsets for each individual part (in most cases ‘parts’ are played

<sup>6</sup> The analysis uses three pieces from the IEMP North Indian Raga corpus (Clayton et al., 2018): NIR\_ABh\_Puriya, NIR\_PrB\_Jhinjhoti and NIR\_DBh\_Malhar. A different set of performances is used for the coordination analysis in the next section.

- by a single player on a single instrument; the exception is the *shqashiq* part in TS which is played by three players in unison, whose onsets cannot be distinguished).<sup>7</sup>
2. Manual annotation of metrical downbeats (i.e., the strongest beat of the musical meter, usually counted as “1”).
  3. Alignment of selected event onsets with metrical (beat and subdivision) positions calculated from manual annotations. This was done by dividing the duration of the cycle into the relevant number of beats or subdivisions, either equally or—in cases with a consistent non-isochronous subdivision—based on average relative timings across each piece. (This process is described in more detail in Clayton et al., in press.)
  4. Output of selected onset times labeled by metrical position, with additional information including local event density (number of onsets for each instrument per second, calculated over 2-s windows). Labeled onset times were then manually checked, with problematic onsets either removed or relabeled.
  5. For the corpora in which only one pair of onsets was available, namely NIR and TS, performances were cut into Segments averaging 120s and 50s respectively; boundaries were set at the nearest metrical downbeat to the desired average duration, which was calculated in order to generate a similar number of data points for the analysis of these corpora as was available for the other corpora with a higher number of instrument pairs.

Across the collection approximately 155,000 annotated event onsets were available for analysis (selected from a significantly larger number of raw extracted onset times, since only those unambiguously falling on defined metrical positions were employed in the analysis).

#### ASYNCHRONY ANALYSES

This section summarizes a set of asynchrony analyses of the data described above. The aim is to give an overview of the patterns found in asynchrony data from groups of

musicians in diverse musical genres, and to explore some ways in which these data vary. This summary is not intended to give a definitive account of SMS in any one genre, but rather to point to trends and preliminary findings for further analysis.

Our starting point for synchronization analysis is provided by the measures defined by Rasch (1988). The primary measure of the **precision** or **tightness of synchronization** in any example is termed *Asynchronization* (low asynchronization = high precision), and this can be calculated pairwise (as the standard deviation of the asynchronies of any pair of instruments in the music) or groupwise (calculating the RMS of the pairwise measurements to provide a group measure of synchrony). The terms “pairwise asynchronization” and “group asynchronization” are based on Rasch’s descriptions of his measures. The **Mean absolute asynchrony**, calculated from the same asynchrony data, offers an alternative measure of precision (see *Metrical Position* with the section *Relationship of Asynchrony to Event Density and Metrical Position* for details).

As well as these measures of the precision of synchrony, we also calculate the **relative positions** of the different instrumental parts. Again, Rasch provides a starting point: **Mean relative onset** is calculated as the mean position of an instrument’s onsets relative to the average position of the group.

In addition to Rasch’s measures, we also calculate the **Mean pairwise asynchrony**, which calculates the relative position of two instruments directly from their simultaneous onsets. This measure is useful in determining, for example, whether the relative positions of two instruments varies with other factors such as tempo or event density. These measures are summarized in Table 4.

*Group Asynchronization (A) and Pairwise Asynchronization (Precision).* **Group asynchronization** provides the simplest way to compare the precision of synchronization across an ensemble. Rasch’s (1988) group asynchronization values for performances by three Western classical trios (one of recorders, one of reed instruments, and the last of strings) ranged from 29–51 ms. Our values range somewhat lower, 15.6–34 ms, the highest value being from the ESQ corpus (see Table 5). This may indicate that Western chamber ensembles tend to be looser in synchrony than most of the cases in our collection.<sup>8</sup>

<sup>7</sup> Our analysis focuses on physical onsets, based on acoustic rise times, rather than perceptual onsets (p-centers), which are also affected by duration, pitch, and timbre. This is a pragmatic choice in as much as the analysis includes several instruments for which new research would be required in order to reliably estimate p-centers. Whereas for most drum sounds analyzed here there is likely to be no significant difference between physical and perceptual onset, for others (e.g., bowed instruments) the difference is likely to be more significant, and this factor should be taken into account in future studies of synchronization of sounds with more gradual acoustic rise times.

<sup>8</sup> One difference between Rasch’s calculations (1988) and ours is the temporal resolution of the onset detection: for Rasch this was limited to a 5 ms bin size, a limitation that does not apply to our onset detection. This may have an effect on the asynchronization calculations.

TABLE 4. *Measures of Synchronization*

PRECISION	Pairwise asynchronization	= <i>SD</i> of the onset time differences of simultaneous sounds of two parts (such as between A1 and B1 in Figure 3; Rasch, 1988, p. 73)
	Group asynchronization (A)	= RMS of all Pairwise asynchronization values (Rasch, 1988, p. 74)
	Mean absolute asynchrony	Calculated as the mean of all unsigned asynchrony values
	Mean vector length ( <i>r</i> )	Circular statistics measure of precision; scale from 0 – 1
ACCURACY (RELATIVE POSITION)	Mean relative position	Mean of one instrument's onsets relative to the group's mean position
	Mean pairwise asynchrony	Relative position of two instruments calculated as the mean difference in their onsets, i.e., the signed asynchrony values
	Mean vector angle ( $\mu$ )	Circular statistics measure of relative position (i.e., relative phase)

TABLE 5. *Group Asynchronization (A) and Pairwise Synchronization Across Six Corpora*

Corpus	Number of parts	A (ms)	Range of pairwise asynchronization (ms)
MJ	4	15.6	12.6 – 18.8
UC	4	18.1	16.4 – 20.4
CSS	7	24.4	13.1 – 33.4
TS	2	28.0	18.7 – 34.7 [Segments]
NIR	2	29.1	18.5 – 54.0 [Segments]
ESQ	4	35.2	31.6 – 38.2

Note: NIR and TS performances were split into roughly equal Segments (c. 120 s and c. 50 s respectively), defined in order to generate a similar number of data points for these as for the other corpora, for which the range of Pairwise asynchronization figures refer to different instrument pairs. (See Table 6.)

Comparing the values in our collection descriptively, the tightest values are provided by drum ensembles of African origin, MJ (15.6 ms) and UC (18 ms). The two genres featuring plucked stringed instruments with drum or percussion accompaniment are looser (NIR, 29.1 ms and TS, 29.6 ms). CSS, which combines a percussion ensemble with plucked and strummed strings, lies between these two ranges (24.8 ms). Each of these examples shows notable variation however, between Segments of performances (TS, and particularly NIR), and between instrument pairs (CSS). Looking at the range of **Pairwise asynchronization** values for the groups of more than two parts reveals a wider range for CSS. The tightest pair, the conga and cowbell, exhibit a lower pairwise asynchronization value ( $SD = 13.1$  ms) than the group average for the two Afrogenic drum ensembles, while the loosest, the tres (lute) and trumpet, is much looser (33.4 ms).

*Mean Relative Onsets and Melody Lead (Accuracy).* Distinct from the question of how tightly synchronized the groups are is the mean position of each instrument with respect to the group. In Rasch's (1988) study the ranges are small, up to about 5 ms. They display a tendency for lead melodic instruments to fall ahead of

other instruments. For example, in a string trio the violins both play a couple of milliseconds early, the viola and cello late. In our study the range of **Mean relative onsets** is somewhat larger in the CSS and especially the TS corpora (see Figure 5). The NIR, MJ, and UC examples show all instruments within about 3 ms of the mean. The CSS group shows, on average, the cowbell and guitar playing about 5 ms after the mean onset and the tres 9 ms early (the tres plays a mixture of rhythm and lead material, so this could reflect a "melody lead").

Note that in the MJ corpus "jembe 1" can be considered the lead instrument, and it does play slightly ahead of the other drums, on average; the situation is less clear-cut in the UC drum ensemble as the "lead" can switch between instruments. In a string quartet the first violin is regarded as the overall leader, but the melody often moves between instruments: in our ESQ data violin 2 is slightly ahead of violin 1 on average, although both are ahead of the lower-pitched instruments. These summary data can of course hide variability within each corpus. In the NIR corpus the mean pairwise asynchrony across all examples is close to 0 ms: however, if we break the pieces down into Segments of about 2 minutes each ( $M = 120.2$  s,  $SD = 8.2$ ), the mean asynchrony between the melody instrument and the tabla accompaniment varies between +13.8 ms (melody lag) and -16.2 ms (melody lead). Thus, examples of both melody lead and melody lag are seen in NIR instrumental duos: more analysis would be needed to explore the musical factors that account for these variations (see Clayton et al., 2019).

#### RELATIONSHIP OF ASYNCHRONY TO EVENT DENSITY AND METRICAL POSITION

If we look in more detail at the data for individual genres or pieces, it becomes clear that local patterns of asynchrony vary with a number of other factors. In this section we consider two factors that can be investigated in all corpora, namely event density and metrical

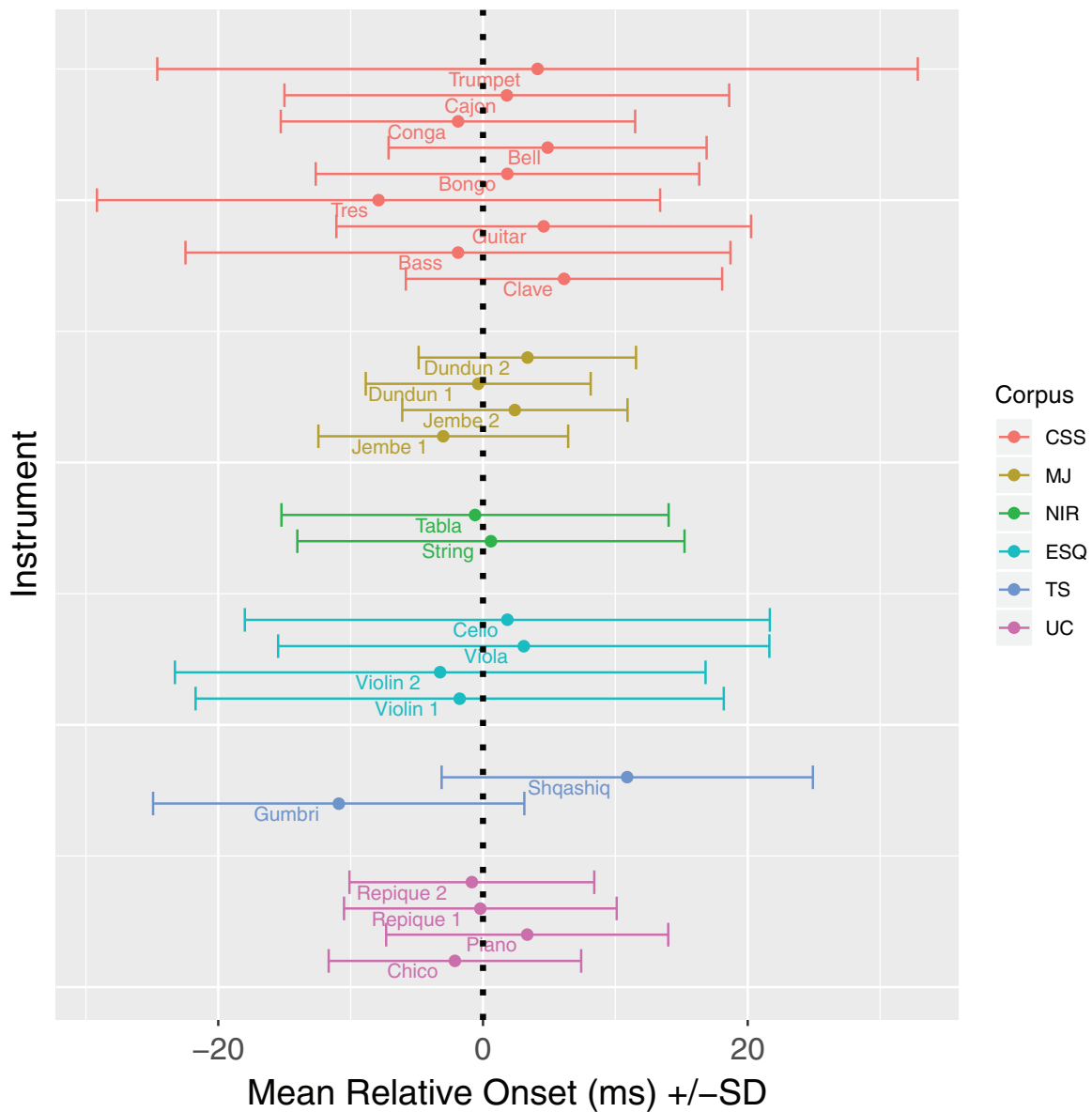


FIG. 5. Mean relative onsets ( $\pm$  SD) of instruments in six corpora.

position. These are unlikely to be the only factors that are associated with differences in asynchronies between musicians, but they serve to illustrate the kind and scale of differences that we encounter.

*Tempo and Event Density.* To explore the possible relationship between variability in asynchrony and tempo, several different factors need to be taken into account. Tempo (usually calculated in beats per minute) is a measure of the rate at which a conductor or listener would mark the “beat” of the music. This measure is problematic, however, since the selection of a reference

pulse level as the beat is subjective: two individuals may select different pulse levels in a 2:1 ratio, in which case one would estimate “tempo” as twice as high as the other. It is not possible to determine the appropriate beat level—referred to in musicological literature as the “tactus”—in a completely objective way, which creates problems for comparative analysis. (Within genres however, it is often possible to test for the dependency of asynchrony on tempo: see Clayton, Jakubowski, & Eerola, 2019.) Our judgement of how fast a piece of music seems to be is also dependent, however, on the rhythmic



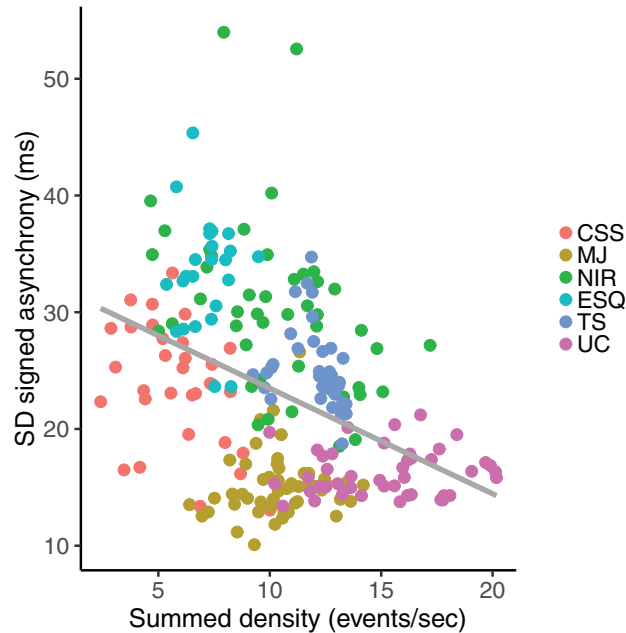


FIG. 6. Pairwise asynchronization (*SD* of pairwise asynchrony) plotted vs. event density for each pair of musicians across all corpora.

event density (i.e., the number of distinct event onsets per second), and this factor *can* be estimated objectively. (The limitation on this is the need to define exactly what qualifies as an event; onset detection algorithms require the setting of thresholds, on which the number of events counted depends. According to the analysis reported in Clayton et al. (in press), the mean proportion of event onsets missed in this process across samples from all 6 corpora was 4%, and the proportion of false positives 6.5%; the NIR corpus produced the highest error rates (up to 23%) due to the large dynamic range of the instrumental sounds.)

Figure 6 plots the Pairwise asynchronization figures for our collection against the summed event densities across the different corpora. The data represented by each point differ between corpora, with the aim being to compare a similar number of data points for each corpus (see Table 6).

We explored the relationship between asynchrony and event density by correlating Pairwise asynchronization with the mean summed event density (i.e., the sum of the mean event densities of the instrument pair being considered) across summary data (pieces, takes, or Segments). Analyzing the summary data points in this way suggests that there is an overall correlation between higher density and lower Pairwise asynchronization,  $r(234) = -.33, p < .001$ ; thus, examples with a greater number of rhythmic events per second exhibit tighter

TABLE 6. Number and Derivation for Summary Data Points Used in Comparative Analysis

Corpus	Summary data points	Method
NIR	45	Each piece divided into roughly equal Segments, rounded up to nearest metrical cycle (duration $M = 120.2s$ , $SD = 8.2s$ ).
UC	45	One data point for each pair of instruments in each take
MJ	51	One data point for each pair of instruments in each take
CSS	32	One data point for each pair of instruments across the corpus <sup>9</sup>
TS	39	Each piece divided into roughly equal Segments (duration $M = 50s$ , $SD = 3.7s$ ), rounded up to nearest metrical cycle
ESQ	24	One data point for each pair of instruments in each take

TABLE 7. Correlations Between Asynchronization (*SD* Pairwise Asynchrony) and Summed Event Density

Data		Correlation of Pairwise asynchronization with Summed event density
By instrument types	All data	$r(234) = -.41, p < .001^{***}$
	Drum/ percussion only	$r(101) = .09, p = .396$
	Melody + drum pairs	$r(101) = -.23, p = .019^*$
	Melody + melody pairs	$r(28) = -.16, p = .389$
By corpus	NIR	$r(43) = -.40, p = .007^{**}$
	UC	$r(43) = .05, p = .730$
	MJ	$r(49) = .12, p = .385$
	CSS	$r(30) = -.40, p = .024^*$
	TS	$r(37) = -.32, p = .047^*$
	ESQ	$r(22) = .05, p = .814$

synchronization. The patterns vary between corpora and between instrument types, however (see Table 7). Pairs involving drums and percussion instruments only, for instance, show no such correlation,  $r(101) = .09, p = .37$ . Tighter synchronization in faster passages could be due to the so-called subdivision benefit related to the scalar variability of a central timekeeper (Repp, 2003); however, the lack of such an effect with drum pairs remains to be explained. The corpora that show an overall correlation between summed event density and Pairwise asynchronization are the NIR,  $r(43) = -.40, p = .007, p < .01$ ; CSS,  $r(30) = -.40, p = .024$ ; and TS,

<sup>9</sup> All pairings were included for which at least 50 data points were available, with the exception of the cowbell-cajon pair, since these instruments are played by the same musician.

$r(37) = -.32, p = .047$ —those which feature a combination of melody and percussion instruments.

The correlation between mean pairwise asynchrony (relative position) and summed event density was calculated for cases featuring a melody instrument with a drum or percussion instrument, which includes the NIR and TS corpora and some pairs of CSS only. The overall correlation for this data is  $r(101) = -.30, p = .002$ , indicating that, in general, the melody instrument plays further ahead with an increase in event density. We find this correlation to be significant in the TS, where the gumbri lute plays further ahead in denser passages,  $r(37) = -.67, p < .001$ ; for NIR,  $r(43) = .22, p = .147$ . There may be significant correlations for specific instrument pairings in the other corpora that would be revealed by more detailed analysis.

