Interpersonal neural synchrony in joint music-making and conversation: toward an integrative Marr-level account

Juan-Pablo Robledo[*1,2], Ian Cross[3], Michelle Phillips[4], Josephine F. Kearney[5], Jason R. Taylor[5]

[1]Laboratoire INTERPSY, Département de Psychologie, Université de Lorraine, Nancy, France
[2]Millennium Institute for Care Research (MICARE), Santiago, Chile
[3]Centre for Music and Science, Music Department, The University of Cambridge, UK
[4]Royal Northern College of Music, UK
[5]Division of Psychology, Communication, and Human Neuroscience, School of Health Sciences, University of Manchester, UK

[*]Corresponding author: Juan-Pablo Robledo. Email: juan-pablo.robledo-del-canto@univ-lorraine.fr

Abstract

Joint music-making and conversation are two fundamental forms of human interaction. A growing number of hyperscanning studies have examined interpersonal neural synchrony (INS) in music and language separately, yet few have sought to connect findings across these domains or interpret them within a unified theoretical framework. Recent reviews have underscored the need for systematic approaches that move beyond isolated reports of brain regions or oscillatory patterns. This article answers that call by applying David Marr's tripartite framework—computational, algorithmic, and implementational—to INS observed during dyadic interaction in musical and conversational contexts. The aim is to move from descriptive to mechanistic accounts of synchrony.

The article introduces INS as a domain-general mechanism supporting social coordination, situates it within the rise of hyperscanning in social neuroscience, and explains how Marr's framework bridges music and language interaction. To minimize interpretive pitfalls, it discusses methodological caveats specific to hyperscanning and emphasizes the need to distinguish intra-individual from truly interaction-specific processes. The review then synthesizes evidence from musical interaction, including dynamic synchrony, task manipulations, and sensorimotor coupling, followed by findings from conversational studies. Overlapping neural networks are mapped across domains, and Marr's framework is used to interpret these shared systems in terms of the problems they solve, the processes they instantiate, and their neural realizations summarized in a comparative table. The final section outlines future directions and testable hypotheses concerning role asymmetry, mutual adaptation, and dynamic reconfiguration.

Keywords

Hyperscanning, Music, Conversation, Synchrony, Coordination, EEG, Marr-levels

Table of contents

1. Introduction

1.1. Music, language, and the rise of interpersonal neural synchrony in social neuroscience

In recent years, the concept of *interpersonal neural synchrony* (INS) has emerged as a powerful lens through which to understand the dynamic nature of social interaction. INS refers to the temporal alignment of neural activity between individuals during real-time engagement, whether verbal, non-verbal, cooperative, or observational. This synchrony, observable via techniques such as EEG, MEG, or fNIRS hyperscanning, has been shown to accompany a wide range of socially relevant behaviors—joint attention (Leong et al., 2017), affective attunement (Kinreich et al., 2017), coordination (Lindenberger et al., 2009), and verbal turn-taking (Nguyen et al., 2023).

Music and language are both culturally ubiquitous, inherently social and share a number of biological and cognitive features (Besson & Schon, 2011). In terms of function, while language can refer unambiguously and represent and communicate information about complex states of affairs, music seems to be simply expressive and its role that of entertainment; we might enjoy it but, unlike language, it scarcely seems to be a necessary part of human life. But language does not simply encode meaning. Jakobson (1960) identified five functions of language in addition to the referential; these are the emotive, poetic, conative, metalingual and phatic functions. This last, the phatic, acts to establish, maintain and enhance social connections in conversational interactions in which meanings are less important than the fact and act of communication itself (Malinowski, 1923). Enfield (2015: 214) notes that these six functions are non-exclusive and "theoretically well-grounded in the core elements of a communicative act involving two people and a shared code". Similarly, music does not just exist as auditory form that affords pleasure. It is an interactive, participatory medium that may be as much concerned with "…the social relations being realized through the performance" (Turino, 2008:35) as it is with the sound produced. When we look across cultures, music is not limited in its function to entertainment; it does many different things in different societies — including those of the contemporary west — and many of the things that we use it for are shared with language. Music's functions include; the expression and elicitation of affect; the socialisation of infants; the evocation of complex meanings; the underpinning of narrative (in story, film or computer game); the framing of private or collective ritual; the creation and maintenance of interpersonal connections; the articulation of collective identity; and the expression of law. Rather than music and language constituting discrete categories, they are best understood as cultural products and social tools with overlapping functions and features that reflect genetically-shaped human communicative capacities.

The investigation of their interconnection—how music and language shape and influence one another—remains in its early stages. For over two decades, scientists have documented how musical training enhances language-related cognitive functions through overlapping neural and perceptual mechanisms. A growing body of research supports the claim that training in music improves linguistic cognitive function, particularly in the domains of phonological awareness, syntax processing, and auditory discrimination. Studies have shown that musical experience sharpens the brain’s ability to encode speech (Kraus & Chandrasekaran, 2010: Patel, 2011) and improves sensitivity to pitch, rhythm, and prosody—core components of both music and language (Besson et al., 2011; Magne et al., 2006). These effects appear early: musically trained children show superior neural and behavioral responses in tasks involving speech segmentation, grammatical judgments, and verbal memory (Jentschke & Koelsch, 2009; Moreno & Besson, 2006). Meta-analyses and intervention studies confirm that music training can support reading acquisition and phonological processing, including in children with dyslexia (Forgeard et al., 2008; Gordon et al., 2015). Rhythm-based approaches in particular have proven effective in improving temporal prediction and syllabic parsing in speech (Thomson et al., 2013). The overlap between musical and linguistic processing—especially in auditory and syntactic domains—suggests shared neural mechanisms and highlights music as a promising tool for enhancing language development and rehabilitation.

While most of the studies listed in the previous paragraph explored the mid- and long-term cognitive and neural benefits of music training on language abilities in isolated individuals, a growing number of researchers are now turning their attention to the dynamic interplay between music and language as the most interactionally rich forms of real-time human communication. This shift reflects a broader move away from offline, decontextualized tasks toward ecologically valid paradigms that capture spontaneous, temporally coordinated behavior. In this regard, although the role of neural oscillators in music and language has been studied mainly at a unipersonal level (Rietmolen et al., 2024), the application of hyperscanning to musical interaction is slowly growing (Cheng et al., 2024). As a result, there is increasing interest in studying how musical and linguistic real-time exchanges unfold across shorter timescales—such as during conversational turn-taking or joint improvisation—highlighting their shared reliance on predictive timing, mutual adaptation, and interpersonal alignment (Hawkins et al., 2013, Robledo et al., 2021). This emerging perspective sets the stage for hyperscanning approaches, which are uniquely suited to investigate the neural mechanisms supporting these tightly coupled social interactions.

1.2. Bridging musical and linguistic domains through Marr's framework

Despite being traditionally separated by disciplinary boundaries there is mounting evidence that the cognitive and neural foundations of these two domains are deeply intertwined. Both are temporally structured, socially embedded systems that rely on dynamic interaction between individuals (Hawkins et al., 2013). They share core features such as hierarchical organization, predictive timing, turn-taking, and mutual adaptation (Koelsch, 2011; Patel, 2008). These shared structural and functional features may imply that language and music are not only processed by partially overlapping neural systems but may also rely on common computational principles—that is, they may solve analogous functional problems of dynamic interpersonal alignment and predictive coordination. In this regard, INS, as revealed through hyperscanning studies, provides a promising entry point for probing these speculative parallels empirically.

Although neuroscientific data is still scarce, some of it supports such a convergence. Hyperscanning during joint musical improvisation (e.g., Lindenberger et al., 2009; Müller et al., 2013) and verbal dialogue (e.g., Jiang et al., 2016; Sun et al., 2024) reveals remarkably similar patterns of cross-brain coupling, especially in low-frequency bands and across regions like the inferior frontal gyrus, medial prefrontal cortex, and motor areas. This makes sense, as they are both spontaneous, unscripted forms of interaction. These findings support the view that INS may reflect a shared mechanism underlying real-time interpersonal coordination across modalities. Moreover, neuroimaging studies suggest that similar predictive and integrative processes operate in both domains: just as a musician anticipates a partner's phrase, a speaker anticipates a listener's reaction or a partner's turn. These predictive dynamics appear to rely on similar frontotemporal networks and oscillatory signatures (e.g., beta for motor prediction, theta for attention and working memory), however, such correspondences should be taken as conceptual analogies rather than direct evidence of formal predictive-coding computations.

Yet the existence of shared mechanisms does not imply identical functions. Furthermore, a number of articles implicitly warn about a current fragmentation in hyperscanning research and the need for integrative approaches. For example, Hamilton (2021), writing from a broader, multi-domain perspective, critiques the lack of coherence across hyperscanning studies and emphasizes the importance of stronger theoretical grounding and behavioral integration to avoid purely correlational interpretations of interbrain synchrony. In the domain of musical interaction, Lender et al. (2023) report that INS can increase even when behavioral coordination is disrupted, suggesting a dissociation between neural synchrony and performance. Similarly, Kurihara et al. (2022), working in the context of rhythmic motor coordination, find that INS correlates positively with task instability, again implying that increased INS may reflect compensatory effort rather than successful interaction. These findings, spanning musical, motor, and broader domains, underscore the need for comparative frameworks to clarify under which conditions INS indexes coordination versus cognitive compensation, thereby calling—albeit implicitly—for systematic synthesis across paradigms, tasks, and frequency bands.