**Metrical Position.** Is there any relationship between asynchrony and metrical position? For example, are pairs of musicians more precisely synchronized on downbeats than on other beats, or tighter on main beats than offbeats or subdivisions? If we regard musical meter as a form of attentional behavior, as London argues (2012), and if greater attention on specific moments in time is related to tighter synchronization at those points, then this is what we would hypothesize. Patel, Iversen, Chen, and Repp (2005) found that listeners synchronized more accurately (with smaller mean pairwise asynchronies) on beat 1 than on other beats when tapping to certain “metrical” stimuli (taking the form of sequences of tones with IOI patterns likely to evoke a metrical percept). Keller and Repp (2005) found that when tapping offbeats in time with metrically structured pacing sequences (sequences in which every 2nd, 3<sup>rd</sup>, or 4th tone is different to the others) finger taps are delayed at downbeats relative to other beats: they explain this by hypothesizing “a regularly applied, meter based phase-resetting mechanism stabilizes syncopation” (p. 292). Keller and Repp’s finding is echoed by Friberg and Sundström’s demonstration (2002) that in a selection of jazz recordings, soloists played more accurately and more precisely on offbeats than on downbeats (they tend to play later and with greater variability than the ride cymbal on downbeats, the effect being larger at slower tempi). Polak et al. (2016) refer to the possible variation of asynchrony with subdivision position in the MJ corpus, demonstrating an effect of subdivision position on variability of the IOIs (the first subdivision of each beat is found to be less variable in duration than the others). It remains to be demonstrated, however, whether an effect of metrical position on synchronization can be generalized across musical genres in performance.

In this subsection we present the results of analysis of the effect of metrical position on synchronization across five corpora (The ESQ corpus is not included here due to the small amount of data, but full results for all corpora and pairings can be found in Supplementary Table 1 online at [mp.ucpress.edu](http://mp.ucpress.edu)). Do either (a) precision or (b) relative position vary by metrical position? “Metrical position” here can be represented by diverse factors in the different corpora. We compared “strong” and “weak” metrical positions at one or more levels in each of the corpora, as follows:

- UC, CSS, and MJ corpora, all at two levels: (1) Beat 1 vs. the other beats, and (2) The four main beats vs. their subdivisions. Results are presented for each pairing separately.
- NIR: Only examples in the 16-beat cycle “teental” were analyzed, but these were separated into slow (vilambit), medium, and fast (drut) tempo pieces. They were analysed at two levels: (1) The four “vibhag” beats (those strong beats marked by hand gestures) vs. other beats, and (2) Beat 1 vs. the other three “vibhag” beats. Results are presented for different tempi separately.
- TS, at one level: Beat one (the metrical downbeat) vs. other beats. Separate analyses are given for the four different rhythmic patterns on which the eight pieces are based.

Since in many cases there are a greatly different number of data points for weak as opposed to strong metrical positions, we randomly sampled (without replacement) an equal number of asynchronies from the larger set as the number in the smaller set, for each analysis. This analysis was repeated 1,000 times, and mean differences, 95% confidence intervals (CIs), and  $p$  values were computed in each case. In this analysis we used mean absolute asynchronies for the precision measure, in order to use an analogous sampling procedure for both signed and absolute asynchronies (rather than using a single, summarized  $SD$  calculation for each piece). The full results of these analyses are reported in Supplementary Table 1.

The analysis of the difference in **precision** (mean absolute asynchrony) between strong and weak metrical positions shows many examples of significant differences across the corpus (see Figures 7 and 8). However, these differences are fairly evenly divided between examples in which strong beat positions are more precisely synchronized than weak beats and vice versa (tighter synchronization on weak beats). Many of the differences are also small in magnitude. Examples of some of the larger effects are:

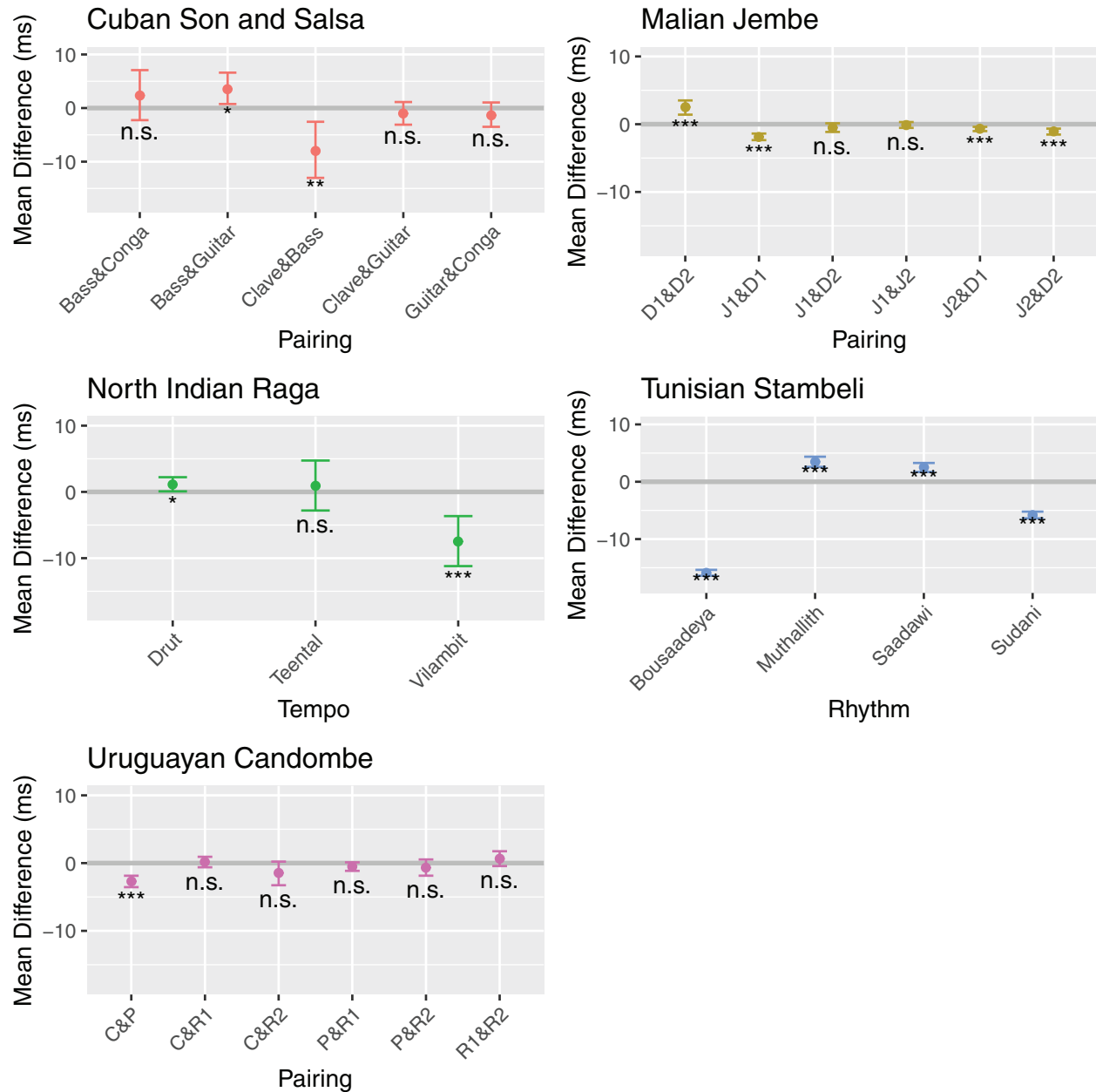


FIG. 7. Mean difference in ms ( $\pm$  95% CI) between beat 1 vs. other beats for each corpus (UC, MJ, and CSS, by pairing; NIR, by tempo; TS, by rhythm; for CSS, only the pairings in the nominal “rhythm section” are shown). Positive mean difference values denote tighter synchrony (smaller absolute asynchronies) on beat 1 than other beat positions.  $p < .05 = *$ ,  $p < .01 = **$ ,  $p < .001 = ***$ . UC: C = chico, P = piano, R = repique. MJ: D = dundun, J = jembe. NIR: “vilambit” = slow, “teental” refers to medium tempo, “drut” = fast).

- The clave and bongo are *more* precisely synchronized on beat one than the other beats in the CSS corpus (mean absolute asynchrony 12.3 ms vs. 18.2 ms).
- The gumbri and shqashiq are *less* precisely synchronized on the main downbeat than on the

weaker beats in the bousaadeya rhythm in the TS corpus (mean absolute asynchrony 47.6 ms vs. 31.7 ms).

The analysis of the difference in **Mean relative position** between strong and weak metrical positions also shows

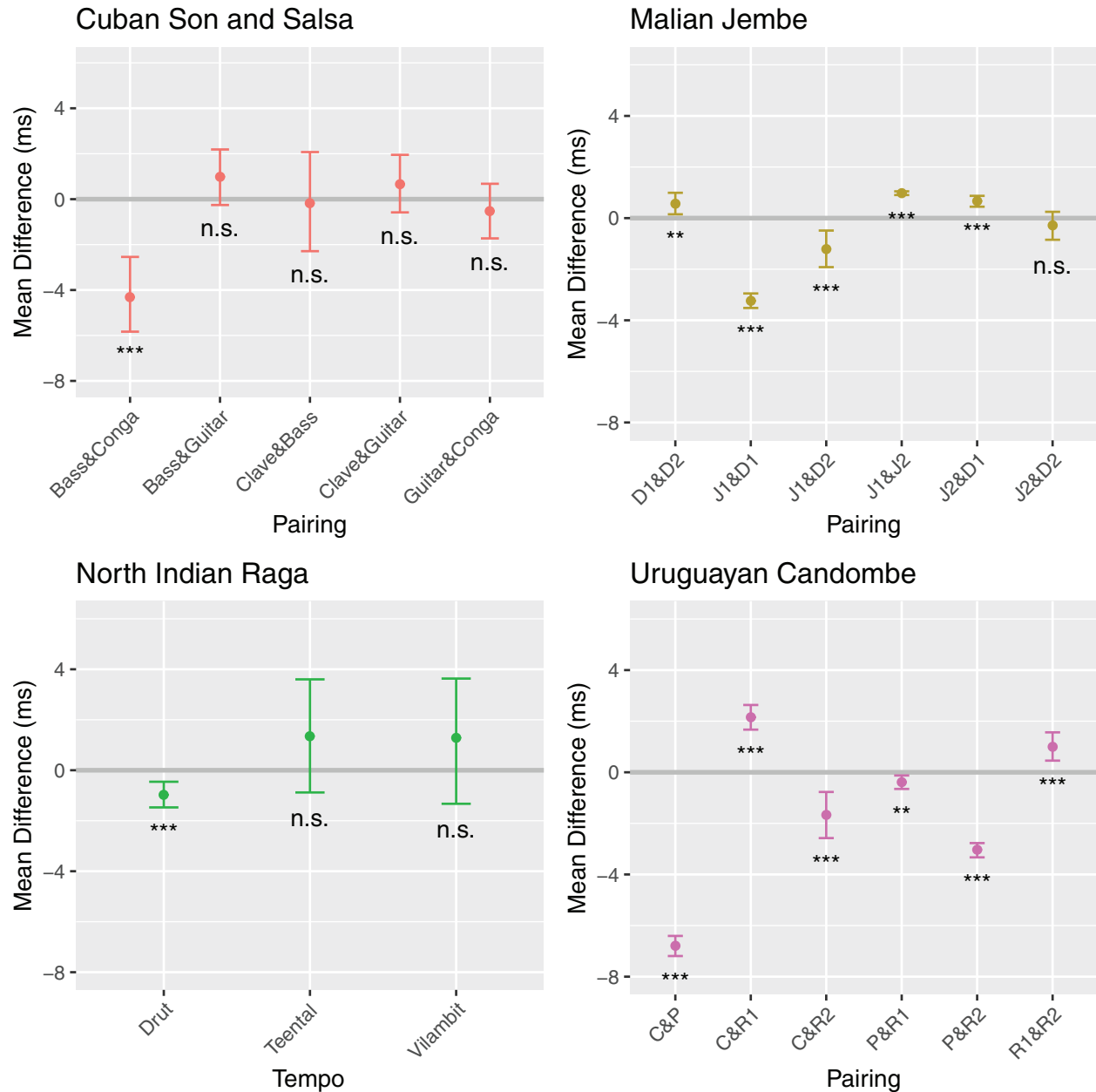


FIG. 8. Mean difference in ms ( $\pm$  95% CI) between beat vs. subdivision for each corpus and pairing. Key as Figure 7.

many examples of significant differences across the corpus. In most cases these differences are of just a few milliseconds, but a few examples stand out:

- Between gumbri lute and shqashiq cymbals in two of the four rhythmic patterns in TS: the gumbri is further ahead by 17.2 (sudani) and 18.8 (bousaa-deya) ms on the “strong” than the “weak” metrical position.

- The chico drum is 13.6 ms further ahead of the piano on the main beats than on their subdivisions, in UC.

These results are presented with the aim of both reporting the size of effects found in these corpora, and establishing where clear patterns seem to emerge cross-culturally and where not (in particular, that tighter synchrony on metrical downbeats does not seem to be



a general tendency). It is beyond the scope of this paper to attempt to interpret the significance of differences observed in each corpus, but see Clayton et al. (2019) for more detailed interpretations of the NIR results.

#### SUMMARY

A number of factors can be identified that contribute to variations in synchronization, which could therefore contribute to our model of IME. The main factors identified to this point are:

1. Instrument type. We have clearly seen that pairwise synchronization between drums or percussion instruments is tighter than that involving melodic instruments, while pairs including plucked string instruments are tighter than those involving bowed instruments (NIR and TS vs. ESQ). This progression is from short envelopes with clearly defined onsets to more sustained sounds with less defined onsets. As noted above, further work on perceptual onsets is necessary here; another factor that needs to be explored empirically is the time for which sounds are sustained (which is clearly much shorter for most drums than for bowed string tones).
2. Event density. There is an overall tendency for synchronization to be tighter at higher event densities. The picture is not straightforward when comparing across corpora as other factors, such as instrument type, have bigger effects, and a significant correlation is not found in all corpora. We found no relationship between event density and precision of synchronization in drum/percussion pairs.
3. “Melody lead.” Although there are some exceptions, instruments taking a lead role (which is often the main melody in the texture but can also apply to lead drum parts) tend to play a few milliseconds ahead of the group, on average. Melody lead may be more pronounced at higher event densities, as we observed for the TS corpus.
4. Metrical position. We do not observe a clear pattern of greater precision on strong metrical positions, as would be predicted by a hypothesis that musicians attend more to synchronizing with each other at particular points in the metrical cycle. Nonetheless we do observe numerous differences between precision and/or mean relative position according to metrical position, effects which depend on specific instrument pairs.

The first two factors point to the significance of acoustic factors on the synchronization mechanism. In brief, future studies could test the prediction that shorter

instrument sounds (shorter acoustic envelopes) increase synchronization precision, independent of the music’s cultural, genre-specific characteristics. The relationship between precision and greater event density (and tempo, where tempi can be compared fairly; see Clayton, Jakubowski & Eerola, 2019) is similarly independent of corpus when melodic sounds are involved. We did not observe this effect for drum-only pairings, however. Studies of very low-density percussion genres would help to explore this factor further. “Melody lead” has been explained in terms of the melody part being given more prominence by being performed slightly ahead of its accompaniment. The descriptive results in Figure 5 suggest that this phenomenon could be present cross-culturally, but by no means universally. Indeed, only certain kinds of music have a consistent lead-plus-accompaniment ensemble structure, so we would not expect to find evidence of “melody lead” everywhere; and even where this melody/accompaniment distinction pertains, there may be other factors determining who plays ahead in time. The diverse effects found of metrical position could be related to melody lead (or melody lag, since we find that lead instruments sometimes play significantly *behind* their accompaniments; the lead part may either push ahead or pull behind at stronger metrical positions), or to diverse local factors concerning the rhythmic patterns played by each instrument in relation to the metrical framework.

Beyond the specific factors that have been explored in this comparative analysis, future research is needed to investigate a number of other factors that may be linked to differences in synchronization. Individual variation is likely to play a role (see the section *Evolution and Development*), especially when we investigate less skilled performers than those studied in these examples, all of whom were selected as experts in their respective styles. In many styles increased event density and/or tempo occurs alongside increased dynamic levels, which suggests dynamics may also be investigated. Not least, any number of high-level musical factors may contribute to variability, many of which would be specific to particular kinds of music: for instance the difference between composed and improvised sections, when a particular instrument is given prominence (by taking a solo); relatively homophonic vs. contrapuntal textures (i.e., whether everyone plays the same rhythm or conflicting rhythms), whether a cadence (a formula signaling closure) is approaching, and so on. Detailed investigation of all these factors is beyond the scope of this paper.

Overall, these analyses do not support the hypothesis that strong metrical positions are more precisely synchronized than weak. This analysis of synchronization

between expert musicians contrasts with Patel et al.'s (2005) finding (see above). Clearly however, the task of tapping along with a stimulus is very different from the task of performing music with other people, since in the former case the stimulus does not adjust to the tapper, and this difference in task may explain the contrasting results. Even within corpora, in groups larger than two we find that some pairings are more precisely synchronized on strong positions at the same time as other pairings are less precise. We suggest, therefore, that variations in precision are difficult to explain as global effects of metrical structure per se, but rather may be due to differences in the specific rhythmic patterns being played at different parts of metrical cycles (to put it another way, rhythmic patterns tend not to be randomly distributed across the metrical framework). These differences are likely to result in many cases in small differences in the precision and/or mean relative positions of instruments by metrical position.

The factors we have investigated so far suggest, nonetheless, that synchronization between musicians could in principle be modeled using factors including onset or acoustic envelope characteristics, role (lead vs. accompanist or group member), event density, and metrical structure. Where any of these factors either change (e.g., the “lead” role switching between musicians) or continuously vary (e.g., event density), we would expect to find changes in synchronization over time: synchronization may get tighter or looser, or the relative positions of instruments change either continuously or suddenly. This could be explored systematically in future studies of musicians performing under more constrained conditions that allow some of these variables to be manipulated.

#### INTERPERSONAL COORDINATION IN MUSIC ENSEMBLES: CONTINUOUS DATA FROM ANCILLARY MOVEMENTS

*Overview.* In order to explore coordination empirically, we next present a case study of the temporal relationships between co-performers' ancillary movements, using video recordings of professional musicians from three corpora (MJ, NIR—of which a different selection of recordings is used from that in the section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis*—and the “Improvising Duos” corpus, comprising standard jazz and free improvisation). Due to the slower timescales over which ancillary movements evolve in comparison to sound-producing movements, it is possible to track these movements from standard video recordings (despite the lower sampling rate of video in comparison to motion capture systems, for instance; Jakubowski et al., 2017). This

allows us to make use of field recordings collected in diverse locations throughout the world. (Although a huge number of music recordings from around the world is freely available, for example via web sharing platforms, the methods presented here rely on static shots, which are relatively rare, as well as expert annotation; nonetheless, our approach expands the range of usable recordings considerably beyond the set of available motion capture data sets.) Our primary aims in this case study are 1) to demonstrate a relevant quantitative method for measuring coordination between co-performers (using automated tracking of ancillary movements and cross-wavelet transform (CWT) analysis), and 2) to examine how such movement coordination varies as a function of musical structure (i.e., at section boundaries).

#### METHOD

*Materials.* Three corpora of video recorded musical performances were used. These were chosen to partially coincide with the audio recorded performances utilized in the onset analysis reported in the section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis*, although it was not possible to use exactly the same materials, as some audio recordings did not have corresponding video recordings (or corresponding video recordings did not meet the required specifications for the automated motion tracking procedures, outlined below).