Marr-level account of INS in music and conversation

In the same vein, in their recent systematic review of hyperscanning research in musical contexts, Cheng et al. (2024) call for more structured approaches to synthesizing findings in musical hyperscanning research, emphasizing the need for systematic thematic categorization across study types (e.g., performance, listening, therapy). Rather than aggregating data through meta-analysis, they advocate for organizing studies by musical domain to better account for methodological diversity and conceptual fragmentation. The authors also highlight the lack of unified cognitive interpretations of INS, noting that the variety of synchrony metrics and experimental paradigms has led to inconsistent explanations of what INS represents—ranging from shared attention to compensatory effort. To address this, they recommend grounding future research in Predictive Coding Models and computational frameworks that can formalize INS as an emergent property of joint information processing, thereby linking interbrain measures to clearly defined cognitive functions.

One element that could respond to these challenges is a well-established theoretical framework capable of distinguishing superficial similarities from deeper architectural correspondences. Marr's (1982) three-level framework provides just such a tool. At the computational level, one may ask: what problem are these systems solving? In both joint music-making and conversation, this may involve constructing and maintaining a shared temporal and intentional model between individuals. At the algorithmic level, both systems may use oscillatory phase alignment, internal simulations, and prediction–error correction to implement interpersonal coordination. At the implementational level they appear to be realized via cross-brain synchrony in homologous cortical regions. Framing both domains within Marr's hierarchy thus provides a solid standpoint for researchers to compare them rigorously—not only in terms of *where* they occur in the brain, but *why* and *how* the brain orchestrates them.

Critically, this kind of cross-domain theoretical mapping is not merely of academic interest. It offers the potential to develop general models of social cognition (Frith & Frith, 2012), applicable to diverse contexts—from musical performance to therapy, from teaching to conversation. By understanding how neural systems support the emergence of coordinated behavior in one domain, it may be possible to predict or influence another. Furthermore, understanding the interaction between domains (e.g., how musical improvisation may enhance verbal synchrony in conversation) opens novel paths for experimental intervention, especially with developmental or clinical populations.

1.3. Marr's levels of analysis

Understanding complex cognitive systems requires more than cataloguing their neural correlates or behavioral outputs. It demands a layered explanation—one that accounts not just for *where* in the brain things happen, but *what* problems the system is designed to solve, and *how* it solves them. David Marr, a foundational figure in cognitive science, addressed this challenge in his seminal work *Vision* (1982), where he introduced a three-level framework for analyzing information-processing systems: the computational, algorithmic, and implementational levels. Though originally developed for the visual system, Marr's hierarchy has since become a touchstone for theorizing across all domains of cognition, including language (Poeppel, 2012), music (Koelsch, 2011; Patel, 2008), and more recently, interpersonal interaction (Frith & Frith, 2012; Redcay & Schilbach, 2019).

At the computational level, one asks: *What is the goal of the system? What problem is it solving, and why is that problem important?* This level is concerned with defining the nature of the task and the logic behind it. In the context of interpersonal synchrony, this could involve the problem of establishing shared temporal expectations or maintaining coordinated turn-taking in real time. The computational level helps identify the functional rationale behind a cognitive operation—whether it be recognizing a face, interpreting a sentence, or aligning with a musical partner's timing. Crucially, this level is independent of the particular neural or algorithmic details used to implement the solution.

The algorithmic level deals with the *procedures and representations* that transform input into output. Here, the focus is on the format in which information is encoded and the step-by-step operations used to manipulate it. In musical or conversational synchrony, for instance, this might involve internal generative models of a partner's behavior, predictive simulations using phase-aligned oscillations, or mechanisms for computing prediction errors. Representations at this level may include rhythmic schemas, hierarchical event structures, or temporal binding through cross-modal entrainment. The algorithmic level often maps onto concepts from computational neuroscience (e.g., predictive coding), offering hypotheses about how information is processed dynamically.

Finally, the implementational level asks: *How are the algorithms physically realized in the brain?* This involves the anatomical, electrophysiological, and biophysical substrates of cognition—neurons, circuits, oscillations, neurotransmitters. In the case of INS, implementational-level data come from EEG, fNIRS, MEG, or fMRI hyperscanning studies that show cross-brain coherence in frontal, parietal, and temporal regions during joint tasks (Czeszumski et al., 2020; Dumas et al., 2010; Lindenberger et al., 2009; Cui et al., 2012; Liu et al., 2016). Synchrony in the theta (4–7 Hz) band between medial prefrontal cortices, or beta synchrony between homologous motor areas, are examples of phenomena located at this level. This level also accounts for task constraints (e.g., noise, latency, metabolic cost) and implementation bottlenecks, which may shape how abstract computations are realized in practice. This final level in Marr's framework, which is directly related to hyperscanning as a technique, carries important methodological caveats. One perhaps immediately worth mentioning is that EEG, fNIRS, MEG, and fMRI—all non-invasive modalities—provide only indirect measures of implementational mechanisms. To illustrate, while the spatial relation between an EEG electrode and its cortical source is partly determined by proximity, it is equally influenced by the orientation of the underlying dipole and by volume conduction through intervening tissues (Nunez & Srinivasan, 2006). Therefore, although it may be tempting to associate an electrode's signal with activity in the brain area directly beneath it, such assumptions require careful qualification. Since this article aims to support the design of robust future hyperscanning studies, these and related caveats will be discussed in greater detail in the following section.

These limitations underscore the importance of theoretical scaffolding when interpreting neural data. Marr's framework further cautioned that these three levels are logically distinct but interdependent. A complete theory of any cognitive function must ideally specify *what* it does, *how* it does it, and *what neural machinery enables that process*. Importantly, Marr cautioned against the illusion that a complete description at one level (e.g., a high-resolution brain map) automatically implies understanding at the others. As he famously wrote:

"Even if one already has a theory at the neural level... it is still necessary to characterize the nature of the information-processing task being solved, or one will not be able to understand why the system is doing what it is doing" (Marr, 1982, p. 27).

This warning is particularly relevant in the age of large-scale neuroimaging and hyperscanning. Without a strong computational-level hypothesis—what problem inter-brain synchrony is helping to solve—one risks collecting rich datasets with impoverished theoretical grounding (Hamilton, 2021). Likewise, without algorithmic clarity, one cannot distinguish whether a given pattern of synchrony reflects mirroring, prediction, error correction, or something else entirely (Mu et al., 2018). In the context of a novel area of research such as the hyperscanning of interaction, Marr's levels provide a critical scaffold for integrating findings across domains (music and speech) and measurement techniques (EEG, fNIRS, etc.), offering a structured way to link high-level functions like co-regulation and coordination with their underlying neural dynamics. By asking not only *where* synchrony happens but also *why* and *how*, researchers can move from descriptive neuroscience (Yuste, 2015) toward mechanistic explanation (Craver, 2007)—an essential step in both basic science and applied domains like social neuroscience and music cognition.

1.4. Methodological caveats

1.4.1. Sensor ≠ Brain Area

As previously mentioned, a fundamental methodological caveat in hyperscanning research is that activity recorded at a single EEG electrode or MEG sensor cannot be confidently attributed to the cortical region directly beneath it (Nunez & Srinivasan, 2006). Consequently, inferences from scalp electrodes to localized cortical generators are not straightforward; the location of a signal peak cannot be assumed to indicate the location of a source in the nearest patch of cortex. Source estimation techniques, such as beamforming or distributed inverse solutions, exist to address this issue, but these approaches each make different assumptions and cannot be fully validated in the absence of ground truth data (Grech et al., 2008). This is not an issue for haemodynamic methods such as fNIRS, but the low temporal resolution (and often limited spatial scope) of such methods can result in other inferential limitations, as discussed below.

This problem is often framed as one of spatial resolution: fMRI and fNIRS famously have high spatial resolution, MEG somewhat lower, and EEG much lower still. The truth, however, is that spatial resolution in M/EEG is fundamentally undefined. An infinite number of possible source configurations could have produced the signals one has recorded on M/EEG sensors, and it is impossible to say which are more likely than others without making many assumptions. That is not to say that source estimation is hopeless and shouldn't be attempted. Rather, it is to say that one *must at least attempt it* before making any conclusions about the neural sources of M/EEG signals, but then one must then acknowledge that the results are heavily dependent on the assumptions made by their model of choice, and ideally one would also use principled means to decide upon a model (e.g., model comparison).

Rather than treating sensor-level peaks as evidence for specific cortical loci, researchers should acknowledge the indirect nature of such measures. As Hamilton (2020) has emphasized, oversimplified mappings from electrodes to brain regions risk reinforcing purely correlational interpretations of interbrain synchrony. A more conservative interpretation is to use spatial patterns across sensors, especially in the context of experimental manipulations, as tentative indicators of differential recruitment of cortical resources. For example, if one condition yields a centrally maximal topography while a more demanding variant yields a frontal shift, it is reasonable—though still speculative—to infer that additional frontal resources were engaged. However, such inferences rely on contrasts between conditions, not on the absolute location of a single peak.

1.4.2. Brain Activity ≠ Cognitive Process

Another interpretive risk concerns the temptation to assume a transparent relationship between brain regional activity and cognitive processes. Such an assumption risks encouraging reverse inference: inferring cognitive processes from patterns of brain activation (Poldrack, 2006; Henson, 2006). Reverse inferences are notoriously weak, particularly when based on small-scale, surface-level hyperscanning data. By contrast, forward inference, in which researchers manipulate elements of a task or stimuli and observe the corresponding differences in brain response between conditions, yields stronger conclusions about the neural substrates of specific computations (Poldrack, 2011). Many of the reviewed hyperscanning studies appropriately employ forward inference by contrasting conditions that differ in interactional demands, such as joint versus solo tasks (e.g., Kurihara et al., 2022; Lender et al., 2023).

A Marr-inspired framework would be well served by placing cognitive processes and task demands at the forefront (computational and algorithmic levels), before situating regional activations or interbrain synchrony patterns at the implementational level. This process-first organization would not only avoid the pitfalls of reverse inference but would also align with the broader aim of systematizing hyperscanning evidence across domains of music and language (Hamilton, 2020).