The first corpus was a subset of the MJ corpus described above, comprising three performances of the piece “Maraka” played by trios: jembe 1 (soloist), jembe 2 (accompanist who plays ostinato rhythms), and dundun (bass drummer who plays a repertoire-specific timeline pattern).<sup>10</sup> The same three performers played in all three recordings; however, for MJ\_Maraka\_3 the two jembe players switched roles (the soloist took on the accompanist role and vice versa). The mean duration of these video recordings was 160.7 s ( $SD = 26.7$ , range = 145 to 341 s).

The second video corpus, “Improvising Duos,” comprised 15 performances of free improvisation and 15 performances of the jazz standard *Autumn Leaves*, which have been used in previous research on aspects of visual interpersonal communication in music performance (Eerola et al., 2018; Moran et al., 2015; video corpus published as Moran, Jakubowski, & Keller, 2017). Within this corpus, five different duos performed the free improvisations (a style of music that

<sup>10</sup> Labelled MJ\_Maraka\_1, MJ\_Maraka\_2 and MJ\_Maraka\_3 in (Polak et al., 2018).

deliberately avoids a regular musical pulse) and six duos performed the standard jazz improvisations (*Autumn Leaves*, in a 4/4 meter with a regular pulse). These duos comprised 12 different instruments (e.g., saxophone, piano, double bass, drums, etc.) and the mean duration of the 30 performances was 157.0 s ( $SD = 55.7$ , range = 98.3 to 336.5 s).

The third corpus was a set of six North Indian classical music performances (three featuring a vocal soloist and three featuring an instrumental soloist) drawn from the same NIR corpus described in the previous section (Clayton et al., 2018); the selection of recordings used in this section is different from that in the synchronization analysis above, however.<sup>11</sup> As these performances are much longer in duration than the pieces in the other corpora, we focused on only the slow tempo sections. In each case only the soloist and tabla player was studied, to allow instrumental and vocal examples to be compared directly, although in some cases the ensembles were larger, including harmonium players (with vocal only) and/or one or more players of the accompanying plucked lute *tanpura*. The mean duration of these sections of the video recordings was 1711.1 s ( $SD = 758.2$ , range = 796.5 to 2901.9 s).

Although recorded in different settings using different equipment, all videos met the necessary criteria for the automated movement tracking procedures that we implemented. Specifically, the videos in all three corpora were recorded with fixed cameras (e.g., on a tripod) and a constant camera angle (e.g., no zooming or panning), with no substantial changes in lighting during the course of each performance. Performers were well separated in space, such that their movements did not occlude one another. All videos were recorded at a sampling rate of 25 Hz.

**Movement Data Extraction.** The movements of each performer were tracked using dense optical flow (OF) estimation in EyesWeb XMI 5.7.0.0 ([http://www.infomus.org/eyesweb\\_ita.php](http://www.infomus.org/eyesweb_ita.php)). OF is a computer vision technique that performs two-dimensional (x and y) movement tracking on video data by estimating the apparent velocities of objects. The implementation of OF used here is based on the algorithm of Farnebäck (2003) and has been validated for use in tracking ancillary movements from musical performers in Jakubowski et al. (2017). Movement tracking was implemented for

all performers in the “Improvising Duos” and MJ corpora, and for the soloist and the tabla player in the NIR corpus. First, in the video frame, a region of interest (ROI) was manually defined around the head and shoulders of each performer. Movement within each ROI was then automatically tracked using the OF algorithm (for a detailed description of the method see Jakubowski et al., 2017). The choice to track head and upper body movements is motivated by previous work, in which it was found that 97% of ancillary movements that were deemed as communicative by expert musicians fell into this category (e.g., head nods, body sway; Eerola et al., 2018). Of course, video recordings are two-dimensional and motion tracking using OF does not capture motion in the third dimension. It is possible therefore that if coordinated movements are primarily in this dimension then coordination could be underestimated using this method. Future studies could explore the use of multiple camera angles to attempt to mitigate this issue. Other options include retaining the two dimensions of movement separately, and or retaining phase information (see Eerola et al., 2018). Doing so in studies of individual corpora may prove productive, although to do so on a comparative basis is beyond our scope here.

**Quantifying Movement Coordination Between Co-performers.** A pairwise measure of co-performer coordination was calculated using cross-wavelet transform (CWT) analysis (See the data analysis section in *Measuring Entrainment in Musical Ensembles*). The aim of the present work was to examine how CWT Energy varied over time in relation to the musical structure, with a particular focus on section boundaries, in order to investigate the role of ancillary movements in coordinating musical transition points.

Before implementing the CWT analysis, the x- and y-coordinates of each performer from the OF data were detrended using linear regression and converted to polar coordinates, and the radial coordinates ( $\rho$ ) from each performer were retained for the subsequent analysis. The CWT analysis was applied across a broad frequency range from 0.3 to 2.0 Hz (in line with Eerola et al., 2018), in order to capture a wide range of co-occurring periodic movements. The first and last two seconds<sup>12</sup> of each performance were excluded, in order to avoid artefacts within the CWT analysis. The

<sup>11</sup> The analysis uses six pieces from the NIR corpus (Clayton et al., 2018). Three feature a vocalist: NIR\_VK\_Multani, NIR\_VS\_Bhoop and NIR\_SCh\_Malhar (khyal in vilambit ektal in each case). Three feature an instrumental soloist: NIR\_NGh\_Tabla (vilambit teental), NIR\_PrB\_Jhinjhoti (rupak tal) and NIR\_ABh\_Puriya (gat in vilambit teental).

<sup>12</sup> This was extended to five seconds for the Western Improvisation corpus, as this was an experiment in which Motion Capture data were also collected and performers were instructed to assume a “T-pose” for calibration purposes at the start and end of each video clip, which was not a part of the actual performance.

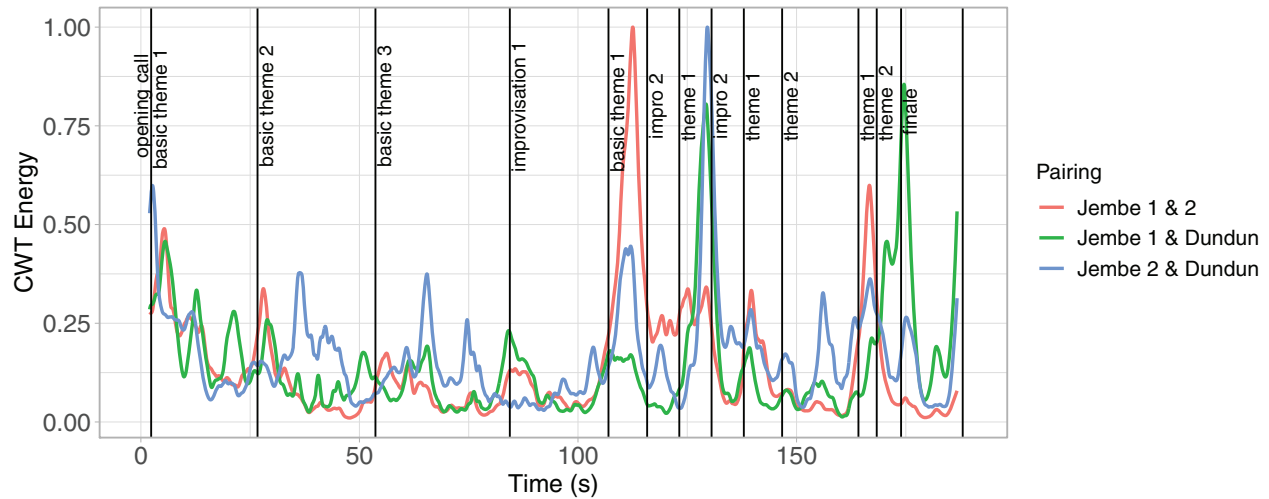


FIG. 9. Pairwise CWT Energy for the three performers in MJ\_Maraka\_1 from the MJ corpus plotted over time (onsets of music structural sections labeled via vertical lines).

resultant CWT Energy measure of pairwise movement coordination was normalized to a range of 0 to 1 for each performance.

*Annotations of Musical Structure.* For each of the three video corpora, meaningful boundary points within the musical structure were annotated by the expert researcher who had made the original video recordings. The MJ performances were divided into sections based on the theme that was being played, and also contained sections marked as improvisation. The “Improvising Duos” performances were demarcated into solo and joint sections, to represent transitions between sections where the performers were playing together vs. sections where one performer was soloing. The NIR music was labeled into terms of the repetitions of the tala (metrical) cycles, which are a key feature of this music (Clayton, 2000). All annotations of musical structure were performed in ELAN (Sloetjes & Wittenburg, 2008).

## RESULTS

*Describing Movement Coordination in Relation to Musical Structure.* Figure 9 shows one full *Maraka* performance (MJ\_Maraka\_1), selected at random from the MJ corpus for the purpose of visualising and describing our approach. The time series curves show the CWT Energy of each pair of performers over time, and the annotated music structural sections are labeled via vertical lines. This plot gives a descriptive indication that coordination of periodic ancillary movement increases at some points of structural transition (e.g., going into the second instance of improvisation 2, and the beginning of the finale section). However, this visualization

also suggests this relationship may vary depending on the pairing that is being examined. For instance, the jembe 2 and dundun player often move concurrently in a periodic fashion in the *middle* of structural sections (e.g., basic theme 2 and basic theme 3).

Another randomly selected example of a full performance is presented in Figure 10; this is NIR\_VK\_Multani from the NIR corpus. The CWT Energy curve shows pairwise movement coordination between the vocalist and tabla player, and in this case vertical lines represent the start of each metrical cycle. This plot shows that there are many clear instances where movement coordination between the two performers increases at the start of a metrical cycle.

*Testing Variation in Movement Coordination in Relation to Musical Structure.* We next sought to test whether CWT Energy was systematically different at musical section boundaries than non-boundaries for each piece in each corpus. We applied the same analysis technique for each piece as follows. First, the mean CWT Energy in a window centered around each section boundary was computed. We used a window size of 10% of the mean section length of the piece, to take account of the differing section lengths across different musical styles (mean window size in MJ = 1.43s, “Improvising Duos” = 2.49 s, NIR = 3.05 s). To select an equal quantity of CWT Energy data for comparison to the data within the section boundary windows, we shifted each section boundary to a randomly selected timepoint, with the conditions that: a) each boundary could move forward/backwards in time up to 50% of the window size before/after the adjacent section



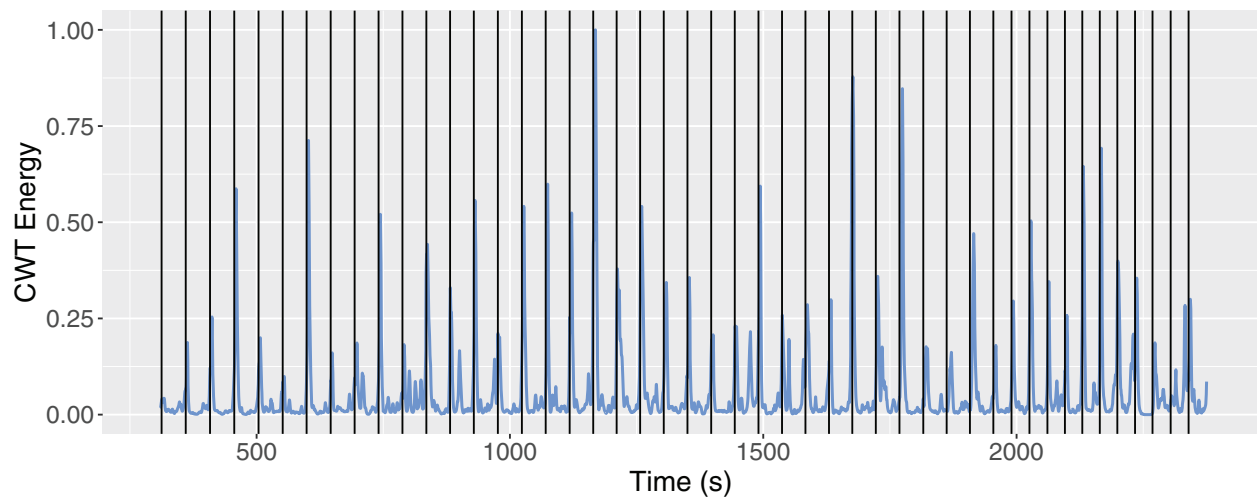


FIG. 10. Pairwise CWT Energy for the vocalist and tabla player in VK\_Multani, khyal in vilambit ektal, from the NIR corpus, plotted over time (onsets of metrical cycles labeled via vertical lines).

boundary, and b) each boundary had to move forward/backward in time at least the length of the window size. This preserves the number and relative spacing of section boundaries, whilst ensuring that “non-boundary” sections will not contain any of the same datapoints as the section boundary windows. We then compared the mean CWT Energy around boundaries vs. non-boundaries, using the same window size for each. The process was repeated 1,000 times for each piece, with a new round of random sampling of non-boundaries each time. We calculated the pairwise mean difference between CWT Energy in each section boundary window and its corresponding non-boundary for each iteration and computed an overall mean difference and 95% confidence interval (CI) across all iterations of the analysis for each piece in each corpus. Figures 11 through 13 show the results of this analysis for each of the three corpora separately. Full results can be found in Supplementary Table 2 (online at mp.ucpress.edu).

As seen in Figure 11, in the MJ corpus, movement coordination (as measured by mean CWT Energy) was higher at section boundaries than non-boundaries for most performer pairings across the three pieces, with an overall mean difference in CWT Energy (computed as mean CWT Energy at boundaries minus non-boundaries) across the corpus of 0.04. However, this difference between the mean CWT Energy at boundaries minus non-boundaries was not statistically significant across the corpus as a whole (95% CI: [-0.05, 0.12]). There were some variations related to performer pairing, as the mean difference in CWT Energy between boundaries/non-boundaries was lower for the

jembe 1 and 2 pairing ( $M_{\text{diff}} = -0.005$ ) than the other two pairings (jembe 1 & dundun:  $M_{\text{diff}} = 0.065$ , jembe 2 & dundun:  $M_{\text{diff}} = 0.061$ ). This could indicate that the two jembe players rely more on other cues to coordinate (e.g., eye contact, auditory cues), or rely on peripheral vision of one another’s hands due to sitting next to each other, rather than ancillary movement cues at points of transition.

Figure 12 shows the results for the “Improvising Duos” corpus. This corpus shows much diversity, with some clear examples of increased movement coordination at section boundaries (e.g., Piece 25), but other pieces showing no clear tendency—or even the opposite pattern (Piece 12). This large degree of variability across pieces may relate to the 22 different performers implicated in this corpus, who might each have their own personal style for communicating points of musical transition. Across the corpus as a whole there was a small overall difference in means in the direction of increased movement coordination at section boundaries as compared to non-boundaries, but this difference was not statistically significant ( $M_{\text{diff}} = 0.03$ , 95% CI: [-0.07, 0.11]). At the individual piece level, there was a significant increase in movement coordination at section boundaries for five pieces (Piece 1, 4, 18, 25, and 27), or three pieces following Bonferroni correction (Piece 1, 25, and 27). In addition, the difference in movement coordination at section boundaries versus non-boundaries was somewhat more pronounced for jazz standards ( $M_{\text{diff}} = 0.04$ ) as compared to free improvisations ( $M_{\text{diff}} = 0.02$ ). This does not necessarily indicate that the free improvisers are not coordinating via

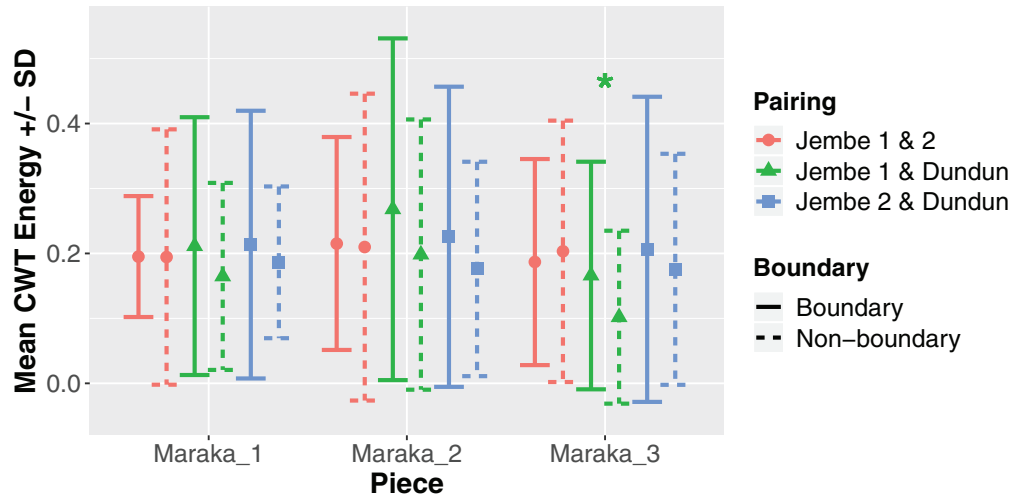


FIG. 11. Mean CWT Energy (+/- standard deviation) within windows around boundaries vs. non-boundaries by piece and instrument pairing (MJ corpus, Maraka\_1, Maraka\_2 and Maraka\_3).  $p < .05 = *$ . Note that no pairwise comparisons here exceeded a Bonferroni-corrected threshold for nine tests of  $p < .0056$ .

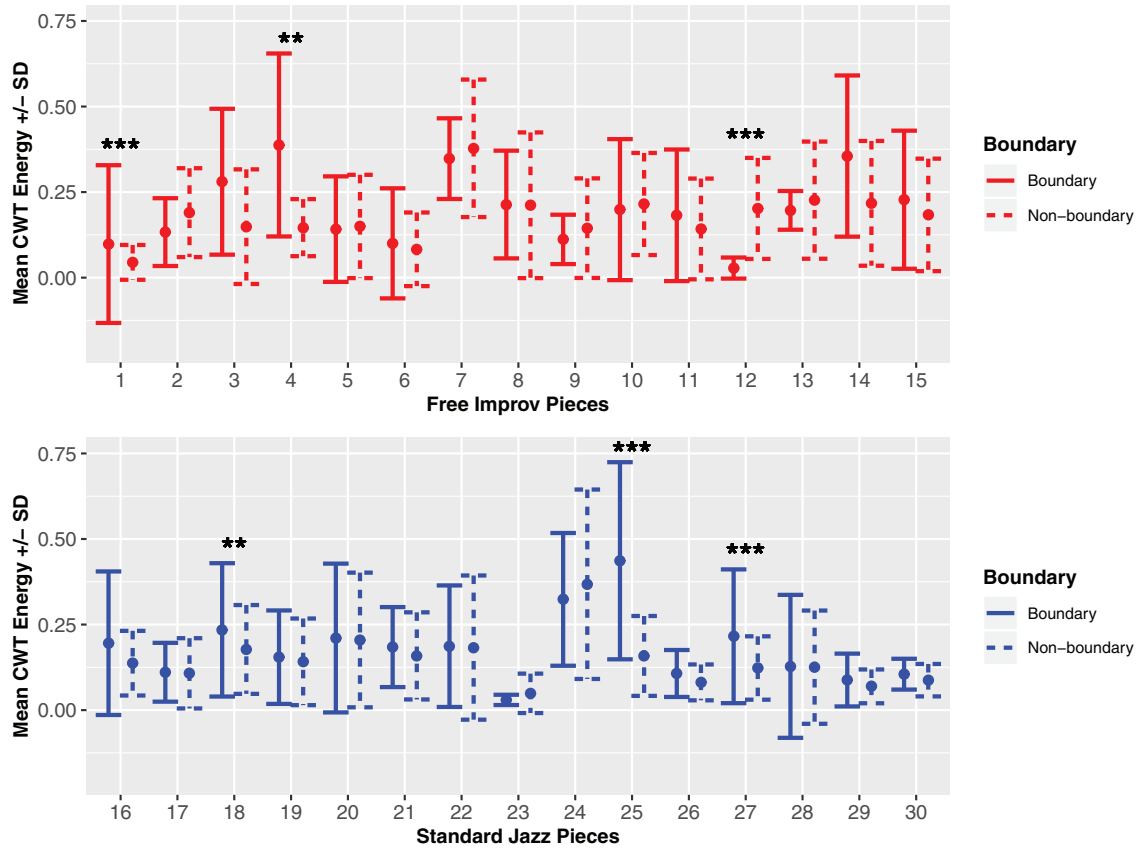


FIG. 12. Mean CWT Energy (+/- standard deviation) within windows around boundaries vs. non-boundaries by improvisation style (Improvising Duos corpus).  $p < .01 = **$ ,  $p < .001 = ***$ . Note that only the comparisons marked \*\*\* exceed a Bonferroni-corrected threshold for 15 tests of  $p < .0033$ .