1.4.3. Brain Region ≠ Mechanism

Closely related to the above is the distinction between "where" neural activity occurs and "how" a mechanism is instantiated. Knowing that a task engages the inferior frontal gyrus (IFG) simultaneously in two interacting participants, for example, provides limited mechanistic insight without further specification of how the IFG contributes to computations such as turn-taking or error monitoring. Oscillatory dynamics, by contrast, come closer to an implementation-level explanation. Synchronization in specific frequency bands—e.g., theta oscillations for predictive processing or beta (13–30 Hz) oscillations for motor coupling—can be directly related to interactional mechanisms (Mu et al., 2018; Kawasaki et al., 2013; Nguyen et al., 2023). While oscillatory measures obtained from EEG or MEG are themselves indirect, since sensors outside the head integrate activity from many populations of neurons, they provide a more precise bridge between algorithmic-level processes and their neural implementation than mere regional localization using e.g., fNIRS.

This problem relates directly to temporal resolution, of both the technology (sampling rate) and the phenomenon being measured (electromagnetic fields or haemodynamic responses). It is important to separate the two, since technology may provide ever-increasing sampling rates, but beyond a certain point this will provide no further information if the signal being sampled is very slow. The haemodynamic response is famously sluggish; a burst of neural activity on the order of milliseconds results in a haemodynamic response that peaks around 6 seconds later and resets a further 20 or 30 seconds later. This response can therefore be thought of as a low-pass filter on the neural signal, with most of the passband being below 0.1Hz (10-second cycle). M/EEG, on the other hand, can measure from direct current (DC) to 100 Hz or higher (particularly for MEG, for which volume conduction is less of a problem). Indeed, M/EEG analysis often begins with a *high*-pass filter in the neighbourhood of 0.1Hz. This means that, for all intents and purposes, inter-brain coherence in fNIRS (<0.1Hz) and M/EEG (>0.1Hz) are measuring *completely different phenomena*. An exception here may be envelope correlations in M/EEG, where the amplitude envelope of a signal (e.g., theta power) may rise and fall at much lower frequencies, and synchronization of such slow modulations of higher frequency power between individuals may track similar phenomena as haemodynamic methods.

Accordingly, oscillatory synchrony should be foregrounded at the implementational level, with regional attributions framed as informed but speculative complements. For example, theta synchrony in IFG may serve different interactional functions than theta synchrony in parietal cortex, but the key mechanistic claim is the role of theta synchrony per se. Embedding regional speculations within algorithmic-level accounts, rather than treating them as primary evidence, would avoid conflating anatomical labels with mechanistic explanations.

1.4.4. Common activity ≠ Coordinated activity

A further conceptual challenge concerns the distinction between processes that are generic and domain-specific (e.g., semantic processing in language, motor coordination in music) and those that are specifically required for interpersonal interaction (e.g., prediction of a partner's actions, maintenance of shared working memory). Both types of processes appear in hyperscanning studies, but they are not equally mechanistic in terms of explaining interactive behaviours. Evidence for domain-generic processes demonstrates the expected neural recruitment for task performance, vis-à-vis single-participant studies, while evidence for interaction-specific processes provides more direct support for the added value of hyperscanning (Sun et al., 2024; Jiang et al., 2016).

Making this distinction explicit is critical for claims at Marr's implementational level. For instance, solo rehearsal of a musical passage and joint performance will both recruit motor coordination networks, but the joint context additionally recruits mechanisms for temporal alignment and partner prediction (Ramírez-Moreno et al., 2023).

2. Hyperscanning during musical interaction

2.1. Foundations and early work

The use of hyperscanning to investigate joint musical interaction has opened a new frontier in the neuroscience of social behavior. Early research in this area demonstrated that the neural activity of two individuals can become temporally aligned—not merely in response to shared sensory input, but as a function of their active engagement in coordinated musical behavior. Music, with its inherently social and temporally dynamic nature, provides an ideal testbed for exploring how brains align during real-time interaction. Unlike linguistic dialogue, where semantic content and syntactic rules heavily structure exchange (Jackendoff, 2002; Levinson, 2016), musical collaboration often unfolds within looser frameworks, driven by shared rhythmic, melodic, and affective dynamics. This allows researchers to investigate core aspects of social coordination—such as joint timing, prediction, adaptation, and mutual attention—without the confound of propositional content.

The foundational work of Lindenberger and collaborators (2009) represents one of the first empirical demonstrations of INS during music performance using EEG hyperscanning. In their study, dyads of guitarists engaged in a prepared duet while their neural activity was simultaneously recorded from 16 electrodes placed according to the standard 10–20 EEG system. The analysis focused on two synchrony metrics: inter-brain phase coherence (IPC) and phase locking value (PLV). The authors found that INS emerged robustly during both the preparatory phase and the active performance, with the strongest synchrony localized over frontal and central electrodes, particularly in the alpha (8–12 Hz) and theta 4–7 Hz bands. Notably, this synchrony was not confined to stimulus onset or rhythm matching; rather, it persisted across time windows and appeared to reflect ongoing coordination efforts.

Although conforming to a rather stereotyped and overlearned model, allocating clear roles to each performer and thus somewhat limited in its generalisability, Lindenberger's study established several key principles that have shaped subsequent research. First, it provided clear evidence that joint musical behavior can drive spontaneous, sustained INS in the absence of artificial constraints or external pacing cues. Second, the authors inferred that frontal and central areas can be particularly sensitive to dyadic coordination, likely due to their role in motor planning, action simulation, and social cognition (though, as discussed earlier, such regional inferences remain tentative at the sensor level). Third, the study set a precedent for using low-frequency oscillations (especially in the theta band) as signatures of interpersonal engagement. These early insights laid the groundwork for later investigations that would expand the frequency range and anatomical scope of synchrony research in music.

Following Lindenberger's study, a broader methodological infrastructure began to form. Researchers increasingly adopted hyperscanning protocols using EEG and fNIRS —the former for its high temporal resolution, and the latter for its portability and better spatial specificity. Tasks evolved from tightly scripted duet performances to more naturalistic joint improvisations, tapping, and ensemble play. Analysis techniques also became more refined, incorporating time–frequency decomposition, wavelet coherence, and cross-brain connectivity metrics. Crucially, the focus of research began to shift from merely documenting synchrony to interpreting its functional role. Was INS a marker of behavioral coordination, a predictor of performance success, or a byproduct of shared environmental input? These questions would animate the next wave of studies, particularly those exploring frequency-specific INS during different types of musical interaction.

The turn to more generalisable and temporally fluid tasks—such as musical improvisation—prompted researchers to examine how synchrony unfolds in less constrained contexts. Rather than aligning to a fixed score or metronome, participants had to listen, predict, and adapt in real time, creating an emergent structure through mutual responsiveness. In such settings, INS was no longer just a correlate of externally driven behavior, but a possible mechanism of co-regulation, enabling individuals to build and maintain shared internal models. This hypothesis—tentatively suggested in early work (Dumas et al., 2010)—would become a central theoretical thread in the field.

Thus, the foundational work in hyperscanning and music-making established both a methodological paradigm and a theoretical challenge: how to move from empirical observations of INS to mechanistic models of joint cognition. The subsequent sections of this review build upon this early work, incorporating task manipulations, frequency-specific analyses, and region-level synthesis to construct a richer picture of how interpersonal synchrony is instantiated during musical interaction.

2.2. Dynamics of musical synchrony

The shift from scripted musical performance to joint improvisation introduced new complexity—and new scientific opportunity—into hyperscanning research. Unlike pre-composed duets, musical improvisation is inherently generative, adaptive, and socially contingent. It demands that performers engage not only in temporal synchronization but in real-time construction of shared meaning through musical gestures. From a neural perspective, improvisation is therefore a privileged context in which to examine how synchrony reflects, supports, and perhaps even enables emergent coordination.

Marr-level account of INS in music and conversation

This insight is clearly illustrated in the work of Müller et al. (2013), who investigated neural synchrony during dyadic guitar improvisation. Their findings refined the earlier results of Lindenberger et al. (2009) by emphasizing the frequency-specific nature of INS and its dependence on the quality of coordination. Using EEG hyperscanning, Müller and colleagues found that inter-brain synchrony increased significantly in the delta (1–4 Hz) and theta bands during periods of smooth, coordinated co-improvisation—what they termed "harmonious" musical interaction. In their study, "harmonious" referred to moments when performers spontaneously aligned in timing, phrasing, and dynamics without external cues. These synchronized phases were characterized by mutual adaptation in timing, phrasing, and dynamics, suggesting that INS was tracking (or facilitating) the emergence of a shared temporal structure between performers.

Crucially, Müller et al. demonstrated that this synchrony was not uniformly distributed across the scalp. It was strongest over midline electrodes—notably Fz, Cz, and Pz—in the delta, and theta bands, suggesting the involvement of frontal executive regions, motor planning areas, and posterior integrative networks. However, and in keeping with the methodological cautions outlined earlier, the spatial relationship between an EEG electrode and its cortical source is not straightforward: while proximity plays a role, the recorded signal is also shaped by the orientation of the underlying neural dipoles (typically perpendicular to the cortical surface) and by volume conduction through the intervening tissues (Nunez & Srinivasan, 2006).

The midline distribution led the authors to suggest that participants were engaging both cognitive control (frontal) and sensorimotor simulation (central) in service of real-time coordination. Yet, as noted earlier, such interpretations should be regarded as tentative, since electrode-level topographies provide only indirect evidence of underlying cortical sources and associated cognitive operations (see methodological caveats above). Furthermore, delta and theta rhythms—traditionally associated with low-frequency entrainment, temporal prediction, and attention (Arnal & Giraud, 2012) —appeared to function as neural substrates for the construction of joint timing frameworks. These findings reinforced the idea that INS is not merely an epiphenomenon of behavior, but part of a neurodynamic system that enables interaction.

A further fine-grained view of musical interpersonal synchrony emerges from Ramírez-Moreno et al. (2023), who extended the investigation into higher frequency bands. In this study, musicians performed in small ensembles, and moments of coordinated entry, improvisational convergence, and synchronous phrasing were annotated and analyzed using bispectral synchrony measures. The results revealed that beta and gamma (30–50 Hz) band synchrony increased significantly during these coordinated segments, particularly across posterior and lateral electrodes, which the authors interpret as reflecting possible engagement of visual, auditory, and spatial processing networks; such inferences, however, remain indirect at the electrode level. These findings introduce two key insights. First, higher-frequency INS is not only observable in musical tasks but is functionally relevant: the authors interpreted beta and gamma activity as reflecting movement intention, action simulation, and multisensory integration. Second, the spatial distribution of synchrony—posterior and lateral rather than purely central—tentatively suggests engagement of visual, auditory, and spatial processing networks during complex coordination.

Together, the studies by Müller and Ramírez-Moreno reveal that musical synchrony unfolds across multiple temporal and anatomical scales. Delta and theta INS emerge during the foundation of shared temporal frameworks, enabling musicians to align their internal models. Beta and gamma INS appear to reflect more immediate sensorimotor coupling, including predictive motor planning and joint action simulation. From a functional standpoint, this multiband synchrony supports a model in which low-frequency oscillations scaffold the overarching timing structure, while higher frequencies fine-tune the execution and integration of gestures.

A key theoretical implication of these findings is a conceptual transition–from what can be interpreted as stimulus-driven to model-driven synchrony. In duet improvisation, there is no external pacing source; coordination must arise from the musicians themselves. This places greater cognitive demands on anticipation, error correction, and adaptive alignment, all of which are supported by neural systems engaged in prediction and control. The synchrony observed in such contexts is thus best interpreted as a signature of mutual prediction and dynamic coupling—a view compatible with predictive coding frameworks (Andersen et al., 2023; Friston & Frith, 2015) and theories of joint action (Sebanz et al., 2006), although current hyperscanning methods do not yet permit direct estimation of cross-brain prediction-error dynamics. Interpretations invoking predictive-coding should therefore be regarded as heuristic rather than formally model-based.

Moreover, both studies emphasize that synchrony is not binary—it does not merely occur or not—but is graded, dynamic, and sensitive to qualitative aspects of interaction. "Harmonious" improvisation (as previously defined), for instance, is marked by higher INS than asynchronous or awkward phases, even within the same dyad. This implies that neural synchrony is modulated by subjective and relational variables, such as flow (Chabin et al., 2022), rapport (Kinreich et al., 2017), or emotional congruence (Dikker et al., 2017).

Taken together, this body of work extends the original findings of Lindenberger et al. (2009) by showing that INS in musical interaction is shaped by the nature of the task, the quality of coordination, and the frequency and region of neural activity involved. More importantly, these results suggest that synchrony is not simply reflective of joint behavior, but may be dynamically coupled to, and in some respects constitutive of, the interactive process itself. In improvisation, especially, synchrony may serve as a neural signature of real-time coordination mechanism, dynamically linking two agents into a coherent, adaptive system.

2.3. Task manipulations and sensory–motor models

While foundational studies in musical hyperscanning have demonstrated the presence of inter-brain synchrony (INS) during joint performance, more recent work has adopted experimental manipulations to clarify the cognitive mechanisms and functional significance of this synchrony. Instead of simply observing coordination as it unfolds, these studies strategically vary contextual parametres to explore how different aspects of musical engagement—such as familiarity with a partner, the presence of rhythmic structure, or coordination difficulty—modulate INS. This line of inquiry advances the understanding of INS by linking it to task-specific demands and representational strategies used during real-time interaction. One of the most compelling examples of this approach is provided by Gugnowska et al. (2022), who manipulated familiarity with a partner's musical part in a duet piano task. Participants were asked to rehearse four Bach-based duets before undergoing EEG hyperscanning. For two of these duets, participants only had access to their own part, leaving them unfamiliar with the content of their partner's contributions. For the remaining two, both parts were provided in advance, enabling participants to anticipate their partner's actions. This within-subject design provided a direct contrast between conditions of high and low predictive certainty.

The results indicated that inter-brain synchrony was significantly higher in the unfamiliar condition, particularly in the gamma band, and that this synchrony peaked between 4.9 and 5.3 seconds after the trial onset. Importantly, this increase in INS was not attributable to differences in gamma-band power, but specifically to cross-brain coherence, suggesting that it was the relational dynamics between performers—not individual cognitive load—that drove the effect. Behaviorally, unfamiliarity was also associated with stronger mutual adaptation, as indicated by more negative lag-zero cross-correlations in timing. This behavioral adaptation can be viewed as a manifestation of mutual engagement—a dynamic state in which interacting partners continuously monitor, predict, and adjust to one another's actions to sustain a shared temporal and attentional focus (Schilbach et al., 2013; Redcay & Schilbach, 2019). The authors interpret this finding as evidence for a shift from internally rehearsed motor plans to externally guided auditory monitoring. When participants were unable to rely on pre-learned expectations, they turned instead to perceptual cues provided by their partner, leading to a stronger need for joint attention and moment-to-moment adjustment.

This study contributes two essential insights. First, it implicates gamma-band synchrony in musical adaptation, a frequency range typically associated with multimodal integration and local cortical binding. Second, it demonstrates that INS is not solely a byproduct of sensorimotor alignment, but is also shaped by cognitive strategies and interactional uncertainty. These findings suggest that synchrony may be recruited flexibly—or become more salient—in response to representational demands, particularly when internal predictive models are weakened or unreliable.

A related but distinct question concerns the role of musical structure, particularly rhythm, in modulating INS. This issue was addressed in a set of experiments by Hu et al. (2022), who investigated how the presence and strength of metrical accents influenced INS during a finger-tapping task. Participants performed synchronized tapping under three conditions: no metre (uniform tones), weak metre (minimally accented downbeats), and strong metre (clearly accented downbeats). The results showed that INS, measured using fNIRS, was significantly higher in the metreed conditions, and especially in the strong metre condition. This effect was recorded over scalp regions corresponding approximately to the left middle frontal cortex (lMFC), a region the authors interpret as supporting attention and working-memory functions. They propose that the lMFC may serve as a putative hub for constructing and updating temporal predictions. (MacDonald et al., 2000).

Interestingly, the correlation between INS and behavioral synchrony emerged only in the strong metre condition. This finding contrasts with Hu et al. (2022), where inter-brain synchrony was stronger under metrically uncertain conditions. However, these results are not necessarily contradictory: the underlying neural dynamics likely reflect distinct frequency regimes. Whereas Hu used fNIRS, capturing slow hemodynamic fluctuations below 0.1 Hz, the present EEG findings concern oscillatory coupling in the 30–40 Hz range—two fundamentally different temporal scales and physiological substrates of coupling. This suggests that clearer temporal structure does not merely facilitate behavioral alignment but may also enhance the precision of shared internal timing models. The authors argue that metre provides a scaffold for mutual prediction, enabling participants to maintain aligned expectations about when a partner's action will occur. The lMFC may serve as a hub for constructing and updating these temporal predictions. These findings point to the possibility that INS may reflect neural processes supporting the construction and maintenance of shared temporal frameworks during interaction. Given these radically different temporal regimes, the term inter-brain synchrony necessarily refers to distinct physiological substrates across modalities. Parallels drawn between fNIRS and EEG findings should therefore be interpreted at a functional rather than mechanistic level, acknowledging that what is labeled as 'synchrony' in each case reflects distinct neural and vascular processes operating at separate timescales.

Marr-level account of INS in music and conversation

Whereas the studies above examined how increased predictability affects synchrony, Kurihara et al. (2022) approached the problem from the opposite direction. Their goal was to explore how INS behaves under conditions of decreased behavioral stability. In their anti-phase tapping task, participants were instructed to tap alternately with their partner, producing a pattern in which one individual taps while the other pauses. This form of coordination is inherently unstable and typically leads to greater variability in timing. Contrary to what might be expected, the authors found that INS between the left temporal and central regions increased as coordination became more unstable. This effect was specific to the theta frequency band and was not observed in conditions with more stable performance. Notably, this pattern parallels the finding by Hu et al. (2022) of enhanced INS under metrically uncertain conditions, albeit at a much slower hemodynamic timescale. Together, these observations suggest that elevated synchrony may sometimes index compensatory engagement or predictive recalibration rather than fluent coordination, with the direction of the effect varying across neural frequency bands and measurement modalities. Although the authors do not mention it as such, this unstable coordination can be at least partially characterized as a form of repair. In the context of musical performance (be it solo or group performance), repair refers to the processes by which performers identify, manage, and resolve in real time problems, breakdowns, or deviations from the intended flow of the performance. The notion is borrowed from conversation analysis (e.g., Sacks, Schegloff & Jefferson, 1974), where repair designates the practices that speakers use to deal with trouble in speaking, hearing, or understanding.

These findings suggest that higher INS can arise not only from successful coordination but also from cognitive effort in the face of coordination failure and resulting repair. The increased theta synchrony likely reflects the deployment of attention and monitoring mechanisms, consistent with previous literature on the role of theta in cognitive control and error processing (Cavanagh & Frank, 2014; Cohen, 2011). From a regional perspective, synchrony between the left temporal lobe and central areas may indicate interaction between auditory integration and motor planning systems, supporting the idea that individuals were actively trying to adapt to their partner's unstable timing.