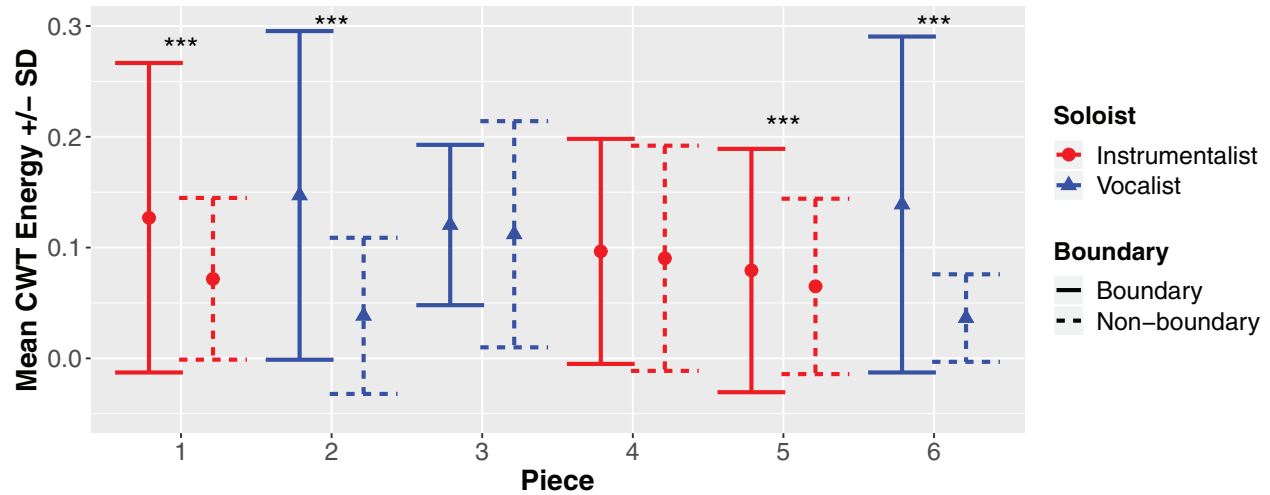


FIG. 13. Mean CWT Energy (+/- standard deviation) within windows around boundaries vs. non-boundaries by piece and soloist type (NIR corpus: 1 = ABh\_Puriya, vilambit teental (sitar); 2 = VK\_Multani, vilambit ektal (vocal); 3 = VS\_Bhoop, vilambit ektal (vocal); 4 = PrB\_Jhinjhoti, rupak tal (sarod); 5 = NGh\_Tabla, vilambit teental (tabla solo); 6 = Sch\_Malhar, vilambit ektal (vocal)).  $p < .001 = ***$ . Piece 2 is the same as that illustrated in Figure 10. Note that all significant comparisons marked \*\*\* exceed a Bonferroni-corrected threshold for six tests of  $p < .0083$ .

movement cues at music structural boundaries; one possibility could be that the unplanned and unpredictable nature of this music means that the performers actually need to coordinate *more often* than at these larger structural boundaries (e.g., subsections) in order to produce a coherent performance. This is consistent with Schögl's analysis (1999) of jazz duets, which found coordination to be higher in the seconds immediately before "points of change." On this basis we might expect that in music with more frequent transitions, we also find more frequent moments of higher coordination. Thus, movement coordination may be more similar at boundaries versus non-boundaries in the free improvisation than in the standard jazz, whereas in standard jazz a shared sense of the musical pulse can facilitate coordination between structural boundary points.

Finally, the results for the NIR corpus are shown in Figure 13. Four of the individual pieces in this corpus showed a significant difference in terms of mean CWT Energy being higher at section boundaries than non-boundaries (Pieces 1, 2, 5, and 6), with a significant mean difference across the corpus as a whole ( $M_{\text{diff}} = 0.05$ , 95% CI: [0.03, 0.06]). The colors in the figure show the pieces that feature an instrumental versus vocal soloist, which indicate that there is a more pronounced difference in CWT Energy at boundaries for the vocal ( $M_{\text{diff}} = 0.07$ ) than the instrumental performances ( $M_{\text{diff}} = 0.03$ ). As demonstrated in Clayton et al. (2019), this effect is largely apparent where the metrical boundary coincides with a cadential boundary in the

instrumental pieces; the larger effect in the vocal pieces may be due to the fact that a higher proportion of metrical boundaries coincide with cadences in the vocal than the instrumental pieces.

In sum, we have demonstrated one of many possible approaches for quantifying and testing variations in co-performer movement coordination as a function of musical structure. More detailed study of ancillary movement coordination could include, for example, expert annotation of other points of change than section boundaries, analysis of gaze patterns, or performer interviews. Such additional elements could allow exploration of different types of transitions; for example, do tempo changes differ from changes in texture, harmony or other (genre-specific) features? Analysis of gaze patterns and performer interviews could help to elucidate the extent to which musicians are conscious of increased coordination (or increased negotiation or information exchange) at these points. More detailed analysis could also explore the temporal structure of coordination in more detail, for example the point at which greater movement coherence begins in relation to the occurrence of a transition, or the use of any explicit cues by musicians. It is beyond the scope of this paper to go into this level of detail on a comparative basis.

#### SUMMARY

In this section we outlined some of the wide variety of data types available for studying IME, both discrete and continuous, and how to apply different methods in

practice. Our analyses exemplified the two data types, while also focusing on different modalities and timescales. This should not however be taken as implying that, for example, only acoustic information is relevant at short timescales, or only continuous movement data is available for longer-term processes.

In the section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis* we analyzed synchronization using onset timing data from a diverse selection of six musical genres. The Group and Pairwise asynchronization figures suggest a spread from relatively precise synchrony for the Afrogenic drum ensembles (MJ, UC) to lower precision in melodic genres such as the ESQ, some Segments of NIR and TS, and some pairings in CSS. Precision of synchronization in our corpora is correlated with event density for examples including a melody instrument, but not drum/percussion only pairings. We found many examples of significant effects of metrical position on precision, but these tend to cancel each other out when the group as a whole is considered, with some pairings more precise and others less so on theoretically strong metrical positions. We thus found little support for the idea that synchrony is universally more precise on strong metrical positions. Mean relative positions (mean time differences between one instrument and the average of the group) tend to be less than 5 ms in most cases, with a few outliers: the gumbri plays significantly ahead in TS, for example, and plays further ahead at higher event densities. Metrical position again has an effect on mean relative position in various cases, but there is no clear overall tendency.

In the section *Interpersonal Coordination in Music Ensembles: Continuous Data From Ancillary Movements* we explored the coordination of ancillary movement between pairs of performers in relation to musical structure in three contrasting corpora. These three corpora contrast musically in almost every respect: nonetheless, each corpus contained examples of pieces where movement coordination increased at section boundaries. However, it is clear that this pattern varies notably across styles and performers. The NIR corpus showed the clearest support for the idea that movement coordination increases at section boundaries, with the MJ corpus providing limited evidence in this regard. This may be due to the fact that in NIR, eye contact and the expression of pleasure at cadences is a common feature of performance practice, whereas this is not the case in MJ. These results also suggest other factors such as genre, instrument, ensemble size, and spatial positioning of the ensemble, may affect these results in, as of yet, unexplored ways. More analysis needs to be carried out

on this question, including exploring the role of different rates of movement (i.e., isolating different frequency bands in the CWT data, rather than taking account of all movements across a broad frequency range).

What does this tell us about the cultural variability, or lack thereof, in IME? The fact that the precision of synchronization between drum/percussion pairs is similar in the different examples, and that melody/drum pairs become more precise at higher densities while drum/percussion pairs do not, suggests that to this extent synchrony varies with physical factors such as the envelope type of the typical instrument sounds or the speed at which people play. These factors can be described as “cultural” in the broad sense outlined in the section *The Role of Social and Cultural Factors in a Model of IME*, and we have also seen how recent synchronization models incorporate a role for top-down influence of internal representations, that may include culturally specific representations of aspects of musical structure. However, although we cannot rule it out on the basis of our analysis, we have seen no evidence that these factors are cultural *choices* (e.g., in the sense that group A plays more precisely because they value the feeling of being tightly synchronized more than group B). We may hypothesize that for a given musical structure, taking into account the type of sounds produced, the complexity of their interrelationship, speed of performance etc, a performance ceiling exists for precision of synchronization: for example, event onset perception is limited by the shape of envelope onsets and it is impossible to be more precise than this variability allows. Then we might ask, what scope is there for groups to deliberately vary the precision of synchronization? Must each individual perform at their optimal level, or can they play more “loosely” for aesthetic reasons? We have not collected evidence that could answer this question. If there exists a musical genre in which performers consistently perform suboptimally in terms of synchronization precision by choice, it would be necessary to ask them to record the same music with greater precision in order to demonstrate this. Anecdotally, musicians in many different traditions are aware of the possibility of playing more mechanically or more regularly than they in fact choose to do. Are they similarly able to turn synchronization precision up and down? That is less certain.

We are clearly observing a common phenomenon (SMS) that employs a very similar cognitive and motor architecture cross-culturally, albeit one that can be modified by learning. The common tendency to move in a more coordinated fashion at structural boundaries may be a side-effect of musicians paying each other



greater attention at these points, just as experimental research has found individual movements become more coordinated in the presence of visual attention (Kawase, 2014). At the same time, all of these same factors do evidence significant cultural variability. Musical groups may be more or less tightly synchronized than others: this is linked to physical factors, but people can make *choices* as to what instruments to play and at what speeds. Cultural factors influencing IME, in this sense, must include choice of instrumental sounds and articulation, speed of playing, complexity of musical texture and musical transitions, and group leadership and hierarchy. We may all share a tendency to move together when coordinating activity, but some genres such as NIR clearly make this an explicit part of the performance practice by increasing mutual attention and eye contact at specific moments (this is both recognized by performers and observed empirically; Clayton & Leante, 2015; Moran, 2013), while this is less obviously the case in some other musical cultures and styles. Considerable variety is therefore built on shared processes.

Perhaps the most significant observation to come out of this analysis is not the quantitative findings, but our reflection on the process of exploring synchronization. In order to understand how precisely musicians synchronize we need to specify the metrical structure that they all relate to. In some cases, such as NIR, assigning event onsets to metrical positions is a time-consuming and often subjective task that is not possible without making informed decisions as to what kind of pattern a musician is intending to play. The same process in MJ music may be less subjective, but here we have to build into the model the fact that in some pieces subdivision IOIs are non-isochronous. In both of these contrasting examples, in order to understand how musicians synchronize we have to build into our model an aspect of their shared knowledge, whether the peculiarity of a metrical structure or experience of the kinds of rhythmic variations musicians are likely to play in a given musical style. If this knowledge is essential in order to *measure* synchronization, we also propose that it is necessary for the musicians in order to synchronize effectively. This dependency of synchronization—let alone coordination of structural transitions—on detailed culturally shared knowledge needs to be incorporated in our model of IME.

### Modeling IME

We concluded our review section by proposing that while existing models offer a great deal to studies of IME, a more comprehensive picture should be possible

by combining different levels of explanation to include both a neurophysiological mechanism for synchronization, a broader understanding of the different components of IME, and a model of the relationships between entrainment, social, and cultural factors. Further, we have suggested the need to consider a wider variety of the *musical* factors that allow individuals to coordinate and anticipate the progress of musical performances. The latter, we suggested, should include more explicit acknowledgement of culturally shared knowledge representations, which would allow the larger model to address the ways in which IME is culturally mediated as well as socially effective.

The analyses in the third section of this paper (*Measuring Entrainment in Musical Ensembles*), while exemplifying only two of many possible approaches, helped to establish the principle that IME can be analyzed at different levels in the same example, using both discrete and continuous data, and that these analyses can be compared cross-culturally. They help to demonstrate the fact that IME is dependent on culturally specific and shared knowledge at different levels, from internalizing timing patterns and hierarchies up to anticipating the kinds of rhythmic variation that are likely to occur in a given style. The analytical results demonstrate that while similar mechanisms can be observed in action in highly contrasting examples, many of the choices that people make, from the size of groups and the choice of instruments to the assignment of leadership roles, have a quantifiable effect on IME. What these analyses also show is that to a significant extent cultural variability is not simply—or even necessarily at all—a question of aesthetic choice, but may be determined by the behavior of the underlying systems: for example, if cultures who use drum ensembles rather than bowed lute ensembles (e.g., string quartets) are more precisely synchronized, this difference may be explainable in terms of the acoustic properties of the sounds produced. (This does not imply a neurological difference between cultures but may reflect the influence of different acoustic signals on a common system.) This is cultural in the sense that the choice of instrument type is down a particular group of people passing instrument designs and playing techniques down the generations; it nonetheless depends on cognitive and motor systems operating in ways that may be less culturally variable. Bearing all of this in mind, in the following section we present an expanded model of IME.

### AN IME MODEL

Figure 14 illustrates complementary aspects of the conceptual model we propose of interpersonal musical

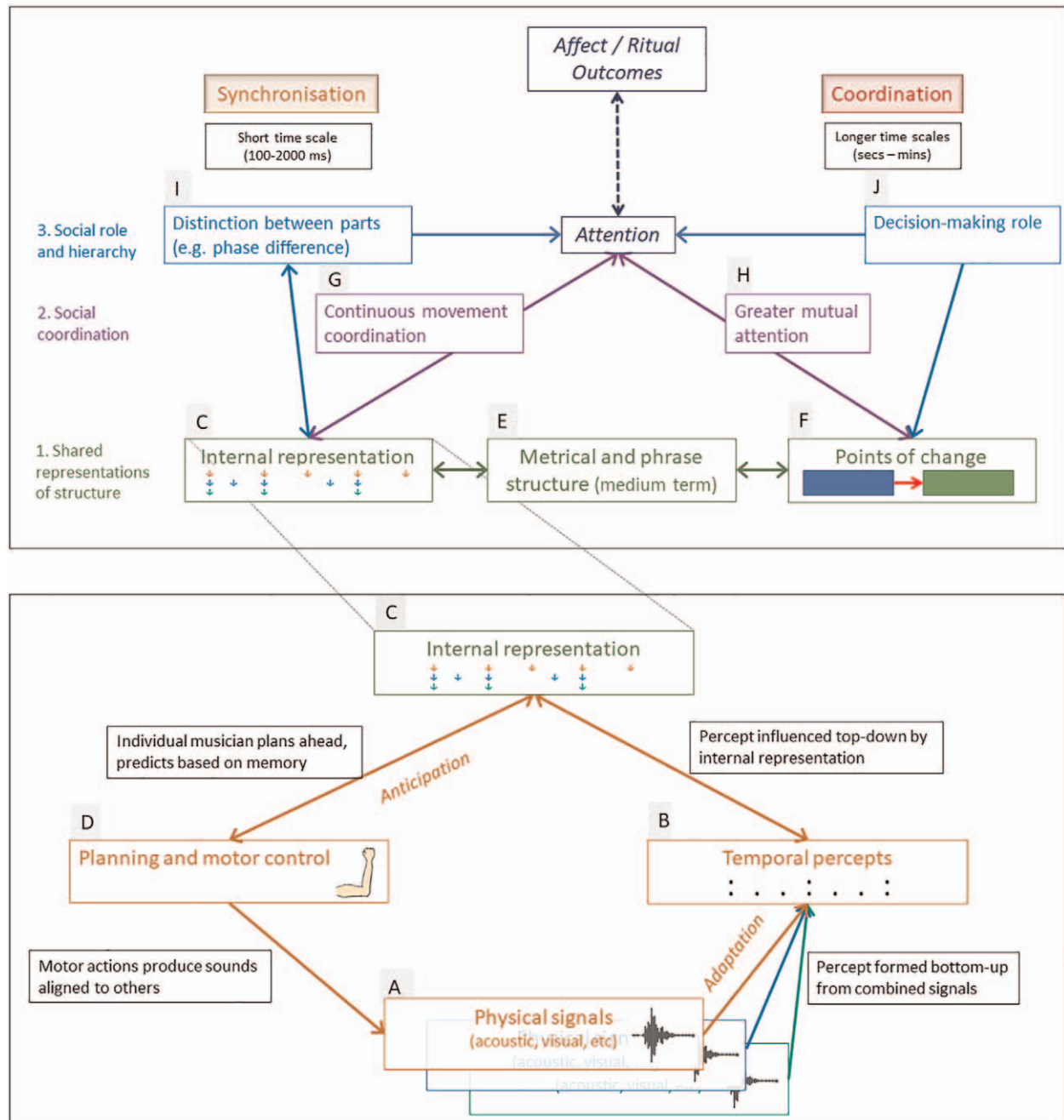


FIG. 14. Model of Interpersonal Musical Entrainment that accounts for synchronization at short timescales and coordination at longer timescales. The lower pane illustrates the synchronization part of the model (boxes A-D), while the upper pane depicts elements related to shared representations of musical structure (C-F), social coordination (G & H), and social role and hierarchy (I & J).

entrainment (IME). This model focuses on the act of performance; individuals who can listen and entrain to the music but not influence it in any way (e.g., audience members) are not represented here, and it is not intended as a model of a listener responding to

recorded music (for approaches that encompass these scenarios, see Leman, 2007, 2016; Trost et al., 2017). It assumes current understanding of the neurophysiological processes involved in musical entrainment while adding both a clear role for longer-term processes and

structures and consideration of which aspects of IME may be culturally determined, and which may be influenced by or have an effect on social interaction. The synchronization part of the model is illustrated in brown boxes (A-D, lower pane); the points of change illustrated in Box F fall under “coordination,” with Box E (metrical and phrase structure) relating to both components, extending the synchronized short-term patterns to cover longer time-spans. Some key factors of social coordination (purple, Boxes G and H) and social role and hierarchy (blue, Boxes I and J) are also illustrated.