Taken together, these studies strongly suggest that INS is a highly flexible phenomenon, modulated by both environmental structure and internal strategies. Familiarity, rhythmic metre, and task difficulty each influence the degree and distribution of synchrony across brains, suggesting that inter-brain coupling probably reflects not merely sensorimotor coordination but the interplay of prediction, attention, and adaptation. In some cases, synchrony may be facilitated by external structure (as in the strong metre condition), while in others it appears to be recruited to cope with increased uncertainty or reduced predictability (as in the unfamiliar and anti-phase conditions). However, it remains difficult to determine whether these divergent patterns arise from genuine functional differences in how synchrony supports coordination, or simply from differences in task demands and measurement modalities. The fNIRS studies capture slow fluctuations (<0.1 Hz) associated with large-scale vascular coupling, whereas EEG-based measures reflect fast oscillatory synchrony in the theta and gamma bands. Consequently, apparent discrepancies in directionality—greater INS under predictability versus unpredictability—may reflect distinct neural processes sampled at different temporal scales rather than contradictory findings. Furthermore, given the typically small samples used in these hyperscanning studies (Müller (n=8 duets), Ramírez-Moreno is a case study), Kurihara (n=24), Hu (n=20), Gugnowska (n=22)), these findings should be regarded as preliminary and in need of replication

3. Hyperscanning during conversation

Research on INS during conversation has expanded rapidly over the past decade, driven by the increasing ecological sophistication of hyperscanning paradigms. In contrast to joint music-making, where coordination is largely rhythmic and often non-verbal, verbal conversation involves turn-taking, semantic exchange, and socially regulated timing mechanisms. Yet, despite these domain differences, studies consistently report that synchronous neural activity emerges during conversational interaction, with temporal and spatial properties comparable to those found in musical improvisation. These findings could mean that INS is a core neurocognitive mechanism supporting dynamic social coordination, and that it generalizes across both linguistic and non-linguistic modalities.

A foundational contribution in this area was made by Kawasaki et al. (2013), who demonstrated that conversational alignment at the level of speech rhythm was associated with increased theta-band phase synchrony in EEG recordings between dyad members. Using a rhythmic speech coordination task, they observed enhanced INS in fronto-temporal regions during successful rhythmic alignment compared to non-synchronized baselines. Specifically, increased inter-brain synchrony was found over frontal regions (F3, Fz, F4 — frontal midline and lateral) and temporal regions (T7, T8 — left and right temporal scalp regions, suggesting coordinated engagement of predictive speech planning and auditory-motor integration systems. These findings suggested that low-frequency neural coupling supports the temporal regulation of turn-taking, likely by aligning expectations about when a conversational partner is likely to begin or end their utterance. The theta-band findings paralleled those from musical coordination, reinforcing the idea that shared timing models are a general feature of cooperative interaction.

Building on early demonstrations of rhythmically mediated synchrony during speech coordination (Kawasaki et al., 2013), Ahn et al. (2018) used EEG hyperscanning to examine neural coupling during naturalistic turn-taking. They observed increased theta- and alpha-band phase synchronization between interlocutors' frontal and temporal cortices precisely at speaker–listener transition points, suggesting that inter-brain coupling may index the predictive alignment of turn boundaries. These findings support the view that conversational coordination depends on a shared temporal framework, dynamically updated as roles alternate. Together with more recent work using fNIRS and fMRI (e.g., Sun et al., 2024), this study underscores that INS is not invariable but evolves with the real-time contingencies of dialogue.

Subsequent studies expanded on this by exploring the social functions of neural synchrony in more naturalistic dialogue. Sun et al. (2024) used fMRI hyperscanning to examine spontaneous conversation in adult dyads and found that speaker–listener neural coupling emerged reliably in default mode network (DMN) and language areas, including the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), and bilateral temporal lobes. Notably, the strongest synchrony occurred when participants exhibited high mutual engagement (as previously defined), suggesting that INS reflects not merely linguistic decoding but the degree of interactive attunement. These findings are consistent with a model suggesting that conversational INS may reflect alignment between shared internal models —representing not only the surface structure of dialogue, but also partner intentions, emotions, and attention —consistent with the notion of mutual engagement as a dynamic, reciprocal alignment of mental states (Schilbach et al., 2013; Redcay & Schilbach, 2019).

Nguyen and collaborators (2023) extended these findings developmentally, showing that neural synchrony emerges even in preverbal mother–infant interactions, and that it correlates with proto-conversational turn-taking. Using fNIRS hyperscanning, they demonstrated increased INS recorded over regions consistent with the medial prefrontal cortex when infants and mothers engaged in face-to-face vocal exchange. This synchrony was strongest during the early moments of interaction and was associated with more frequent turns and mature cortical responses in the infant brain. These data suggest that neural coupling supports the foundational architecture of social communication, even before the acquisition of formal language. Importantly, this early synchrony may scaffold the development of later language skills and social cognition.

In the adult domain, Jiang et al. (2016) investigated how leader–follower dynamics emerge in cooperative tasks. Using fNIRS during a joint decision-making game, they found that dyads in which one participant visibly assumed a leadership role in terms of behavioral dominance (i.e., decision frequency, response timing) displayed stronger inter-brain coherence between the leader's and follower's dlPFC and premotor regions; however, given the spatial limitations of fNIRS, these regional designations should be interpreted with caution (see methodological caveats above). These patterns were predictive of behavioral dominance and initiative, and occurred prior to overt decision-making. The authors proposed that INS in these regions reflects the anticipatory modeling of partner behavior, supporting social hierarchy and coordination. This complements findings from musical interaction, where increased dlPFC activity has been associated with mutual adaptation and cognitive load during improvisation (Abiru et al., 2016; Tachibana et al., 2019), and where inter-brain synchrony in prefrontal regions—including the dlPFC—has been observed during coordinated ensemble play and improvisational convergence (Ramírez-Moreno et al., 2023)

Another key contribution comes from Liu et al. (2017), who used fNIRS to compare neural synchrony during turn-based cooperation and competition. They found that cooperative conditions elicited significantly stronger INS over regions corresponding approximately to the bilateral inferior frontal gyrus (IFG), posterior superior temporal sulcus (pSTS), and inferior parietal lobule (IPL). Competitive conditions, by contrast, showed reduced synchrony or even desynchronization in the same areas. These spatial patterns should of course be viewed as approximate indicators of cortical engagement rather than precise localizations, given the resolution constraints of fNIRS. These results underscore the sensitivity of INS to social context, and support the idea that synchrony reflects a dynamic cognitive state, not just sensory overlap or shared motor planning. The frontal and parietal regions identified here are also known to support mentalizing, joint attention, and action prediction, reinforcing their role in cooperative cognition.

Mu et al. (2018) provided a comprehensive review of such studies, noting that the most consistent findings across conversational hyperscanning involve frontal and parietal regions, and temporo-parietal junction (TPJ) areas, modulated by task structure and communicative intent. They argue that inter-brain coherence is strongest when individuals share a common frame of reference, whether through mutual goals, temporal predictability, or shared affect. This also suggests that conversational partners who are already familiar with each other can rely on this familiarity to explore a broader range of conversational topics (and thus diverge in terms of INS), rather than focusing primarily on convergence (Speer et al., 2024). Crucially, Mu and collaborators emphasize that different frequency bands may index different components of interaction: delta and theta for timing and prediction, alpha for attentional alignment, and beta for sensorimotor simulation of speech.

Hirsch et al. (2017) and Wass et al. (2020) contribute additional depth to these findings by emphasizing the role of alignment in developmental and affective contexts. Hirsch et al. used fNIRS to show that conversation between adults is marked by INS in inferior frontal and temporal regions, particularly when speech is accompanied by mutual gaze. Wass et al., focusing on infants, found that attentional synchrony between parent and child predicted subsequent neural alignment, highlighting a bidirectional relationship between behavioral coordination and INS. These studies converge on the idea that interpersonal synchrony is not static, but dynamically regulated by attention, intention, and affective engagement.

Taken together, this body of work establishes a robust empirical foundation for the study of conversational neural coupling. Across paradigms and age groups, findings converge on the involvement of a distributed frontotemporal network, modulated by task structure, social role, and affective engagement. The parallels with musical interaction are substantial: both domains engage similar frequency bands, rely on predictive timing, and exhibit modulation of INS by role, familiarity, and context. These findings point toward a shared computational architecture for interpersonal alignment, instantiated through oscillatory coupling and dynamically updated internal models.

In light of this evidence, it becomes increasingly plausible to view conversation not merely as a linguistic process but as a temporally extended, mutually regulated joint action, possibly underpinned by the same neural systems that support coordination in other domains such as music, dance, and cooperative movement. This theoretical shift lays the groundwork for a Marr-level mapping of neural synchrony across communicative modalities, as explored in the next section.

4. Marr-level mapping of INS across domains

The empirical convergence observed in hyperscanning studies across joint music-making and conversation raises a pressing theoretical question: what is the cognitive and neural function of INS? The consistent involvement of frontal, temporal and parietal networks and cross-brain oscillatory coupling during dyadic interaction suggests that INS is not domain-specific, but rather supports a cross-domain cognitive operation. Marr's (1982) framework offers a principled scaffold for interpreting these phenomena. When applied to the data reviewed above, Marr's levels help shift the focus from merely observing where and when synchrony occurs to understanding why it arises, how it functions cognitively, and how it is physically realized in the brain.

Table 1 summarizes the effort in this section to tentatively map the evidence reviewed in the preceding sections into an integrated, panoramic view. In line with the earlier discussion of the advantages of forward—as opposed to reverse—inference (Section 1.4.2), the computational level is placed in the leftmost column, with the subsequent levels progressing toward the right. The table (and the remainder of this section) also includes brain regions that can, at present, only be putatively associated with the elements described above. By "putative," we mean regions that are most likely to be involved but which, given the caveats outlined earlier, cannot yet be regarded as firmly established in scientific terms. Bearing in mind yet another caveat, this section also takes care to distinguish between self-focused and interaction-specific processes. As a general approach, literature at the individual level (self-focused) is first presented, followed by its relation to INS phenomena (interaction-specific processes).