*Short-term Psychological Factors: Synchronization.* The lower panel (boxes A-D) illustrates the short-term processes involved in SMS. The synchronization process is understood here to be founded on the embodied structures and processes described in existing theories, including both bottom-up and top-down processes. The relationship between physical signal, temporal percepts (including meter), internal representation of the music’s short-term structure and planning and motor control is drawn schematically in the figure’s four main boxes. Physical signals (box A)—auditory and in other modalities—can give rise to a perception of temporal structure, such as meter (bottom-up processing), but this percept (box B) can also be influenced by existing representations of the music’s short-term structure (box C, top-down processing). The structural representation is also updated on the basis of what is perceived. Referring to the represented structure the individual musician can anticipate motor actions, which are planned in such a way that the resulting sounds align temporally with those predicted for the rest of the group (box D); in fact, it is the combined physical signals produced by all participants that inform perception (although a musician may be more influenced by some parts than others). The basic synchronization process described as “adaptation” in the ADAM model may be regarded as a loop between boxes A and B; influence from box C adds the “anticipation” part of that model. Most entrainment models effectively include boxes A, B, and D in some way—acoustic signals, their perception and motor production; models including a top-down element such as ADAM, rich BPS, and other dynamic models including Active Sensing, include all four of these elements.

The physical signals comprise the only evidence available to the musicians that they are coordinating appropriately, allowing them to judge prediction errors and adjust. Participants may infer from this that they all share the same representations and perceptions of the music (although in fact their internal representations may not need to be identical in order for them to be entrained appropriately).

What is clear from this part of our model, and the previous models on which it draws, is that low-level processes such as error correction and neural resonance interact with higher-level processes involving the representation of musical structures and the planning of performance. What needs to be investigated is how higher-level processes interact with the low-level neural processing; and also, how these representations are learned and shared within communities (their cultural dimension). Both the acquisition of such representations in development and music training, and their historical development, are important topics related to IME.

Our synchronization analyses in the section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis* explored IME by measuring the timing of auditory events in relation to a known metrical structure. This analysis showed that precision tends to increase with higher event density and be higher for percussion instruments than for melodic instruments (perhaps because of their shorter sound envelopes). The fact that these patterns are observed across different corpora may suggest that these factors arise from constraints at the neurophysiological level without significant input from culturally shared knowledge representations: more temporally dense input with clearly defined onset times leads to more precise synchronization. If this proves to be the case, these factors can be regarded as aspects of bottom-up processing. We did not observe this relationship with drum pairs, however: more work needs to be done on the interaction between acoustic envelope and event density in order to explain this finding. Another limitation of the analysis in the section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis* is that the possible differences between physical and perceptual onsets are not taken into account. Whether the lower precision of bowed instruments simply reflects a margin of error in the perceptual estimation of onset times, or other factors are involved, remains to be explored. For instance, if the temporal difference between onset and p-center for a given instrument is fairly constant (recent estimates suggest that differences are generally less than 40 ms with variability rarely over  $SD \sim 20$  ms, see Danielsen et al., 2019; London et al., 2019), then a fixed asynchrony difference when comparing slow-onset and fast-onset instruments would be expected, with the slow-onset instrument perceived to sound relatively late.

#### SOCIAL AND CULTURAL FACTORS: SYNCHRONIZATION AND COORDINATION

Another aspect of variability in the synchronization analysis concerns mean relative position—in other

terms, phase differences between the different musicians' parts. Although the picture is complicated, in many cases where the musical structure involves a clear leader, such as a melodic soloist accompanied by a percussion part, the "lead" part tends to play ahead in time. (This was also observed with the "lead" drum in the MJ corpus, and Clayton et al. (2019) have shown a similar result in terms of the switch in mean relative position in some of the Indian performances as the lead moves between melody part and tabla). This is one way in which socio-musical factors are represented in our model: if an individual performs a leadership role, which may be inherent in a particular part of the musical texture, this may affect the synchronization by driving a (usually small) phase difference between individual parts (see Varni et al., 2010). This factor is represented in the upper part of Figure 14, which illustrates processes taking place over longer time spans than SMS: it focuses on social and cultural aspects of IME, that tend to be left out of existing entrainment models.

Level 1 focuses on shared representations of structure. Internal representations of short-term (e.g., metrical) structure guide top-down perception (Box C); such representations are learned in the context of social interactions and culturally shared. Longer-term temporal structures are also learned and represented (E). This may include long metrical cycles (such as many Indian talas), or groups of short metrical cycles (such as 4- or 8-cycle patterns), which extend the patterns that can be inferred in a bottom-up process. If the music contains points of change (F), for example of texture, meter, or tempo, these must be coordinated between the group; this process may include socially agreed cues or reference to an external representation such as a score, or may be influenced by other factors such as audience response.

"Social coordination" (Level 2) represents ways in which the coordination between individuals is reflected in a performance. For example, a regular meter or groove may be accompanied by continuous coordinated movement such as swaying or foot-tapping (G). Points of change may be accompanied by greater mutual attention (H) and an increase in coordinated movement, as shown in the section *Interpersonal Coordination in Music Ensembles: Continuous Data From Ancillary Movements*, and may also be signaled by single cueing gestures.

Level 3, "Social role and hierarchy" represents ways in which the distinctions of role or the hierarchy between individuals may be reflected in IME. For example, (I) if one musician plays ahead of the others to make their part stand out, this results in a mean phase difference between the parts. An individual may also take on the

role of deciding when a transition should be effected and/or how (J). All of these factors may be regarded as cultural to the extent that they are shared between a specific group of people, however defined; they are social to the extent that they are activated in the course of interaction between individuals, especially if they are shaped by or contributing to social distinctions or social affiliations. Areas represented in this part of the model also explicitly point up to higher-level social and ritual effects of IME, drawing on Collins' Interaction Ritual Chains model. Areas represented in this part of the diagram are as follows:

Gross body movements and visual information become more salient at longer time spans (to the right), while acoustic information offers greater precision in note-to-note synchronization. Coordination clearly often involves, however, in addition to coordinated movement, the exchange of symbolic information (auditory and visual cues, i.e., aspects of the physical signals with culturally agreed meanings). We have hypothesized that coordinated body movement is linked to increased mutual visual attention, which can occur both deliberately and without any conscious intention. The tendency to focus more attention on co-performers at points of transition seems to be common across cultures, although since the nature of musical structures and the transitions they involve differ greatly, the frequency and dynamics of these moments may still differ cross-culturally. We also noted, however, many contrasts between the expression of this tendency between genres. This part of the model also tentatively links the social dimensions of IME, via attention, to the "ritual outcomes" suggested by Collins' Interaction Ritual Chains model, which also involves the mediation of affective responses. More work is required to flesh out the interconnections between these phenomena, however.

#### IME AND CULTURALLY SHARED KNOWLEDGE

These discussions suggest that culturally shared knowledge, learned through exposure to and training in specific kinds of music, may be influential in IME at several levels:

- In modeling the relationship between acoustic information and metrical structure (e.g., understanding that a bass drum sound will emphasize beat 1 in much Western popular music).
- In modeling specific temporal structures (e.g., if a pattern of one longer and two shorter beats is to be recognized, an internal representation of this pattern, or at least a general representation of non-isochronous beat durations, is necessary).



- In extending metrical structures to create patterns over longer time spans.
- In the progression of performances, including choices that may be made, who may make them and how they may be signaled.
- In the management of transitions, including how changes of meter, tempo, or texture are affected.
- In understanding the relationship of internal musical processes to the wider context (e.g., that certain kinds of music should inspire listeners to dance, to cry, or to sit silently, and that there are appropriate musical responses to these eventualities).

It is not possible at this point to specify exactly how many or which mechanisms are involved in IME. Specifying how and where this knowledge is encoded is also beyond the scope of this paper. Nonetheless it is clear that all of this musical knowledge cannot be reduced to a single form of learning or representation: these mechanisms cover a wide range of timescales, and they differ in the extent to which the individual learns them explicitly or implicitly, or whether one is conscious that they are being affirmed or contradicted. Picking up the beat in a piece of music can be so fast and instinctive that one does not realize that a specific piece of knowledge, a representation acquired through learning, is being deployed; on the other hand, if a musician becomes irritated that a co-performer is trying to lead a transition that she considers inappropriate to the style or occasion, that musician is much more likely to be aware that expert knowledge is being deployed and that she has a choice as to how to respond to the violation of expectations.

The last item on our list involves knowledge of the relationships between the performance and the wider context that affords it meaning. In discussing this we return to Collins' interaction ritual chains theory to broaden the discussion of the sociocultural significance of IME (Figure 1). Collins' "ritual ingredients" include group assembly (the co-presence of performers that allows them to mutually entrain); barriers to outsiders (which may be related to the cultural specificity of knowledge, or to a refusal to entrain with outgroups as in Lucas et al. 2011); and a mutual focus of attention (the activities of making music, associated activities such as dance and other ritual actions) which is linked to shared mood (entrainment has been linked to affective entrainment, see section *Prosocial Behavior & Affect*). According to Collins these elements help to generate Durkheim's collective effervescence, an overflowing of positive affect that motivates the repetition of ritual and generates, amongst other things, group solidarity.

Some aspects of Collins' model help in understanding IME: for example, the longitudinal perspective, whether considered over evolutionary or developmental time, as patterns of entrainment are repeated. Collins suggests a fundamental role for affect in IME's affordance of social effects: this is an idea that has been considered in psychology too and deserves to be explored further, as does Launay et al.'s (2016) observation of the link between IME and the release of specific neurohormones. There nonetheless remain aspects that need to be clarified if affect is to be properly integrated into our musical model. In Collins' diagram, the different "ingredients" are combined to generate collective effervescence, but how do the different elements *function*? Is mutual attention between participants more or less important than joint attention to an agreed goal? How important is the recognition of an out-group? How does entrainment generate emotion or affect, and how does Collins' feedback loop in which entrainment reinforces a "transient emotional stimulus" function? Some of the implications of Collins' model remain to be tested.

Do patterns of IME reflect patterns of social organization? The "ingredients" combine to generate the outcomes, from group solidarity to symbols of social relationship and standards of morality, but Collins' model does not offer a specific mechanism by which particular ways in which groups assemble and share attention lead to particular kinds of solidarity or standards of morality. Partly this is down to a simple omission of a higher-level feedback loop: groups assemble and interact having already established some form of group solidarity and identity (for example, attending a ritual of a specific religious group or a musical event featuring a known artist)—consistent with Collins's perspective but omitted from the diagram—so any such relationship would have to emerge through the operation of this feedback loop over an extended time period. To what extent the specifics of his "ritual outcomes" really emerge from the entrainment process, and how significant this is in relation to the many other forces influencing social institutions and practices, is impossible to quantify. The question of leadership, buried in Collins' "group assembly" category but addressed in the Goffmanian aspect of his model, is instructive here. Leadership is often an important aspect of group interaction that, as we have seen, influences IME. Yet leadership is itself complex. Even clearly defined "leaders" can be contested: a drummer may resist the instruction to speed up in a way he objects to, for example. In this way, we would expect IME not simply to embody a notional group hierarchy, but to reflect micro-social issues around leadership and its contestation. In this

way, the dynamics of a musical ensemble need not passively reflect a given social structure but may embody also the tensions inherent in that structure.

Another way that the processes embodied in our model may impact socially is simply through the convergence they expect of their participants and listeners. A metrical structure, for instance, allows musicians to coordinate and also provides a temporal structure to which listeners may entrain (and perhaps move). The shared expectation that a dance tune will be repeated a certain number of times is useful, even essential, in coordinating some kinds of social dance: the musical structure built, we suggest, on top of a neurophysiological mechanism, this facilitates a broader form of joint action (see review section *Background and Operational Definitions*) and therefore has a social utility. As Dueck (2013) points out, however, music's metrical structure recruits hearers to interact in particular ways; examples of music that appear to defy expectations of regular metrical structure may sometimes be viewed as an embodiment of resistance. Who is in charge, then, is a broader issue than "which individual is the leader?" but implies a wider issue of how groups of people coordinate their actions in pursuit of a shared goal, who controls that process and how this control may be resisted.

We have seen an increasing body of literature addressing the group-bonding and prosocial behavior benefits of IME (see review sections *Background and Operational Definitions* and *Social and Cultural Dimensions of Musical Entrainment*). We suggest here that these effects may be just one set of symptoms of a much bigger issue around social temporal coordination: the mutual entrainment evident in group music-making is surely felt to be felicitous in many, perhaps most of its manifestations, and its effects on people's perceptions of each other can be measured experimentally. And yet, it is equally true that many musical performances involve a delicate balance between group bonding and individual expression, as individuals find ways to bind together in pursuit of shared goals while asserting their own agency within the group (Keller, Novembre, & Loehr 2016). The kind of group ritual experiences described by Durkheim as "collective effervescence" may involve a temporary loss of this self-consciousness and awareness of social distinction (i.e., *communitas*), but this is far from being the way that all music is experienced. It may also be, as Mogan et al. (2017) suggest, that this "effervescent" mechanism becomes more salient when large group sizes are involved. This, in turn, would suggest that more attention be given to the sort of large group interactions that take place both in large ensembles (orchestras, choirs), but also between audience members.

#### PROSPECTS FOR RESEARCH IN IME

Different aspects of the IME model presented here highlight the need for further research into IME at a number of different levels. The following suggestions all refer back to the proposals in this paper, and are not intended to be exhaustive.

- **Neural.** While we have not sought to advance current models of neural entrainment, our synchronization analysis suggests possible future research, including on the effect of different aspects of the auditory signal (e.g., envelope shape) on neural resonance or error-correction, and interaction between envelope shape and event density.
- **Learning.** How are temporal hierarchies and patterns that moderate IME learned? Via statistical learning or induced by multiple levels of periodicity? Can we follow the learning process and track its effects on synchronization dynamics? Can large-scale processes and structures be designed for experimental processes as a way of testing learning? What kinds of patterns are easy/difficult for different populations to learn? Is there a critical period (e.g., in early childhood) during which it is easiest to learn metrical/rhythmic patterns?
- **Autonomic functions.** What is the impact of IME on functions including respiratory and cardiac rhythms?
- **Perceptual factors.** How are different aspects of IME perceived by listeners? Do cultural differences impact the ability to discriminate different aspects of IME, and are differences in preferences for different patterns of IME related to cultural difference and shared knowledge? Are these distinctions linked to aesthetic appreciation (via affect? via vestibular activity?)
- **Sociomusical dynamics,** including the exercise of leadership, differentiation, definition of out-groups, etc. Can these processes be better understood by bringing together observation, feedback from performers, and empirical analysis of performance?
- **Affective.** What is the significance of affect in motivating IME and its social effects, and the perceptual and physiological correlates of IME-related affect? To what extent is the sharing of positive affect a function of group size?
- **Individual differences.** How do less-skilled or less-experienced individuals perform, and what are the effect of personality traits on IME processes?

- Leadership. How does this relate to (a) perceptions of which part is most important or prominent, (b) perceptions of which individual has the most authority?
- Information and coupling. Can the role of different modalities be better understood through experimental intervention in musical performance, for example by restricting mutual information (e.g., poor visibility, poor audibility)?
- Coordination constraints, including movement constraints (body size, instrument mobility, etc.). What effect do coordination errors, as opposed to synchronization errors, have on synchronization, or on musical experience?
- Interaction between SMS and coordination mechanisms. To what extent can each mechanism impact the other?

This amounts to a considerable variety of research approaches, involving different data types and different analytical approaches. In many of these areas, moreover, what is required is the coordination of input from different disciplinary positions: topics such as explicit representation of temporal structures and performance processes, or the rich topic of leadership, require ethnographic and observational methods as well as empirical analysis of timing or physiological and brain imaging approaches. IME research cannot reach its potential if people's motivations, values and knowledge are not considered alongside the functioning of their neuron populations, or if musical knowledge, however learned and however articulated (or not) is treated as a minor detail.

#### CONCLUSIONS

We have explored some key elements of current understanding of IME, including models of different aspects of this rich and multi-layered phenomenon from the neurological to the sociological, from Neural resonance theory, ADAM, and rich BPS through to Interaction ritual chains. We have highlighted the need to include representations of diverse culturally shared knowledge alongside other, less culturally mediated neural processes; for consideration of different timescales, their different modalities and mechanisms and the relationships between them, and advocated cross-cultural comparative analysis as an essential part of the process of developing such a conceptual model. We therefore drew a distinction between synchronization and coordination and their different functions and timescales, and presented examples of comparative analysis drawing on a diverse and richly annotated set

of corpora from India, Mali, Uruguay, Cuba, Tunisia, the UK, and Germany. These analyses also allowed us to demonstrate some of the contrasts between analysis of discrete and continuous data, and between audio and video data. Our synchronization analysis (section *Interpersonal Synchronization in Music Ensembles: Onset-based Comparative Analysis*) highlighted acoustic factors which impact on IME, as well as the importance of culturally shared knowledge in establishing a framework within which synchronization can be conceptualized, performed, and measured. Our coordination analysis (section *Interpersonal Coordination in Music Ensembles: Continuous Data From Ancillary Movements*) pointed to the fact that, for all their extraordinary diversity, very different genres seem to show a tendency for movements to become more coordinated at structural boundaries in the music, albeit to varying degrees that appear to relate to the style, pairing, and performer/instrument implicated.

Building on these analytical findings and the earlier theoretical discussion, in this final section we have presented a new model of IME. This model builds explicitly on several earlier models, while integrating a greater role for culturally shared knowledge and learning. Aspects of the latter, which can give rise to cultural variability, are highlighted at the same time as common neurophysiological mechanisms are acknowledged. The role of affective entrainment in potentially mediating the social effects of IME is included, tentatively placing this factor with the social aspects of IM. This perspective is expanded with reference to Collins' Interaction ritual chains theory, which alongside the importance of affect also builds in an appreciation of the longitudinal aspect of IME, encompassing not just repetition and learning but also the affective motivation for that repetition. Finally, we have outlined a handful of areas that the development of this model suggests as important areas for future research. Some of these strands of research are already in progress. Some are not so well developed, especially those concerning longer timescales, the integration of cognitive and social perspectives, or the role of knowledge representations in entrainment. The theoretical discussion and empirical results presented in this paper are intended to address the disjunction between different perspectives on IME, and to contribute to better understanding of the relationships between neurophysiological processes underpinning entrainment, its psychological importance for the individual, and its sociological significance, as well as to enhance understanding the diverse ways in which IME manifests in different musical cultures.

### Author Note

This work was supported by the Arts and Humanities Research Council [grant number AH/N00308X/1].

The following individuals kindly shared their data (audiovisual files, annotations, and in some cases extracted onset data) and responded to queries about its interpretation. This paper would not have been possible without their contribution: Richard Jankowsky (Tunisian stambeli), Nikki Moran (Improvising duos),

Rainer Polak and Nori Jacoby (Malian jembe), Adrian Poole (Cuban son and salsa), Martín Rocamora and Luis Jure (Uruguayan candombe).

Thanks also to Simone Tarsitani for his work preparing and editing media.