At the computational level, the relevant question here is what problem INS helps to solve. Across musical and conversational contexts, this problem can be framed as that of dynamic interpersonal alignment: the need to establish and maintain a shared temporal, intentional, and attentional space between interacting individuals. As previously reviewed (section 2), in joint musical improvisation, participants must align on rhythm, phrasing, and affect without relying on a fixed external structure. In conversation, interlocutors must regulate turn-taking, co-construct meaning, and maintain semantic and pragmatic coherence in real time. Both domains require each participant to continuously anticipate, adapt to, and influence the behavior of the other. The neural synchrony observed during such interactions is thus best understood as supporting the real-time coordination of joint behavior under uncertainty.

Marr level			Putative brain region	Representative references
Computational	Algorithmic	Implementational		
Maintain flexible coordination strategies, support adaptive control of interaction under role pressure or cognitive load.	Enable task-switching, monitoring, and working memory updates; construct role-based internal models of interaction.	Observed associations via cross-brain theta and beta synchrony	Dorsolateral Prefrontal Cortex (dlPFC)	Jiang et al., 2016; Abiru et al., 2016; Tachibana et al., 2019
Support mutual prediction of verbal and non-verbal actions during joint coordination.	Simulate partner's speech or action gestures; correct mismatch via error-driven adjustments.	Couples across brains in beta/gamma bands during action–perception alignment in EEG, and IFG in fNIRS studies.	Inferior Frontal Gyrus (IFG)	Liu et al., 2017; Hirsch et al., 2017; Yang et al., 2020
Facilitate alignment of affect, intention, and shared attention across individuals.	Represent mental states and monitor joint engagement; integrate social-affective cues.	INS emerges especially in fNIRS and fMRI studies; activated during turn-taking and emotional resonance.	Medial Prefrontal Cortex (mPFC)	Cui et al., 2012; Sun et al., 2024; Nguyen et al., 2023
Coordinate auditory timing and parsing to align interactional rhythm.	Process prosody, syllable structure, and musical beat; synchronize auditory models.	Phase-locked in delta/theta ranges; key site of auditory-driven synchrony across modalities.	Superior Temporal Gyrus / Sulcus (STG/STS)	Kawasaki et al., 2013; Mu et al., 2018; Gugnowska et al., 2022
Execute temporally precise joint action (e.g., tapping, playing, speaking).	Simulate and predict motor timing of partner's actions; drive synchronization.	Beta-band IBS tracks motor coupling; strong signals in EEG studies of music and finger tapping.	Central Motor Areas	Kurihara et al., 2022; Ramírez-Moreno et al., 2023; Hirsch et al., 2017
Support joint spatial attention and integration of shared context.	Track partner's position, goal, or gesture; aid coordination in space and time.	Synchrony observed in EEG/fNIRS during cooperative interaction; modulated by social context.	Inferior Parietal Lobule (IPL)	Liu et al., 2017; Jiang et al., 2016

Table 1. Marr-level tentative mapping of music-making and conversation hyperscanning

This coordination is not trivial. In naturalistic social contexts, participants must manage noise, delays, and partial information, while maintaining communicative efficiency and relational cohesion. INS may contribute to reducing this complexity by reflecting the formation of shared predictive models of the ongoing interaction. As seen in studies of joint tapping, unfamiliar duets, and spontaneous conversation, synchrony often increases when prediction is more difficult or when external cues are absent (Gugnowska et al., 2022; Kurihara et al., 2022; Sun et al., 2024). In this sense, the presence of INS seems to reflect not just successful coordination but the active process of constructing and updating shared internal representations, allowing two individuals to jointly solve the problem of mutual alignment.At the algorithmic level, the task is to specify the cognitive representations and processes that enable this alignment. One such mechanism is the maintenance of internal generative models. These are forward models that simulate the likely timing, content, or structure of a partner's action. In musical performance, such models anticipate a collaborator's rhythmic or harmonic choices; in conversation, they predict turn completion, semantic content, or emotional tone. The brain likely uses these models to anticipate the partner's next move and adjust its own behavior accordingly. Evidence for such predictive processing can be found in Kurihara and collaborators' work (2022): theta-band INS increased under unstable coordination, strongly suggesting that synchrony was not a marker of ease but of increased cognitive effort toward prediction and control.

Closely related is the process of mutual adaptation, which involves rapid, bidirectional updating of behavior in response to prediction errors. Gugnowska (2022) observed stronger gamma-band INS when pianists were unfamiliar with their partner's part—precisely when internal models were weak and adaptation had to be driven by perceptual cues. This suggests that synchrony can function as a mechanism of repair (joint error correction), dynamically aligning representations across agents. In conversation, similar patterns emerge, as INS related to medial prefrontal and temporal regions appears to covary with behavioral manifestations of mutual engagement, implying that shared attentional focus and social attunement are supported by continuous alignment processes (Sun et al., 2024).

Another crucial algorithmic process is covert motor simulation, shown to be supported by mirror neuron systems and sensorimotor feedback loops at the individual level. Studies of musical improvisation suggest that performers internally simulate a partner's phrasing or rhythm to anticipate timing and respond in kind. In conversation, this takes the form of simulating articulatory gestures or turn transitions. The teams of Liu (2017) and Jiang (2016) showed that INS in IFG and premotor cortex was modulated by cooperative dynamics and role differentiation, respectively, underscoring the involvement of action–perception systems in managing social contingencies. These regions enable participants not just to monitor a partner's behavior, but to simulate and adjust to it proactively, making them essential to the algorithmic substrate of INS.

At the implementational level, the question is how these computations and algorithms are realized in the brain's physical dynamics—namely, in specific neural circuits, anatomical structures, and electrophysiological mechanisms. This includes not just the localization of functions to cortical regions being synchronized across interactants, but also the biophysical processes—like cross-brain phase coupling or frequency-specific coherence—that instantiate inter-individual coordination. Neural oscillations provide a natural temporal structure for coding the likelihood of sensory or motor events. Delta and theta rhythms support temporal prediction, rhythmic entrainment, and attentional gating. Beta rhythms are closely tied to motor planning, timing execution, and internal simulation of movement, while gamma rhythms support multimodal integration and feature binding. Importantly, these rhythms are not only internally synchronized but also sometimes interpersonally phase-locked, suggesting that individuals may be coupling their internal timing systems to facilitate shared predictions. This mechanism has been observed in both joint music-making (e.g., Müller et al., 2013; Ramírez-Moreno et al., 2023) and speech (Kawasaki et al., 2013; Sun et al., 2024; Nguyen et al., 2023; Jiang et al., 2016), often tracking moments of high coordination or interactional salience.

Marr-level account of INS in music and conversation

The studies reviewed here consistently point to a distributed set of cortical regions that form a domain-general social alignment network, engaged across joint music-making and conversation though, as previously noted, EEG- and fNIRS-based measures of inter-brain coupling capture different physiological processes). It is also perhaps worth reiterating that the regional associations discussed below are necessarily putative: given the spatial and inverse-problem limitations of current hyperscanning methods, these attributions should be interpreted as heuristic guides to functional organization rather than definitive localizations. The dorsolateral prefrontal cortex (dlPFC) has been, at the individual level, classically associated with executive functions such as task management, behavioral inhibition, and working memory. In the context of dyadic interaction, however, its relevance lies not merely in supporting these functions within each individual, but in enabling their alignment across individuals. Hyperscanning studies thus suggest that dlPFC engagement becomes critical when interlocutors or musical partners must coordinate self-generated output with partner-dependent input, for instance during turn-taking, improvisation, or the emergence of leadership roles (Jiang et al., 2016; Abiru et al., 2016; Tachibana et al., 2019). Thus, while intra-individual executive functions are a necessary substrate, the added value of hyperscanning is to demonstrate that these functions are jointly and dynamically synchronized across brains in order to sustain effective communication and joint musical performance.

The inferior frontal gyrus (IFG), long associated with speech production and motor mirroring, supports action–perception coupling and prediction error correction in both speech and music. These functions have been extensively studied at the individual level, but recent hyperscanning studies suggest that temporally aligned recruitment of these mechanisms across individuals contributes to IFG-based inter-brain synchrony (e.g., Yang et al., 2020; Liu et al., 2017). Similarly, the medial prefrontal cortex (mPFC), classically implicated in shared attention, affective attunement, and theory of mind in single-brain studies, has also shown consistent INS during parent–infant and adult–adult interaction (Nguyen et al., 2023; Sun et al., 2024), suggesting a functional role in inter-individual social alignment.

Auditory and temporal regions—particularly the superior temporal gyrus (STG) and superior temporal sulcus (STS)—are well-established in individual-level research as supporting the tracking and segmentation of acoustic input in both musical and linguistic contexts (e.g., phonological processing, beat detection). In the hyperscanning literature, these regions have also shown inter-brain synchrony (INS), especially in delta and theta bands, during tasks involving speech rhythm coordination (Kawasaki et al., 2013) and joint musical phrasing (Müller et al., 2013). Central motor areas, including the primary motor cortex and supplementary motor areas (SMAs), are known from single-brain studies to govern motor timing and action simulation. Hyperscanning evidence confirms that these same regions frequently exhibit INS during both musical and conversational interaction—particularly in the beta band—suggesting a shared role in coordinating the timing and execution of joint action (Kurihara et al., 2022; Ramírez-Moreno et al., 2023). Parietal regions, including the inferior parietal lobule (IPL), are implicated at the individual level in spatial alignment, attentional orienting, and the integration of multisensory cues. Correspondingly, fNIRS and EEG hyperscanning studies have reported INS in parietal areas during cooperative and competitive interaction, supporting a role in distributed attentional coordination (Liu et al., 2017). Across all three regions, hyperscanning studies consistently report frequency-specific patterns of INS that appear functionally relevant: delta band for beat-level alignment and turn structure, theta for cognitive control and adaptation, beta for timing execution and motor simulation, and gamma for high-level integration and cognitive effort under uncertainty. These patterns are not only anatomically recurrent but also sensitive to interactional variables such as role differentiation (e.g., leader vs. follower), familiarity, and task difficulty, as demonstrated across multiple studies reviewed here (e.g., Gugnowska et al., 2022; Jiang et al., 2016).