*Correspondence concerning this article should be addressed to* Martin Clayton, Durham University, Music Department, Palace Green, Durham DH1 3RL, United Kingdom. E-mail: martin.clayton@durham.ac.uk

### References

- ALÉN, O. (1995). Rhythm as duration of sounds in Tumba Francesca. *Ethnomusicology*, 39, 55–71. DOI: 10.2307/852200
- ARRIGHI, R., ALAIS, D., & BURR, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision*, 6, 260–268. DOI: 10.1167/6.3.6
- AYMOZ, C., & VIVIANI, P. (2004). Perceptual asynchronies for biological and non-biological visual events. *Vision Research*, 44, 1547–1563. DOI: 10.1016/j.visres.2003.12.011
- BABILONI, C., BUFFO, P., VECCHIO, F., MARZANO, N., DEL PERCIO, C., SPADA, D., ET AL. (2012). Brains “in concert”: Frontal oscillatory alpha rhythms and empathy in professional musicians. *NeuroImage*, 60, 105–116. DOI: 10.1016/j.neuroimage.2011.12.008
- BAUER, A. K. R., JAEGER, M., THORNE, J. D., BENDIXEN, A., & DEBENER, S. (2015). The auditory dynamic attending theory revisited: A closer look at the pitch comparison task. *Brain Research*, 1626, 198–210. DOI: 10.1016/j.brainres.2015.04.032
- BERNARDI, L., PORTA, C., CASUCCI, G., BALSAMO, R., BERNARDI, N. F., FOGARI, R., & SLEIGHT, P. (2009). Dynamic interactions between musical, cardiovascular, and cerebral rhythms in humans. *Circulation*, 119, 3171–3180. DOI: 10.1161/CIRCULATIONAHA.108.806174
- BISHOP, L., & GOEBL, W. (2017). Communication for coordination: Gesture kinematics and conventionality affect synchronization success in piano duos. *Psychological Research*, 82, 1177–1194. DOI: 10.1007/s00426-017-0893-3
- BISPHAM, J. (2006). Rhythm in music: What is it? Who has it? And why? *Music Perception*, 24, 125–134. DOI: 10.1525/mp.2006.24.2.125
- BLACKING, J. (1977). Towards an anthropology of the body. In J. Blacking (Ed.), *The anthropology of the body* (pp. 1–28). London, UK: Academic Press.
- BLACKING, J. (1981). Ethnography of musical performance, In J. Blacking (Ed.), *Report of the twelfth [IMS] congress Berkeley 1977* (pp. 383–401). Kassel, Germany: Bärenreiter Kassel.
- BOOTSMA, R. J., & VAN WIERINGEN, P. C. (1990). Timing an attacking forehand drive in table tennis. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 21–29. DOI: 10.1037/0096-1523.16.1.21
- BREGMAN, M. R., IVERSEN, J. R., LICHMAN, D., REINHART, M., & PATEL, A. D. (2013). A method for testing synchronization to a musical beat in domestic horses (*Equus ferus caballus*). *Empirical Musicology Review*, 7, 144–156. DOI: 10.18061/emr.v7i3-4.3745
- BROCHARD, R., ABECASIS, D., POTTER, D., RAGOT, R., & DRAKE, C. (2003). The “ticktock” of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, 14, 362–366. DOI: 10.1111/1467-9280.24441
- BROOKS, K., & BROOKS, K. (2010). Enhancing sports performance through the use of music. *Journal of Exercise Physiology Online*, 13, 52–57.
- BROWN, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception and Psychophysics*, 59, 1118–1140. DOI: 10.3758/bf03205526
- BUHUSI, C. V., & MECK, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, 6, 755–765. DOI: 10.1038/nrn1764
- BUTTERFIELD, M. (2010). Participatory discrepancies and the perception of beats in jazz. *Music Perception*, 27, 157–176. DOI: 10.1525/mp.2010.27.3.157
- CHANG, A., LIVINGSTONE, S. R., BOSNYAK, D. J., & TRAINOR, L. J. (2017). Body sway reflects leadership in joint music performance. *Proceedings of the National Academy of Sciences*, 114, E4134–E4141. DOI: 10.1073/pnas.1617657114
- CHEN, J. L., PENHUNE, V. B., & ZATORRE, R. J. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage*, 32, 1771–1781. DOI: 10.1016/j.neuroimage.2006.04.207
- CIRELLI, L. K., EINARSON, K., & TRAINOR, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants. *Developmental Science*, 17, 1003–1011. DOI: 10.1111/desc.12193



- CIRELLI, L. K., SPINELLI, C., NOZARADAN, S., & TRAINOR, L. J. (2016). Measuring neural entrainment to beat and meter in infants: Effects of music background. *Frontiers in Neuroscience*, 10, 229. DOI: 10.3389/fnins.2016.00229
- CLARK, H. H. (1996). *Using language*. Cambridge: Cambridge University Press
- CLAYTON, M. (1996). Free rhythm: Ethnomusicology and the study of music without meter. *Bulletin of the School of Oriental and African Studies*, 59, 323–332. DOI: 10.1017/S0041977X00031608
- CLAYTON, M. (2000). *Time in Indian music: Rhythm, meter and form in North Indian rag performance*. Oxford, UK: Oxford University Press.
- CLAYTON, M. (2007a). Time, gesture and attention in a khyal performance. *Asian Music*, 38(2), 71–96.
- CLAYTON, M. (2007b). Observing entrainment in music performance: Video-based observational analysis of Indian musicians' tanpura playing and beat marking. *Musicae Scientiae*, 11, 27–59. DOI: 10.1177/102986490701100102
- CLAYTON, M., JAKUBOWSKI, K., & EEROLA, T. (2019). Interpersonal entrainment in Indian instrumental music performance: Synchronization and movement coordination relate to tempo, dynamics, metrical and cadential structure. *Musicae Scientiae* 23, 304–331. DOI: 10.1177/1029864919844809
- CLAYTON, M., & LEANTE, L. (2015). Role, status and hierarchy in the performance of North Indian classical music. *Ethnomusicology Forum*, 24, 414–442. DOI: 10.1080/17411912.2015.1091272
- CLAYTON, M., LEANTE, L., & TARSITANI, S. (2018). *IEMP North Indian Raga* (Open Science Framework). DOI: 10.17605/OSF.IO/KS325
- CLAYTON, M., SAGER, R., & WILL, U. (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. *European Meetings in Ethnomusicology 11* (ESEM Counterpoint 1), 1–82.
- CLAYTON, M., & TARSITANI, S. (2019, May 30). *IEMP String Quartet* (Open Science Framework). DOI: 10.17605/OSF.IO/USFX3.
- CLAYTON, M., TARSITANI, S., JANKOWSKY, R., JURE, L., LEANTE, L., POLAK, R., POOLE, A., ET AL. (in press). The Interpersonal Entrainment in Music Performance data collection. *Empirical Musicology Review*.
- COHEN, E. E., EJSMOND-FREY, R., KNIGHT, N., & DUNBAR, R. I. (2010). Rowers' high: Behavioural synchrony is correlated with elevated pain thresholds. *Biology Letters*, 6, 106–108. DOI: 10.1098/rsbl.2009.0670
- COLLEY, I. D., KELLER, P. E., & HALPERN, A. R. (2018). Working memory and auditory imagery predict sensorimotor synchronization with expressively timed music. *The Quarterly Journal of Experimental Psychology*, 71, 1781–1796. DOI: 10.1080/17470218.2017.1366531
- COLLINS, R. (2004). *Interaction ritual chains*. Princeton, NJ: Princeton University Press.
- CONDON, W., & OGSTON, W. (1971). Speech and body motion synchrony of the speaker-hearer. In D. Horton & J. Jenkins (Eds.), *The perception of language* (pp.150–184). Columbus, OH: Charles E. Merrill.
- CONDON, W. S. (1985). Sound-film microanalysis: A means for correlating brain and behavior. In F. H. Duffy & N. Geschwind (Eds.), *Dyslexia: A neuroscientific approach to clinical evaluation*. Boston, MA: Little, Brown and Company.
- COOK, P., ROUSE, A., WILSON, M., & REICHMUTH, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, 127, 412–427. DOI: 10.1037/a0032345
- DAHL, S., & FRIBERG, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, 24, 433–454. DOI: 10.1525/mp.2007.24.5.433
- DANIELSEN, A., NYMOEN, K., ANDERSON, E., CÂMARA, G. S., LANGERØD, M. T., THOMPSON, M. R., & LONDON, J. (2019). Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 45, 402–418. DOI: 10.1037/xhp0000611
- D'AUSILIO, A., BADINO, L., LI, Y., TOKAY, S., CRAIGHERO, L., CANTO, R., ALOIMONOS, Y. & FADIGA, L. (2012). Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS ONE*, 7, e35757. DOI: 10.1371/journal.pone.0035757
- D'AUSILIO, A., NOVEMBRE, G., FADIGA, L., & KELLER, P.E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*, 19, 111–114. DOI: 10.1016/j.tics.2015.01.005
- DAVIDSON, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, 21, 103–113. DOI: 10.1177/030573569302100201
- DEAN, R. T., & DUNSMUIR, W. T. (2016). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior Research Methods*, 48, 783–802. DOI: 10.3758/s13428-015-0611-2
- DEMOS, A. P., CHAFFIN, R., BEGOSH, K. T., DANIELS, J. R., & MARSH, K. L. (2012). Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology: General*, 141, 49–53. DOI: 10.1037/a0023843
- DISERENS, C. M. (1926). *The influence of music on behavior*. Princeton, NJ: Princeton University Press.

- DOELLING, K. B., ASSANE, M. F., BEVILACQUA, D., PESARAN, B., & POEPEL, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, 116, 10113–10121. DOI: 10.1073/pnas.1816414116
- DOELLING, K. B., & POEPEL, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences*, 112, E6233–E6242. DOI: 10.1073/pnas.1508431112
- DOFFMAN, M. (2013). Groove: Temporality, awareness and the feeling of entrainment in jazz performance. In M. Clayton, B. Dueck, & L. Leante (Eds.), *Experience and meaning in music performance* (pp. 62–85). Oxford, UK: Oxford University Press.
- DRAKE, C., JONES, M. R., & BARUCH, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, 77, 251–288. DOI: 10.1016/S0010-0277(00)00106-2
- DUECK, B. (2013). Rhythm and role recruitment in Manitoban aboriginal music. In M. Clayton, B. Dueck, & L. Leante (Eds.), *Experience and meaning in music performance* (pp. 135–160). Oxford, UK: Oxford University Press.
- DUMAS, G., MARTINERIE, J., SOUSSIGNAN, R., & NADEL, J. (2012). Does the brain know who is at the origin of what in an imitative interaction? *Frontiers in Human Neuroscience*, 6, 128. DOI: 10.3389/fnhum.2012.00128
- DURKHEIM, E. (1995). *The elementary forms of religious life* (K. E. Fields, Trans.). New York: The Free Press. (Original work published 1912)
- EEROLA, T., JAKUBOWSKI, K., MORAN, N., KELLER, P. E., & CLAYTON, M. (2018). Shared periodic performer movements coordinate interactions in duo improvisations. *Royal Society Open Science* 5, 171520. DOI: 10.1098/rsos.171520
- EEROLA, T., LUCK, G., & TOIVIAINEN, P. (2006). An investigation of pre-schoolers' corporeal synchronization with music. In *Proceedings of the 9th International Conference on Music Perception & Cognition, Bologna* (pp. 472–476). Bologna, Italy: ICMPC and ESCOM.
- ELLIS, R. J., & THAYER, J. F. (2010). Music and autonomic nervous system (dys)function. *Music Perception*, 27, 317–326. DOI: 10.1525/mp.2010.27.4.317
- ENGEL, A., HOEFLE, S., MONTEIRO, M. C., BRAMATI, J. E., LIMA, D. O., KELLER, P. E., & MOLL, J. (2014). Feeling the groove: Perceiving asynchronies in drumming sounds typical for carnival parades of Brazilian Samba Schools. Paper presented at the *The Neurosciences and Music - V*. Dijon, France.
- ENGEL, A., HOEFLE, S., MONTEIRO, M. C., MOLL, J., & KELLER, P. E. (2014). The influence of loudness and synchrony on pleasure during listening to the drumming section of a Brazilian Samba School. Paper presented at the *Simpósio de Cognição e Artes Musicais*. Campinas, Brazil.
- ETZEL, J. A., JOHNSEN, E. L., DICKERSON, J., TRANEL, D., & ADOLPHS, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *International Journal of Psychophysiology*, 61, 57–69. DOI: 10.1016/j.ijpsycho.2005.10.025
- FAIRHURST, M. T., JANATA, P., & KELLER, P. E. (2014). Leading the follower: An fMRI investigation of dynamic cooperativity and leader-follower strategies in synchronization with an adaptive virtual partner. *NeuroImage*, 84, 688–697. DOI: 10.1016/j.neuroimage.2013.09.027
- FARNEBÄCK, G. (2003). Two-frame motion estimation based on polynomial expansion. In J. Bigun & T. Gustavsson (Eds.), *Image analysis. SCIA 2003. Lecture Notes in Computer Science* (Vol. 2749). Berlin, Heidelberg: Springer. DOI: 10.1007/3-540-45103-X\_50
- FISHER, N. I. (1993). *Statistical analysis of circular data*. Cambridge, UK: Cambridge University Press.
- FISCHER, R., CALLANDER, R., REDDISH, P., & BULBULIA, J. (2013). How do rituals affect cooperation? *Human Nature*, 24, 115–125. DOI: 10.1007/s12110-013-9167-y
- FITZPATRICK, P., FRAZIER, J., COCHRAN, D., MITCHELL, T. V., COLEMAN, C., & SCHMIDT, R. (2016). Impairments of social motor synchrony evident in autism spectrum disorder. *Frontiers in Psychology*, 7, 1323. DOI: 10.3389/fpsyg.2016.01323
- FRIBERG, A., & SUNDBERG, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America*, 98, 2524–2531. DOI: 10.1121/1.413218
- FRIBERG, A., & SUNDBERG, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19, 333–349. DOI: 10.1525/mp.2002.19.3.333
- FUJIOKA, T., TRAINOR, L. J., LARGE, E. W., & ROSS, B. (2009). Beta and gamma rhythms in human auditory cortex during musical beat processing. *Annals of the New York Academy of Sciences*, 1169, 89–92. DOI: 10.1111/j.1749-6632.2009.04779.x
- FUJIOKA, T., TRAINOR, L. J., LARGE, E. W., & ROSS, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *Journal of Neuroscience*, 32, 1791–1802. DOI: 10.1523/JNEUROSCI.4107-11.2012
- GARCÍA-GARIBAY, O., CADENA-VALENCIA, J., MERCHANT, H., & DE LAFUENTE, V. (2016). Monkeys share the human ability to internally maintain a temporal rhythm. *Frontiers in Psychology*, 7, 1971. DOI: 10.3389/fpsyg.2016.01971
- GIBBON, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325. DOI: 10.1037/0033-295X.84.3.279
- GIBBON, J., & CHURCH, R. M. (1990). Representation of time. *Cognition*, 37, 23–54. DOI: 10.1016/0010-0277(90)90017-E

- GINSBORG, J., & KING, E. C. (2009). Gestures and glances – The effects of familiarity and expertise on singers' and pianists' bodily movements in ensemble rehearsals. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P.-S. Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music* (pp. 159–164). Jyväskylä, Finland: University of Jyväskylä.
- GLOWINSKI, D., BADINO, L., AUSILIO, A., CAMURRI, A., & FADIGA, L. (2012). Analysis of leadership in a string quartet. In *Third International Workshop on Social Behaviour in Music at ACM ICMI 2012* (pp. 763–774). Santa Monica, CA: ACM ICMI.
- GLOWINSKI, D., MANCINI, M., COWIE, R., CAMURRI, A., CHIORRI, C., & DOHERTY, C. (2013). The movements made by performers in a skilled quartet: A distinctive pattern, and the function that it serves. *Frontiers in Psychology*, 4, 841. DOI: 10.3389/fpsyg.2013.00841
- GOEBL, W., & PALMER, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, 26, 427–438. DOI: 10.1525/mp.2009.26.5.427
- GOEBL, W., & PARNCUTT, R. (2001). Perception of onset asynchronies: Acoustic piano versus synthesized complex versus pure tones. Paper presented at the *Meeting of the Society for Music Perception and Cognition*. Kingston, Canada.
- GOFFMAN, E. (1959). *The presentation of self in everyday life*. Garden City, NJ: Doubleday.
- GOMEZ, P., & DANUSER, B. (2007). Relationships between musical structure and psychophysiological measures of emotion. *Emotion*, 7, 377–387. DOI: 10.1037/1528-3542.7.2.377
- GOOD, A., CHOMA, B., & RUSSO, F. A. (2017). Movement synchrony influences intergroup relations in a minimal groups paradigm. *Basic and Applied Social Psychology*, 39, 231–238. DOI: 10.1080/01973533.2017.1337015
- GRINSTED, A., MOORE, J. C., & JEVREJEVA, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, 11, 561–566. DOI: 10.5194/npg-11-561-2004
- GUCKENHEIMER, J., & HOLMES, P. (1983). *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. New York: Springer-Verlag. DOI: 10.1007/978-1-4612-1140-2
- HAAS, F., DISTENFELD, S., & AXEN, K. (1986). Effects of perceived musical rhythm on respiratory pattern. *Journal of Applied Physiology*, 61, 1185–1191. DOI: 10.1152/jappl.1986.61.3.1185
- HALWANI, G. F., LOUI, P., RUBER, T., & SCHLAUG, G. (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, 2, 156. DOI: 10.3389/fpsyg.2011.00156
- HASEGAWA, A., OKANOYA, K., HASEGAWA, T., & SEKI, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific Reports*, 1, 120. DOI: 10.1038/srep00120
- HATTORI, Y., TOMONAGA, M., & MATSUZAWA, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Scientific Reports*, 3, srep01566. DOI: 10.1038/srep01566
- HEIDER, A., & WARNER, R. S. (2010). Bodies in sync: Interaction ritual theory applied to sacred harp singing. *Sociology of Religion*, 71, 76–97. DOI: 10.1093/socrel/srq001
- HIRSH, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 31, 759–767. DOI: 10.1121/1.1907782
- HIRSCH, M. W., SMALE, S., & DEVANEY, R. L. (2004). *Differential equations, dynamical systems, and an introduction to chaos*. San Diego, CA: Academic Press.
- HOLCOMBE, A. O. (2009). Seeing slow and seeing fast: Two limits on perception. *Trends in Cognitive Sciences*, 13, 216–221. DOI: 10.1016/j.tics.2009.02.005
- HOVE, M. J., FAIRHURST, M. T., KOTZ, S. A., & KELLER, P. E. (2013). Synchronizing with auditory and visual rhythms: An fMRI assessment of modality differences and modality appropriateness. *NeuroImage*, 67, 313–321. DOI: 10.1016/j.neuroimage.2012.11.032
- HOVE, M., KELLER, P., & KRUMHANS, C. (2007). Sensorimotor synchronization with chords containing tone-onset asynchronies. *Perception and Psychophysics*, 69, 699–708. DOI: 10.3758/BF03193772
- HOVE, M. J., & RISEN, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, 27, 949–960. DOI: 10.1521/soco.2009.27.6.949
- HURLEY, B. K., MARTENS, P. A., & JANATA, P. (2014). Spontaneous sensorimotor coupling with multipart music. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1679–1696. DOI: 10.1037/a0037154
- ISSARTEL, J., BARDAINNE, T., GAILLOT, P., & MARIN, L. (2015). The relevance of the cross-wavelet transform in the analysis of human interaction—a tutorial. *Frontiers in Psychology*, 5, 1566. DOI: 10.3389/fpsyg.2014.01566
- IVERSEN, J., & BALASUBRAMANIAM, R. (2016). Synchronization and temporal processing. *Current Opinion in Behavioral Science*, 8. DOI: 10.1016/j.cobeha.2016.02.027
- IVERSEN, J. R., REPP, B. H., & PATEL, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences*, 1169, 58–73. DOI: 10.1111/j.1749-6632.2009.04579.x
- JACOBY, N., KELLER, P. E., REPP, B. H., AHISSAR, M., & TISHBY, N. (2015). Lower bound on the accuracy of parameter estimation methods for linear sensorimotor synchronization models. *Timing and Time Perception*, 3, 32–51. DOI: 10.1163/22134468-00002047
- JACOBY, N., & McDERMOTT, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, 27, 359–370. DOI: 10.1016/j.cub.2016.12.031

- JACOBY, N., TISHBY, N., REPP, B. H., AHISSAR, M., & KELLER, P. E. (2015). Parameter estimation of linear sensorimotor synchronization models: Phase correction, period correction, and ensemble synchronization. *Timing and Time Perception*, 3, 52–87. DOI: 10.1163/22134468-00002048
- JACOBY, N., MARGULIS, E. H., CLAYTON, M., HANNON, E., HONING, H., ET AL. (2020). Cross-cultural work in music cognition. Challenges, insights, and recommendations. *Music Perception*, 37, 185–195. DOI: 10.1525/mp.2020.37.3.185
- JAKUBOWSKI, K., EEROLA, T., ALBORNO, P., VOLPE, G., CAMURRI, A., & CLAYTON, M. (2017). Extracting coarse body movements from video in music performance: A comparison of automated computer vision techniques with motion capture data. *Frontiers in Digital Humanities, section Digital Musicology*, 4, 9. DOI: 10.3389/fdigh.2017.00009
- JANKOWSKY, R., TARSITANI, S., & CLAYTON, M. (2019, July 9). *IEMP Tunisian Stambeli* (Open Science Framework). DOI: 10.17605/OSF.IO/SWBY6
- JONES, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83(5), 323–355. DOI: 10.1037/0033-295X.83.5.323
- JONES, M. R. (2019). *Time will tell. A theory of dynamic attending*. New York: Oxford University Press.
- JONES, M. R., & BOLTZ, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96, 459–491. DOI: 10.1037/0033-295X.96.3.459
- JONES, M. R., MOYNIHAN, H., MACKENZIE, N., & PUENTE, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, 13, 313–319. DOI: 10.1111/1467-9280.00458
- JURE, L., ROCAMORA, M., TARSITANI, S., & CLAYTON, M. (2019, May 20). *IEMP Uruguayan Candombe* (Open Science Framework). DOI: 10.17605/OSF.IO/WFX7K
- JUSLIN, P. (2013). From everyday emotions to aesthetic emotions: Toward a unified theory of musical emotions. *Physics of Life Reviews*, 10, 235–266. DOI: 10.1016/j.plrev.2013.05.008
- KARAGEORGHIS, C. I., & TERRY, P. C. (2008). The psychological, psychophysical and ergogenic effects of music. In A. Bateman & J. Bale (Eds.), *Sport: A review and synthesis* (pp. 27–50). London, UK: Routledge.
- KAWASE, S. (2014). Gazing behavior and coordination during piano duo performance. *Attention, Perception, and Psychophysics*, 76, 527–540. DOI: 10.3758/s13414-013-0568-0
- KEIL, C. (1994). Participatory discrepancies and the power of music. In C. Keil & S. Feld (Eds.), *Music grooves* (pp. 96–108). Chicago, IL: University of Chicago Press. (Original work published 1987)
- KEIL, C. (1995). The theory of participatory discrepancies: A progress report. *Ethnomusicology* 39, 1–19. DOI: 10.2307/852198
- KELLER, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva, (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions* (pp. 205–221). Amsterdam, Netherlands: IOS Press.
- KELLER, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 260–282). Oxford, UK: Oxford University Press.
- KELLER, P. E., & APPEL, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*, 28, 27–46. DOI: 10.1525/mp.2010.28.1.27
- KELLER, P. E., KNOBLICH, G., & REPP, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, 16, 102–111. DOI: 10.1016/j.concog.2005.12.004
- KELLER, P. E., NOVEMBRE, G., & LOEHR, J. (2016). Musical ensemble performance: Representing self, other, and joint action outcomes. In E. S. Cross & S. S. Obhi (Eds.), *Shared representations: Sensorimotor foundations of social life* (pp. 280–310). Cambridge, UK: Cambridge University Press.
- KELLER, P. E., & REPP, B. H. (2005). Staying offbeat: Sensorimotor syncopation with structured and unstructured auditory sequences. *Psychological Research*, 69, 292–309. DOI: 10.1007/s00426-004-0182-9
- KING, E., & GINSBORG, J. (2011). Gestures and glances: Interactions in ensemble rehearsal. In E. King (Ed.), *New perspectives on music and gesture* (pp. 177–201). Aldershot, UK: Ashgate Press.
- KIRSCHNER, S., & TOMASELLO, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, 102, 299–314. DOI: 10.1016/j.jecp.2008.07.005
- KIRSCHNER, S., & TOMASELLO, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, 31, 354–364. DOI: 10.1016/j.evolhumbehav.2010.04.004
- KNOBLICH, G., BUTTERFILL, S., & SEBANZ, N. (2011). Psychological research on joint action: Theory and data. *Psychology of Learning and Motivation*, 54, 59–101. DOI: 10.1016/B978-0-12-385527-5.00003-6
- KNOBLICH, G., & SEBANZ, N. (2008). Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 2021–2031. DOI: 10.1098/rstb.2008.0006
- KONVALINKA, I., BAUER, M., STAHLHUT, C., HANSEN, L. K., ROEPSTORFF, A., & FRITH, C. D. (2014). Frontal alpha oscillations distinguish leaders from followers: Multivariate decoding of mutually interacting brains. *NeuroImage*, 94, 79–88. DOI: 10.1016/j.neuroimage.2014.03.003



- KOTZ, S., RAVIGNANI, A., & FITCH, W. T. (2018). The evolution of rhythm processing. *Trends in Cognitive Sciences*, 22, 896–910. DOI: 10.1016/j.tics.2018.08.002
- KURAMOTO, Y. (1989). *Cooperative dynamics in complex physical systems*. Berlin: Springer.
- LABBÉ, C., GLOWINSKI, D., & GRANDJEAN, D. (2017). The effect of ensemble and solo performance on affective experiences induced by music. *Psychology of Music*, 45, 600–608. DOI: 10.1177/0305735616659551
- LABBÉ, C., & GRANDJEAN, D. (2014). Musical emotions predicted by feelings of entrainment. *Music Perception*, 32, 170–185. DOI: 10.1525/mp.2014.32.2.170
- LARGE, E. W. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (Ed), *The psychology of time* (pp. 189–232). Bingley, UK: Emerald.
- LARGE, E. W., & GRAY, P. M. (2015). Spontaneous tempo and rhythmic entrainment in a bonobo (*Pan paniscus*). *Journal of Comparative Psychology*, 129, 317–328. DOI: 10.1037/com0000011
- LARGE, E., HERRERA, J. A., & VELASCO, M. (2015). Neural networks for beat perception in musical rhythm. *Frontiers in Systems Neuroscience*, 9. DOI: 10.3389/fnsys.2015.00159
- LARGE, E. W., & JONES, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119–159. DOI: 10.1037/0033-295X.106.1.119
- LARGE, E. W., & KOLEN, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, 6, 177–208. DOI: 10.1080/09540099408915723
- LARGE, E. W., & PALMER, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26, 1–37. DOI: 10.1207/s15516709cog2601\_1
- LARGE, E. W., & SNYDER, J. S. (2009). Pulse and meter as neural resonance. *Annals of the New York Academy of Sciences*, 1169, 46–57. DOI: 10.1111/j.1749-6632.2009.04550.x
- LAUNAY, J., DEAN, R. T., & BAILES, F. (2013). Synchronization can influence trust following virtual interaction. *Experimental Psychology*, 60, 53–63. DOI: 10.1027/1618-3169/a000173
- LAUNAY, J., & TARR, B., & DUNBAR, R. (2016). Synchrony as an adaptive mechanism for large-scale human social bonding. *Ethology*, 122, 779–789. DOI: 10.1111/eth.12528
- LEMAN, M. (2007). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.
- LEMAN, M. (2016). *The expressive moment: How interaction (with music) shapes human empowerment*. Cambridge, MA: MIT Press.
- LENC, T., KELLER, P. E., VARLET, M., & NOZARADAN, S. (2018a). Neural tracking of the musical beat is enhanced by low-frequency sounds. *Proceedings of the National Academy of Sciences*, 115, 8221–8226. DOI: 10.1073/pnas.1801421115
- LENC, T., KELLER, P. E., VARLET, M., & NOZARADAN, S. (2018b). Reply to Novembre and Iannetti: Conceptual and methodological issues. *Proceedings of the National Academy of Sciences*, 115, E11004–E11004. DOI: 10.1073/pnas.1815750115
- LENC, T., KELLER, P. E., VARLET, M., & NOZARADAN, S. (2019). Reply to Rajendran and Schnupp: Frequency tagging is sensitive to the temporal structure of signals. *Proceedings of the National Academy of Sciences*, 116, 2781–2782. DOI: 10.1073/pnas.1820941116
- LOMAX, A. (1968). *Folk song style and culture*. New Brunswick, Canada: American Association for the Advancement of Science.
- LONDON, J. (2012). *Hearing in time: Psychological aspects of musical meter* (2nd ed.). Oxford, UK: Oxford University Press.
- LONDON, J., NYMOEN, K., LANGERØD, M. T., THOMPSON, M. R., CODE, D. L., & DANIELSEN, A. (2019). A comparison of methods for investigating the perceptual center of musical sounds. *Attention, Perception, and Psychophysics*, 81, 2088–2101. DOI: 10.3758/s13414-019-01747-y
- LOVE, S. A., PETRINI, K., CHENG, A., & POLLOCK, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments. *PLoS ONE*, 8, e54798. DOI: 10.1371/journal.pone.0054798
- LUCAS, G., CLAYTON, M., & LEANTE, L. (2011). Inter-group entrainment in Afro-Brazilian congado ritual. *Empirical Musicology Review*, 6, 75–102. DOI: 10.18061/1811/51203
- MACRITCHIE, J., VARLET, M., & KELLER, P. E. (2017). Embodied expression through entrainment and co-representation in musical ensemble performance. In M. Lesaffre, P. J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 150–159). New York: Routledge.
- MADISON, G., & MERKER, B. H. (2002). On the limits of asynchrony in pulse attribution. *Psychological Research*, 66, 201–207. DOI: 10.1007/s00426-001-0085-y
- MADISON, G., & MERKER, B. H. (2004). Human sensorimotor tracking of continuous subliminal deviations from isochrony. *Neuroscience Letters*, 370, 69–73. DOI: 10.1016/j.neulet.2004.07.094
- MALAPANI, C., & FAIRHURST, S. (2002). Scalar timing in animals and humans. *Learning and Motivation*, 33, 156–176. DOI: 10.1006/lmot.2001.1105
- MARDIA, K. V., & JUPP, P. E. (2009). *Directional statistics*. New York: John Wiley & Sons.
- MATES, J. (1994a). A model of synchronization of motor acts to a stimulus sequence. I. Timing and error corrections. *Biological Cybernetics*, 70, 463–473. DOI: 10.1007/BF00203239
- MATES, J. (1994b). A model of synchronization of motor acts to a stimulus sequence: II. Stability analysis, error estimation and simulations. *Biological Cybernetics*, 70, 475–484. DOI: 10.1007/bf00203240

- MAY, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261, 459–467. DOI: 10.1038/261459a0
- MCAULEY, J. D., JONES, M. R., HOLUB, S., JOHNSTON, H. M., & MILLER, N. S. (2006). The time of our lives: Life span development of timing and event tracking. *Journal of Experimental Psychology: General*, 135, 348–367. DOI: 10.1037/0096-3445.135.3.348
- MCAULEY, J. D., & KIDD, G. R. (1995). Temporally directed attending in the discrimination of tempo: Further support for an entrainment model. *Journal of the Acoustical Society of America*, 97, 3278–3278. DOI: 10.1121/1.411574
- MCNEILL, W. H. (1995). *Keeping together in time: Dance and drill in human history*. Cambridge, MA: Harvard University Press.
- MERCHANT, H., & HONING, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, 7, 274. DOI: 10.3389/fnins.2013.00274
- MERKER, B. (2000). Synchronous chorusing and human origins. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 315–328). Cambridge, MA: MIT Press. DOI: 10.1016/j.cortex.2008.06.011
- MERKER, B. H., MADISON, G. S., & ECKERDAL, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, 45, 4–17. DOI: 10.1016/j.cortex.2008.06.011
- MILES, L. K., NIND, L. K., & MACRAE, C. N. (2009). The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology*, 45, 585–589. DOI: 10.1016/j.jesp.2009.02.002
- MILLS, P. F., HARRY, B., STEVENS, C. J., KNOBLICH, G., & KELLER, P. E. (2019). Intentionality of a co-actor influences sensorimotor synchronisation with a virtual partner. *Quarterly Journal of Experimental Psychology*, 72, 1478–1492. DOI: 10.1177/1747021818796183
- MITHEN, S. (2005). *The singing Neanderthals. The origins of music, language, mind and body*. London, UK: Weidenfeld and Nicholson.
- MOGAN, R., FISCHER, R., & BULBULIA, J. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *Journal of Experimental Social Psychology*, 72, 13–20. DOI: 10.1016/j.jesp.2017.03.009
- MORAN, N. (2013). Social co-regulation and communication in North Indian duo performances. In M. Clayton, B. Dueck, & L. Leante (Eds.), *Experience and meaning in music performance* (pp. 64–94). Oxford, UK: Oxford University Press.
- MORAN, N., HADLEY, L. V., BADER, M., & KELLER, P. E. (2015). Perception of 'back-channeling' nonverbal feedback in musical duo improvisation. *PLoS ONE*, 10, e0130070. DOI: 10.1371/journal.pone.0130070
- MORAN, N., JAKUBOWSKI, K., & KELLER, P. E. (2017). *Improvising duos - Visual interaction collection*, 2011 [moving image]. Edinburgh, Scotland: University of Edinburgh. DOI: 10.7488/ds/2153
- MORELAND, R. L., LEVINE, J. M., & WINGERT, M. L. (1996). Creating the ideal group: Composition effects at work. In J. Davis & E. Witte (Eds.), *Understanding group behavior* (Vol. 2, pp. 11–35). Mahwah, NJ: Lawrence Erlbaum.
- MORILLON, B., & SCHROEDER, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction: Neuronal oscillations and temporal predictions. *Annals of the New York Academy of Sciences*, 1337, 26–31. DOI: 10.1111/nyas.12629
- NAEEM, M., PRASAD, G., WATSON, D. R., & KELSO, J. S. (2012). Electrophysiological signatures of intentional social coordination in the 10–12 Hz range. *Neuroimage*, 59, 1795–1803. DOI: 10.1016/j.neuroimage.2011.08.010
- NAGASAKA, Y., CHAO, Z. C., HASEGAWA, N., NOTOYA, T., & FUJII, N. (2013). Spontaneous synchronization of arm motion between Japanese macaques. *Scientific Reports*, 3, 1151. DOI: 10.1038/srep01151
- NÉDA, Z., RAVASZ, E., BRECHET, Y., VICSEK, T., & BARABÁSI, A.-L. (2000). Self-organizing processes: The sound of many hands clapping. *Nature*, 403, 849–850. DOI: 10.1038/35002660
- NOVEMBRE, G., & IANNETTI, G. D. (2018). Tagging the musical beat: Neural entrainment or event-related potentials? *Proceedings of the National Academy of Sciences*, 115, E11002–E11003. DOI: 10.1073/pnas.1815311115
- NOVEMBRE, G., & KELLER, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: What is it good for? *Frontiers in Human Neuroscience*, 8, 603. DOI: 10.3389/fnhum.2014.00603
- NOVEMBRE, G., SAMMLER, D., & KELLER, P. E. (2016). Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action. *Neuropsychologia*, 89, 414–425. DOI: 10.1016/j.neuropsychologia.2016.07.027
- NOZARADAN, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society B*, 369, 20130393. DOI: 10.1098/rstb.2013.0393
- NOZARADAN, S., PERETZ, I., & KELLER, P. E. (2016). Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Scientific Reports*, 6, 20612. DOI: 10.1038/srep20612
- NOZARADAN, S., PERETZ, I., MISSAL, M., & MOURAUX, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, 31, 10234–10240. 10.1523/JNEUROSCI.0411-11.2011

- NOZARADAN, S., PERETZ, I., & MOURAUX, A. (2012a). Steady-state evoked potentials as an index of multisensory temporal binding. *NeuroImage*, 60, 21–28. DOI: 10.1016/j.neuroimage.2011.11.065
- NOZARADAN, S., PERETZ, I., & MOURAUX, A. (2012b). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *Journal of Neuroscience*, 32, 17572–17581. DOI: 10.1523/JNEUROSCI.3203-12.2012
- NOZARADAN, S., SCHÖNWIESNER, M., CARON-DESROCHERS, L., & LEHMANN, A. (2016). Enhanced brainstem and cortical encoding of sound during synchronized movement. *NeuroImage*, 142, 231–240. DOI: 10.1016/j.neuroimage.2016.07.015
- NOZARADAN, S., ZEROUALI, Y., PERETZ, I., & MOURAUX, A. (2015). Capturing with EEG the neural entrainment and coupling underlying sensorimotor synchronization to the beat. *Cerebral Cortex*, 25, 736–747. DOI: 10.1093/cercor/bht261
- PALMER, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331–346. DOI: 10.1037/0096-1523.15.2.331
- PALMER, C., LIDJI, P., & PERETZ, I. (2014). Losing the beat: Deficits in temporal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130405 DOI: 10.1098/rstb.2013.0405
- PATEL, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, 24, 99–104. DOI: 10.1525/mp.2006.24.1.99
- PATEL, A. D., & IVERSEN, J. R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, 8, 57. DOI: 10.3389/fnsys.2014.00057
- PATEL, A. D., IVERSEN, J. R., BREGMAN, M. R., & SCHULZ, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, 19, 827–830. DOI: 10.1016/j.cub.2009.03.038
- PATEL, A. D., IVERSEN, J., CHEN, Y., & REPP, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research*, 163, 226–238. DOI: 10.1007/s00221-004-2159-8
- PATEL, A. D., IVERSEN, J. R., & OHGUSHI, K. (2006). Nonlinguistic rhythm perception depends on culture and reflects the rhythms of speech: Evidence from English and Japanese. *Acoustical Society of America Journal*, 120, 3167. DOI: 10.1121/1.4787902
- PECENKA, N., ENGEL, A., & KELLER, P. E. (2013). Neural correlates of auditory temporal predictions during sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7, 380. DOI: 10.3389/fnhum.2013.00380
- PECENKA, N., & KELLER, P. E. (2009a). Auditory pitch imagery and its relationship to musical synchronization. *Annals of the New York Academy of Sciences*, 1169, 282–286. DOI: 10.1111/j.1749-6632.2009.04785.x
- PECENKA, N., & KELLER, P. E. (2009b). The relationship between auditory imagery and musical synchronization abilities in musicians. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P.-S. Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music* (pp. 409–414). Jyväskylä, Finland: University of Jyväskylä.
- PECENKA, N., & KELLER, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, 211, 505–515. DOI: 10.1007/s00221-011-2616-0
- PETRINI, K., RUSSELL, M., & POLLOCK, F. (2009). When knowing can replace seeing in audiovisual integration of actions. *Cognition*, 110, 432–439. DOI: 10.1016/j.cognition.2008.11.015
- PHILLIPS-SILVER, J., & AKTIPIS, C., & BRYANT, G. A. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Perception*, 28, 3–14. DOI: 10.1525/mp.2010.28.1.3
- PHILLIPS-SILVER, J., & KELLER, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*, 6, 26. DOI: 10.3389/fnhum.2012.00026.
- PHILLIPS-SILVER, J., & TRAINOR, L. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science*, 308, 1430. DOI: 10.1126/science.1110922
- POLAK, R., LONDON, J., & JACOBY, N. (2016). Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study of Malian jembe drumming. *Frontiers in Neuroscience*, 10, 285. DOI: 10.3389/fnins.2016.00285
- POLAK, R., TARSITANI, S., & CLAYTON, M. (2018, September 26). *IEMP Malian Jembe* (Open Science Framework). DOI: 10.17605/OSF.IO/M652X
- POOLE, A. (2012). *The aesthetics of groove in Cuban dance music: An analysis of son and salsa* (PhD dissertation). Open University, Milton Keynes, England.
- POOLE, A., TARSITANI, S., & CLAYTON, M. (2019). *IEMP Cuban Son & Salsa* (Open Science Framework). DOI: 10.17605/OSF.IO/SFXA2
- POVEL, D.-J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3–18. DOI: 10.1037/0096-1523.7.1.3
- POVEL, D.-J., & ESSENS, P. (1985). Perception of temporal patterns. *Music Perception*, 2, 411–440.
- PRESSING, J. (1998). Error correction processes in temporal pattern production. *Journal of Mathematical Psychology*, 42, 63–101. DOI: 10.1006/jmps.1997.1194