In conclusion, Marr's framework allows the characterization of INS as a multi-level cognitive mechanism. At the computational level, it seems to solve the problem of dynamic interpersonal alignment. At the algorithmic level, it would operate through internal generative models, mutual adaptation, and covert simulation. At the implementational level, it appears to be expressed through task-sensitive, frequency-specific coupling between functionally specialized brain regions.

5. Final remarks and future directions

The present synthesis depicts an informed panoramic view of how interpersonal neural synchrony (INS), as observed in both music and language domains, could support a general-purpose mechanism for dynamic social coordination. Cross-brain phase alignment of cross-domain cortical circuits seem to be flexibly recruited according to task demands. Some may be left with the impression that this physical implementation reflects—not merely corresponds to—the algorithmic and computational properties identified above. That is, the system appears to be organized in a way that supports predictive, adaptive, and mutually regulating behavior in real time, regardless of whether the medium is music, speech, or another social modality. Although such a claim can be indeed envisaged, only a much larger corpus would allow to provide evidence supporting or refuting such a notion. Indeed, as previously mentioned, research on INS in musical interaction remains limited, and studies directly comparing INS across music and language are scarcer still. This relatively small body of work can and must, of course, be systematically related to the broader hyperscanning literature across further domains (perhaps beyond interaction, be it musical or linguistic), as well as to the much larger body of evidence on brain areas and their interconnections (Mu et al., 2018; Hamilton, 2020). While such an integrative effort lies beyond the scope of the present article, the analysis offered here is intended to contribute toward that direction. For instance, future studies should directly compare the same dyads across musical and conversational tasks using equivalent formats (e.g., turn-taking, improvisation), enabling within-subject contrasts of INS signatures. Hybrid paradigms such as musical storytelling or chant-based dialogue could also probe shared mechanisms along a continuum. More broadly, multi-site meta-analytic efforts and harmonized data analysis pipelines may allow for more systematic cross-domain comparison, potentially revealing whether the same cortical networks—especially IFG, mPFC, and STG—support joint prediction and adaptation across modalities. Multimodal approaches (e.g., combined EEG–fNIRS studies) could help reconcile discrepancies across modalities and, ideally, yield unified spatiotemporal models linking regional hemodynamic activity with frequency band–specific oscillatory synchrony.

Skeptics may legitimately argue that synchrony arises simply because interlocutors or performers are exposed to similar multisensory input—seeing and hearing one another—rather than because of genuine alignment of neural processes. This raises the question of whether the presence of INS in both conversational and musical contexts reflects shared underlying mechanisms for joint coordination, or whether it can be fully explained by superficial similarities in the jointly-experienced sensory stimuli generated during interaction. Addressing this challenge requires carefully designed experimental contrasts that separate mere stimulus-driven entrainment from interaction-specific alignment, thereby clarifying whether INS indexes domain-general mechanisms of social coordination or merely shared perceptual input. As Hamilton (2021) has emphasized, adopting a General Linear Model (GLM) framework offers a principled way to achieve this dissociation. By explicitly modelling shared stimulus regressors alongside interaction-specific predictors (e.g., partner behavior or mutual contingency), GLM-based analyses can statistically isolate neural coupling arising from genuine social interaction rather than from common perceptual input. To disentangle social alignment from stimulus-driven entrainment, researchers should implement control conditions where participants receive identical auditory and visual input without engaging in mutual interaction—for instance, by presenting pre-recorded material or measuring across non-interacting dyads. Further approaches could involve brief disruptions of contingent feedback (e.g., jittered audio or visual occlusion) to test whether INS tracks behavioral co-regulation beyond co-perception. Causal perturbation methods, such as transcranial stimulation or experimentally manipulated turn-taking roles, could help clarify whether observed synchrony arises from active mutual prediction or is merely epiphenomenal.

From this common foundation, several further empirically tractable predictions emerge, each corresponding to a specific Marr-level function. Future research can use these predictions to probe whether the neural mechanisms observed may prove to be truly domain-general or shaped by task-specific constraints. Taking into account the methodological caveats discussed above, such predictions can still be pursued prudently, provided that conclusions remain appropriately qualified and grounded in careful experimental design.

One immediate direction concerns the computational role of INS in role asymmetry. Hyperscanning studies suggest that leaders and followers show distinct neural signatures. For example, Jiang et al. (2016) found that leaders during cooperative decision-making exhibited significantly greater INS in the dorsolateral prefrontal cortex (dlPFC) and premotor areas, likely reflecting increased anticipatory modeling and executive control. This suggests that role asymmetry modulates how predictive and adaptive functions are distributed across brains. A testable hypothesis is that leaders exhibit greater top-down monitoring and generative modeling, while followers rely more on reactive adaptation. Such a hypothesis could be tested by manipulating role assignments in musical duets or conversational leadership tasks and tracking corresponding changes in INS.

At the algorithmic level, the present synthesis suggests that mutual adaptation and covert simulation are shared mechanisms across domains. Kurihara et al. (2022) observed increased theta-band INS in left-temporal and central motor regions during unstable anti-phase tapping, suggesting heightened cognitive effort to maintain coordination. Gugnowska et al. (2022) found that gamma-band INS was stronger when pianists were unfamiliar with each other's part, reflecting increased reliance on online perceptual monitoring and perhaps covert simulation. These findings support the hypothesis that INS reflects ongoing predictive processing and partner modeling. This could be tested by introducing controlled perturbations—such as unpredictable tempo shifts in music or prosodic disruptions in speech—and examining whether INS in IFG and motor regions scales with the degree of behavioral correction required.

Another algorithmic-level hypothesis concerns the hierarchical structure of temporal prediction. Coordination in both music and language unfolds over multiple nested timescales, from beats to phrases and syllables to conversational turns. Studies by Sun et al. (2024) and Nguyen et al. (2023) showed that speaker–listener coupling emerges at different cortical sites depending on the linguistic complexity and social context of the exchange. Similarly, Ramírez-Moreno et al. (2023) and Müller et al. (2013) showed that INS during musical improvisation aligns with phrase-level structures. It is plausible that different frequency bands—delta, theta, and beta—encode different hierarchical levels of temporal structure across brains. This hypothesis could be tested using cross-frequency coupling analyses in EEG hyperscanning, and by manipulating interaction structure (e.g., using scrambled musical phrasing or syntactically disrupted speech).

At the implementational level, evidence for dynamic reconfiguration of network topology remains limited but promising. Liu et al. (2017) showed that cooperative versus competitive conditions modulated inter-brain connectivity patterns in temporo-parietal regions. Jiang et al. (2016) used graph-theoretical metrics to show that network centrality was higher for emergent leaders. These studies suggest that while certain hubs (e.g., mPFC, IFG) remain stable, network topology can flexibly reorganize depending on task demands. Future work could explicitly compare different interaction types—such as improvisation versus scripted performance, or open-ended conversation versus structured turn-taking—using graph-based measures of inter-brain networks to test this flexibility.

The scope of the present article was deliberately limited to the intersection of studies on conversation and joint music-making that specifically employed hyperscanning, thereby excluding the considerably more extensive body of neural research on conversation or speech conducted at the single-person level. To give just one example, while most hyperscanning research on (INS) during joint music-making performance has focused on frontocentral, motor, and temporal regions, recent single-brain studies suggest that posterior-right visual areas may also play a role in coordination—particularly in asymmetrical contexts. Zhou et al. (2016) reported that beta-band activity in occipital electrodes (P6, PO8, PO4, Oz, POz) was significantly stronger in followers than leaders during joint movement tasks, with source localization implicating early visual cortex. This suggests a role for visuomotor coupling in followers, who may rely more on visual monitoring to maintain alignment with the leader. In ensemble music performance, similar dynamics are often observed behaviorally: followers attend to the gestures or micro-cues of more dominant players. Future hyperscanning studies should systematically investigate whether such posterior-right modulation emerges during musical interaction, particularly in leader–follower paradigms, duets involving non-verbal cues, or setups that limit auditory access. If confirmed, this would extend current models of INS to include not only predictive timing and motor simulation, but also visual monitoring as a dynamic, role-dependent mechanism of coordination, especially in visually accessible or spatially distributed performance contexts.

This article has sought to take a step toward integration by systematically aligning hyperscanning findings from both joint music-making and conversation within a shared theoretical framework grounded in Marr's levels of analysis. It is hoped that this integrative effort will not only contribute to advancing the state of the art but also inspire and fortify future research to build upon it—both theoretically and empirically—in the ongoing endeavor to bridge disciplinary divides in the cognitive neuroscience of human interaction.

6. References

Abiru, M., Sakai, H., Sawada, Y., & Yamane, H. (2016). The effect of the challenging two-handed rhythm tapping task to DLPFC activation. Asian Journal of Occupational Therapy, 12(1), 75–83. https://doi.org/10.11596/asiajot.12.75

Ahn, S., Cho, H., Kwon, M., Kim, K., Kwon, H., Kim, B. S., et al. (2018). Interbrain phase synchronization during turn-taking verbal interaction—a hyperscanning study using simultaneous EEG/MEG. Human Brain Mapping, 39(1), 171–188.