- PRÖGLER, J. A. (1995). Searching for swing. *Ethnomusicology*, 39, 21–54. DOI: 10.2307/852199
- QUIROGA, R. Q., KREUZ, T., & GRASSBERGER, P. (2002). Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Physical Review E*, 66, 041904. DOI: 10.1103/PhysRevE.66.041904
- RAGERT, M., FAIRHURST, M. T., & KELLER, P. E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PLoS ONE*, 9, e84085. DOI: 10.1371/journal.pone.0084085
- RAGERT, M., SCHROEDER, T., & KELLER, P. E. (2013). Knowing too little or too much: The effects of familiarity with a co-performer's part on interpersonal coordination in musical ensembles. *Frontiers in Psychology*, 4, 368. DOI: 10.3389/fpsyg.2013.00368
- RAJENDRAN, V. G., & SCHNUPP, J. W. (2019). Frequency tagging cannot measure neural tracking of beat or meter. *Proceedings of the National Academy of Sciences*, 116, 2779–2780. DOI: 10.1073/pnas.1820020116
- RANKIN, S. K., LARGE, E. W., & FINK, P. W. (2009). Fractal tempo fluctuation and pulse prediction. *Music Perception*, 26, 401–413. DOI: 10.1525/MP.2009.26.5.401
- RASCH, R. A. (1979). Synchronization in performed ensemble music. *Acustica*, 43, 121–131.
- RASCH, R. A. (1988). Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 70–90). Oxford, UK: Clarendon Press.
- REDDISH, P., FISCHER, R., & BULBULIA, J. (2013). Let's dance together: Synchrony, shared intentionality and cooperation. *PLoS ONE*, 8, e71182. DOI: 10.1371/journal.pone.0071182
- REPP, B. H. (1999). Relationships between performance timing, perception of timing perturbations, and perceptual-motor synchronization in two Chopin Preludes. *Australian Journal of Psychology*, 51, 188–203. DOI: 10.1080/00049539908255356
- REPP, B. H. (2001). Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 600–621. DOI: 10.1037/0096-1523.27.3.600
- REPP, B. H. (2002). The embodiment of musical structure: Effects of musical context on sensorimotor synchronization with complex timing patterns. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and performance XIX* (pp. 245–265). Oxford, UK: Oxford University Press.
- REPP, B. H. (2003). Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. *Journal of Motor Behavior*, 35, 355–370. DOI: 10.1080/00222890309603156
- REPP, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review*, 12, 969–992. DOI: 10.3758/BF03206433
- REPP, B. H. (2006). Rate limits of sensorimotor synchronization. *Advances in Cognitive Psychology*, 2, 163–181. DOI: 10.2478/v10053-008-0053-9
- REPP, B. H., & KELLER, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *Quarterly Journal of Experimental Psychology Section A*, 57, 499–521. DOI: 10.1080/02724980343000369
- REPP, B. H., LONDON, J., & KELLER, P. E. (2008). Phase correction in sensorimotor synchronization with nonisochronous sequences. *Music Perception*, 26, 171–175. DOI: 10.1525/mp.2008.26.2.171
- REPP, B. H., & SU, Y. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin and Review*, 20, 403–452. DOI: 10.3758/s13423-012-0371-2
- RICHARDSON, D., DALE, R., & SHOCKLEY, K. (2008). Synchrony and swing in conversation: Coordination, temporal dynamics, and communication. In Wachsmuth, I., Lenzen, M., & Knoblich, G. (Eds.), *Embodied Communication in Humans and Machines*. New York: Oxford University Press (pp. 75–93). DOI: 10.1093/acprof:oso/9780199231751.003.0004
- RICHARDSON, M. J., GARCIA, R. L., FRANK, T. D., GERGOR, M., & MARSH, K. L. (2012). Measuring group synchrony: A cluster-phase method for analyzing multivariate movement time-series. *Frontiers in Physiology*, 3, 405. DOI: 10.3389/fphys.2012.00405
- RICHARDSON, M., MARSH, K., ISENHOWER, R., GOODMAN, J. R. L., & SCHMIDT, R. (2008). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26, 867–891. DOI: 10.1016/j.humov.2007.07.002
- RIMMELE, J. M., MORILLON, B., POEPEL, D., & ARNAL, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, 22, 870–882. DOI: 10.1016/j.tics.2018.08.003
- RIZZOLATTI, G., & SINIGAGLIA, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience*, 11, 264–274. DOI: 10.1038/nrn2805
- ROUSE, A. A., COOK, P. F., LARGE, E. W., & REICHMUTH, C. (2016). Beat keeping in a sea lion as coupled oscillation: Implications for comparative understanding of human rhythm. *Frontiers in Neuroscience*, 10, 257. DOI: 10.3389/fnins.2016.00257
- SADAKATA, M., OHGUSHI, K., & DESAIN, P. (2004). A cross-cultural comparison study of the production of simple rhythmic patterns. *Psychology of Music*, 32, 389–403. DOI: 10.1177/0305735604046097



- SCHACHNER, A. (2010). Auditory-motor entrainment in vocal mimicking species: Additional ontogenetic and phylogenetic factors. *Communicative and Integrative Biology*, 3, 290–293. DOI: 10.4161/cib.3.3.11708
- SCHACHNER, A., BRADY, T. F., PEPPERBERG, I. M., & HAUSER, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology*, 19, 831–836. DOI: 10.1016/j.cub.2009.03.061
- SCHMIDT, R. C., & RICHARDSON, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, behavioral and social dynamics* (pp. 281–308). Berlin, Heidelberg: Springer.
- SCHMIDT, R. C., RICHARDSON, M. J., ARSENAULT, C., & GALANTUCCI, B. (2007). Visual tracking and entrainment to an environmental rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 860–870. DOI: 10.1037/0096-1523.33.4.860
- SCHÖGLER, B. (1999). Studying temporal co-ordination in jazz duets. *Musicae Scientiae, Special Issue 1999–2000*, 75–91. DOI: 10.1177/10298649000030S106
- SCHROEDER, C. E., WILSON, D. A., RADMAN, T., SCHARFMAN, H., & LAKATOS, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, 20, 172–176. DOI: 10.1016/j.conb.2010.02.010
- SCHULZE, H.-H., CORDES, A., & VORBERG, D. (2005). Keeping synchrony while tempo changes: Accelerando and ritardando. *Music Perception*, 22, 461–477. DOI: 10.1525/mp.2005.22.3.461
- SEBANZ, N., BEKKERING, H., & KNOBLICH, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76. DOI: 10.1016/j.tics.2005.12.009
- SEBANZ, N., & KNOBLICH, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1, 353–367. DOI: 10.1111/j.1756-8765.2009.01024.x
- SEBANZ, N., KNOBLICH, G., & PRINZ, W. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1234–1246. DOI: 10.1037/0096-1523.31.6.1234
- SENN, O., KILCHENMANN, L., BECHTOLD, T., & HOESL, F. (2018). Groove in drum patterns as a function of both rhythmic properties and listeners' attitudes. *PLoS ONE*, 13, e0199604. DOI: 10.1371/journal.pone.0199604
- SENN, O., KILCHENMAN, L., VON GEORGI, R., & BULLERJAHN, C. (2016). The effect of expert performance microtiming on listeners' experience of groove in swing or funk music. *Frontiers in Psychology*, 7, 1487. DOI: 10.3389/fpsyg.2016.01487
- SERRÀ, J., SERRA, X., & ANDRZEJAK, R. G. (2009). Cross recurrence quantification for cover song identification. *New Journal of Physics*, 11, 093017.
- SHAFFER, L. H. (1984). Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology*, 36A, 577–595. DOI: 10.1080/14640748408402180
- SHOCKLEY, K. (2005). Cross recurrence quantification of interpersonal postural activity. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 142–177). Retrieved March 1, 2018, from <http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp>
- SHOCKLEY, K., RICHARDSON, D., & DALE, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1, 305–319. DOI: 10.1111/j.1756-8765.2009.01021.x
- SHOCKLEY, K., SANTANA, M.-V., & FOWLER, C. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326–332. DOI: 10.1037/0096-1523.29.2.326
- SLOETJES, H., & WITTENBURG, P. (2008). Annotation by category-ELAN and ISO DCR. In *6th International Conference on Language Resources and Evaluation (LREC 2008)*. Marrakesh, Morocco: LREC.
- SNYDER, J. S., & LARGE, E. W. (2004). Tempo dependence of middle-and long-latency auditory responses: Power and phase modulation of the EEG at multiple timescales. *Clinical Neurophysiology*, 115, 1885–1895. DOI: 10.1016/j.clinph.2004.03.024
- SNYDER, J. S., & LARGE, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, 24, 117–126. DOI: 10.1016/j.cogbrainres.2004.12.014
- STUPACHER, J., WITTE, M., HOVE, M. J., & WOOD, G. (2016). Neural entrainment in drum rhythms with silent breaks: evidence from steady-state evoked and event-related potentials. *Journal of Cognitive Neuroscience*, 28, 1865–1877. DOI: 10.1162/jocn\_a\_01013
- TAL, I., LARGE, E. W., RABINOVITCH, E., WEI, Y., SCHROEDER, C. E., POEPPPEL, D., & GOLUMBIC, E. Z. (2017). Neural entrainment to the beat: The “missing-pulse” phenomenon. *Journal of Neuroscience*, 37, 6331–6341. DOI: 10.1523/JNEUROSCI.2500-16.2017
- TARR, B., LAUNAY, J., COHEN, E., & DUNBAR, R. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biology Letters*, 11, 20150767. DOI: 10.1098/rsbl.2015.0767
- TARR, B., LAUNAY, J., & DUNBAR, R. I. (2014). Music and social bonding: “Self-other” merging and neurohormonal mechanisms. *Frontiers in Psychology*, 5, 1096. DOI: 10.3389/fpsyg.2014.01096
- TEIXEIRA, E. C., LOUREIRO, M. A., WANDERLEY, M. M., & YEHA, H. C. (2015). Motion analysis of clarinet performers. *Journal of New Music Research*, 44, 97–111. DOI: 10.1080/09298215.2014.925939
- THAUT, M. H., MCINTOSH, G. C., & HOEMBERG, V. (2015). Neurobiological foundations of neurologic music therapy: rhythmic entrainment and the motor system. *Frontiers in Psychology*, 5, 1185. DOI: 10.3389/fpsyg.2014.01185

- TIERNEY, A., & KRAUS, N. (2013). The ability to move to a beat is linked to the consistency of neural responses to sound. *The Journal of Neuroscience*, 33, 14981–14988. DOI: 10.1523/JNEUROSCI.0612-13.2013
- TIERNEY, A., & KRAUS, N. (2014). Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience*, 8, 949. DOI: 10.3389/fnhum.2014.00949
- TIERNEY, A., & KRAUS, N. (2015). Evidence for multiple rhythmic skills. *PloS ONE*, 10, e0136645. DOI: 10.1371/journal.pone.0136645
- TOGNOLI, E., LAGARDE, J., DE GUZMAN, G.C. & KELSO, J.A.S. (2007). From the cover: The phi-complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences*, 104, 8190–8195. DOI: 10.1073/pnas.0611453104
- TOMLINSON, G. (2015). *A million years of music. The emergence of human modernity*. New York: Zone Books.
- TORRENCE, C., & COMPO, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79, 61–78. DOI: 10.1175/1520-0477
- TREISMAN, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the “internal clock.” *Psychological Monographs: General and Applied*, 77, 1–31. DOI: 10.1037/h0093864
- TROST, W. J., LABBÉ, C., & GRANDJEAN, D. (2017). Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia*, 96, 96–110. DOI: 10.1016/j.neuropsychologia.2017.01.004
- UHLIG, M., FAIRHURST, M. T., & KELLER, P. E. (2013). The importance of integration and top-down salience when listening to complex multi-part musical stimuli. *NeuroImage*, 77, 52–61. DOI: 10.1016/j.neuroimage.2013.03.051
- VALLACHER, R. R., & NOWAK, A. E. (1994). *Dynamical systems in social psychology*. New York: Academic Press.
- VAN DER STEEN, M. C., & KELLER, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7, 253. DOI: 10.3389/fnhum.2013.00253
- VAN DER STEEN, M. C., JACOBY, N., FAIRHURST, M. T., & KELLER, P. E. (2015). Sensorimotor synchronization with tempo-changing auditory sequences: Modeling temporal adaptation and anticipation. *Brain Research*, 1626, 66–87. DOI: 10.1016/j.brainres.2015.01.053
- VAN NOORDEN, L., & MOELANTS, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, 28, 43–66. DOI: 10.1076/jnmr.28.1.43.3122
- VARNI G., VOLPE G., & CAMURRI, A. (2010). A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. *IEEE Transactions on Multimedia*, 12, 576–590. DOI: 10.1109/TMM.2010.2052592
- VELASCO, M. J., & LARGE, E. W. (2011). Pulse detection in syncopated rhythms using neural oscillators. *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011*. 185–190.
- VESPER, C., ABRAMOVA, E., BUTEPAGE, J., CIARDO, F., CROSSEY, B., EFFENBERG, A., ET AL. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology*, 7, 2039. DOI: 10.3389/fpsyg.2016.02039
- VESPER, C., BUTTERFILL, S., KNOBLICH, G., & SEBANZ, N. (2010). A minimal architecture for joint action. *Neural Networks*, 23), 998–1003. DOI: 10.1016/j.neunet.2010.06.002
- VICARY, S., SPERLING, M., VON ZIMMERMANN, J., RICHARDSON, D. C., & ORGS, G. (2017). Joint action aesthetics. *PloS ONE*, 12, e0180101. DOI: 10.1371/journal.pone.0180101
- VINES, B. W., KRUMHANS, C. L., WANDERLEY, M. M., & LEVITIN, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, 101, 80–113. DOI: 10.1016/j.cognition.2005.09.003
- VORBERG, D., & SCHULZE, H.-H. (2002). Linear phase-correction in synchronization: Predictions, parameter estimation, and simulations. *Journal of Mathematical Psychology*, 46, 56–87. DOI: 10.1006/jmps.2001.1375
- VORBERG, D., & WING, A. M. (1996). Modeling variability and dependence in timing. In S. W. Keele (Ed.), *Handbook of perception and action* (Vol. 2, pp. 181–262). London, UK: Academic Press.
- VOS, P. G., MATES, J., & VAN KRUYSBERGEN, N. W. (1995). The perceptual centre of a stimulus as the cue for synchronization to a metronome: Evidence from asynchronies. *Quarterly Journal of Experimental Psychology. Human Experimental Psychology*, 48, 1024–1040. DOI: 10.1080/14640749508401427
- VOS, J., & RASCH, R. A. (1981). The perceptual onset of musical tones. *Perception and Psychophysics*, 29, 323–335. DOI: 10.3758/bf03207341
- WALTON, A. E., RICHARDSON, M. J., LANGLAND-HASSAN, P., & CHERO, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Frontiers in Psychology*, 6, 313. DOI: 10.3389/fpsyg.2015.00313
- WANDERLEY, M. M., VINES, B. W., MIDDLETON, N., MCKAY, C., & HATCH, W. (2005). The musical significance of clarinetists’ ancillary gestures: An exploration of the field. *Journal of New Music Research*, 34, 97–113. DOI: 10.1080/09298210500124208
- WILLIAMON, A., & DAVIDSON, J. W. (2002). Exploring co-performer communication. *Musicae Scientiae*, 6, 53–72. DOI: 10.1177/102986490200600103
- WILSON, M., & COOK, P. F. (2016). Rhythmic entrainment: Why humans want to, fireflies can’t help it, pet birds try, and sea lions have to be bribed. *Psychonomic Bulletin and Review*, 23, 1647–1659. DOI: 10.3758/s13423-016-1013-x

- WILTERMUTH, S. S., & HEATH, C. (2009). Synchrony and cooperation. *Psychological Science*, 20, 1–5. DOI: 10.1111/j.1467-9280.2008.02253.x
- WING, A. M., ENDO, S., BRADBURY, A., & VORBERG, D. (2014). Optimal feedback correction in string quartet synchronization. *Journal of The Royal Society Interface*, 11, 20131125. DOI: 10.1098/rsif.2013.1125
- WING, A. M., & KRISTOFFERSON, A. B. (1973a). Response delays and the timing of discrete motor responses. *Attention, Perception, and Psychophysics*, 14, 5–12. DOI: 10.3758/BF03198607
- WING, A. M., & KRISTOFFERSON, A. B. (1973b). The timing of interresponse intervals. *Attention, Perception, and Psychophysics*, 13, 455–460. DOI: 10.3758/BF03205802
- WINKLER, I., HÄDEN, G. P., LADINIG, O., SZILLER, I., & HONING, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, 106, 2468–2471. DOI: 10.1073/pnas.0809035106
- WÖLLNER, C., & CANAL-BRULAND, R. (2010). Keeping an eye on the violinist: Motor experts show superior timing consistency in a visual perception task. *Psychological Research*, 74, 579–585. DOI: 10.1007/s00426-010-0280-9
- ZAKAY, D., & BLOCK, R. A. (1996). The role of attention in time estimation processes. *Advances in Psychology*, 115, 143–164. DOI: 10.1016/S0166-4115(96)80057-4
- ZANTO, T. P., LARGE, E. W., FUCHS, A., & KELSO, J. S. (2005). Gamma-band responses to perturbed auditory sequences: Evidence for synchronization of perceptual processes. *Music Perception*, 22, 531–547. DOI: 10.1525/mp.2005.22.3.531
- ZANTO, T. P., SNYDER, J. S., & LARGE, E. W. (2006). Neural correlates of rhythmic expectancy. *Advances in Cognitive Psychology*, 2, 221–231. DOI: 10.2478/v10053-008-0057-5
- ZARCO, W., MERCHANT, H., PRADO, L., & MENDEZ, J. C. (2009). Subsecond timing in primates: Comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology*, 102, 3191–3202. DOI: 10.1152/jn.00066.2009
- ZENTNER, M. R., & EEROLA, T. (2010). Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, 107, 5768–5773. DOI: 10.1073/pnas.1000121107