Andersen, M. M., Kiverstein, J., Miller, M., & Roepstorff, A. (2023). Play in predictive minds: A cognitive theory of play. Psychological Review, 130(2), 462–479. https://doi.org/10.1037/rev0000369

Anderson, M. L. (2014). After phrenology: Neural reuse and the interactive brain. MIT Press.

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. Trends in Cognitive Sciences, 16(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Aquino, M. P. B. d., Verdejo-Román, J., Pérez-García, M., & Pérez-García, P. (2019). Different role of the supplementary motor area and the insula between musicians and non-musicians in a controlled musical creativity task. Scientific Reports, 9(1), 13006. https://doi.org/10.1038/s41598-019-49405-5

At, A., Spierer, L., & Clarke, S. (2011). The role of the right parietal cortex in sound localization: A chronometric single pulse transcranial magnetic stimulation study. Neuropsychologia, 49(9), 2794–2797. https://doi.org/10.1016/j.neuropsychologia.2011.05.024

Azar, M., Cox, G., & Impett, L. (2021). Introduction: Ways of machine seeing. AI & Society, 36(4), 1093–1104. https://doi.org/10.1007/s00146-020-01124-6

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. Frontiers in Psychology, 2, 94. https://doi.org/10.3389/fpsyg.2011.00094

Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. Research on Language and Social Interaction, 50(1), 71–89. https://doi.org/10.1080/08351813.2017.1262118

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. Nature Reviews Neuroscience, 13(6), 407–420. https://doi.org/10.1038/nrn3241

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences, 18(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Chabin, T., Bénony, H., Roy, M., & Dumas, G. (2022). Flow together: Behavioral and inter-brain signatures of team flow during musical improvisation. Neuropsychologia, 168, 108173. https://doi.org/10.1016/j.neuropsychologia.2022.108173

Cheng, S., Wang, J., Luo, R., & Hao, N. (2024). Brain to brain musical interaction: A systematic review of neural synchrony in musical activities. Neuroscience & Biobehavioral Reviews, 164, 105812. https://doi.org/10.1016/j.neubiorev.2024.105812

Cohen, M. X. (2011). Error-related medial frontal theta activity predicts cingulate-related structural connectivity. NeuroImage, 55(3), 1373–1383. https://doi.org/10.1016/j.neuroimage.2010.12.072

Craver, C. F. (2007). Explaining the brain: Mechanisms and the mosaic unity of neuroscience. Oxford University Press.

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. NeuroImage, 59(3), 2430–2437. https://doi.org/10.1016/j.neuroimage.2011.09.003

Czeszumski, A., Eustergerling, S., Volz, L. J., König, P., & König, S. U. (2020). Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. Frontiers in Human Neuroscience, 14, 39. https://doi.org/10.3389/fnhum.2020.00039

Davidson, J. W. (2005). Bodily communication in musical performance. In D. Miell, R. MacDonald, & D. Hargreaves (Eds.), Musical communication (pp. 215–238). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198529361.001.0001

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., … Poeppel, D. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. Current Biology, 27(9), 1375–1380. https://doi.org/10.1016/j.cub.2017.04.002

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. PLoS ONE, 5(8), e12166. https://doi.org/10.1371/journal.pone.0012166

Enfield, N. J. (2015). Linguistic Relativity from Reference to Agency. Annual Review of Anthropology, 44(1), 207-224.

Forgeard, M., Schlaug, G., Norton, A., Rosam, C., & Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. Music Perception, 25(4), 383–390. https://doi.org/10.1525/mp.2008.25.4.383

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences, 6(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. Annual Review of Psychology, 63, 287–313. https://doi.org/10.1146/annurev-psych-120710-100449

Friston, K., & Frith, C. D. (2015). A duet for one: Predictive coding and active inference in social interaction. Brain Research, 1626, 133–152. https://doi.org/10.1016/j.brainres.2015.03.021

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. Frontiers in Psychology, 6, 1777. https://doi.org/10.3389/fpsyg.2015.01777

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., Xanthopoulos, P., Sakkalis, V., & Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. Journal of NeuroEngineering and Rehabilitation, 5, 25. https://doi.org/10.1186/1743-0003-5-25

Green, A. C., Bærentsen, K. B., Stødkilde-Jørgensen, H., Wallentin, M., Roepstorff, A., & Vuust, P. (2008). Music in minor activates limbic structures: A relationship with dissonance? NeuroReport, 19(7), 711–715. https://doi.org/10.1097/WNR.0b013e3282fd0dd8

Gugnowska, K., Novembre, G., Kohler, N., Villringer, A., Keller, P. E., & Sammler, D. (2022). Endogenous sources of interbrain synchrony in duetting pianists. Cerebral Cortex, 32(18), 4110–4127. https://doi.org/10.1093/cercor/bhab469

Hamilton, A. F. D. C. (2021). Hyperscanning: Beyond the hype. Neuron, 109(3), 404–407. https://doi.org/10.1016/j.neuron.2020.11.008

Hawkins, S., Cross, I., & Ogden, R. (2013). Communicative interaction in spontaneous music and speech. In M. Orwin, C. Howes, & R. Kempson (Eds.), Language, music and interaction (Vol. 3, pp. 285–329). College Publications.

Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. Trends in Cognitive Sciences, 10(2), 64–69. https://doi.org/10.1016/j.tics.2005.12.005

Hu, Y., Zhu, M., Liu, Y., Wang, Z., Cheng, X., Pan, Y., & Hu, Y. (2022). Musical metre induces interbrain synchronization during interpersonal coordination. eNeuro, 9(5). https://doi.org/10.1523/ENEURO.0504-21.2022

Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), Style in language (pp. 350-377). New York & London: John Wiley & Sons, Inc.

Jackendoff, R. (2002). Foundations of language: Brain, meaning, grammar, evolution. Oxford University Press.

Jentschke, S., & Koelsch, S. (2009). Music training modulates the development of syntax processing in children. NeuroImage, 47(2), 735–744. https://doi.org/10.1016/j.neuroimage.2009.04.090

Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., & Feldman, R. (2017). Brain-to-brain synchrony during naturalistic social interactions. Scientific Reports, 7, 17060. https://doi.org/10.1038/s41598-017-17339-5

Kelsen, B. A., Sumich, A., Kasabov, N., Liang, S. H. Y., & Wang, G. Y. (2022). What has social neuroscience learned from hyperscanning studies of spoken communication? A systematic review. Neuroscience & Biobehavioral Reviews, 132, 1249–1262. https://doi.org/10.1016/j.neubiorev.2021.12.017

Khalil, A., Musacchia, G., & Iversen, J. R. (2022). It takes two: Interpersonal neural synchrony is increased after musical interaction. Brain Sciences, 12(3), 409. https://doi.org/10.3390/brainsci12030409

Koelsch, S. (2011). Toward a neural basis of music perception – a review and updated model. Frontiers in Psychology, 2, 110. https://doi.org/10.3389/fpsyg.2011.00110

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. Nature Reviews Neuroscience, 11(8), 599–605. https://doi.org/10.1038/nrn2882

Kurihara, Y., Takahashi, T., & Osu, R. (2022). The relationship between stability of interpersonal coordination and inter-brain EEG synchronization during anti-phase tapping. Scientific Reports, 12, 6164. https://doi.org/10.1038/s41598-022-10049-7

Levinson, S. C. (2016). Turn-taking in human communication—origins and implications for language processing. Trends in Cognitive Sciences, 20(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Lender, A., Perdikis, D., Gruber, W., Lindenberger, U., & Müller, V. (2023). Dynamics in interbrain synchronization while playing a piano duet. Annals of the New York Academy of Sciences, 1530(1), 124–137. https://doi.org/10.1111/nyas.15072

Lindenberger, U., Li, S.-C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. BMC Neuroscience, 10, 22. https://doi.org/10.1186/1471-2202-10-22

Liu, T., Saito, G., Lin, C., & Saito, H. (2017). Inter-brain network underlying turn-based cooperation and competition: A hyperscanning study using near-infrared spectroscopy. Scientific Reports, 7, 8684. https://doi.org/10.1038/s41598-017-09226-w

Liu, T., Saito, H., & Oi, M. (2016). Role of the right inferior frontal gyrus in turn-based cooperation and competition: A near-infrared spectroscopy study. Brain and Cognition, 102, 13–22. https://doi.org/10.1016/j.bandc.2015.12.001

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science, 288(5472), 1835–1838. https://doi.org/10.1126/science.288.5472.1835

Magne, C., Schön, D., & Besson, M. (2006). Music training and pitch processing in children: Electrophysiological evidence. Journal of Cognitive Neuroscience, 18(2), 218–230. https://doi.org/10.1162/jocn.2006.18.2.218

Malinowski, B. (1923). The problem of meaning in primitive languages. In C. K. Ogden & I. A. Richards (Eds.), The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism (pp. 296-336). London: Routledge.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. W. H. Freeman.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. Annual Review of Neuroscience, 24, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

Moreno, S., & Besson, M. (2006). Musical training and language-related brain electrical activity in children. Clinical Neurophysiology, 117(12), 2703–2714. https://doi.org/10.1016/j.clinph.2006.06.061

Müller, V., & Lindenberger, U. (2023). Intra- and interbrain synchrony and hyperbrain network dynamics of a guitarist quartet and its audience during a concert. Annals of the New York Academy of Sciences, 1523(1), 74–90. https://doi.org/10.1111/nyas.14987

Müller, V., Sänger, J., & Lindenberger, U. (2013). Intra- and inter-brain synchronization during musical improvisation on the guitar. PLoS ONE, 8(9), e73852. https://doi.org/10.1371/journal.pone.0073852

Nguyen, T., Zimmer, L., & Hoehl, S. (2023). Your turn, my turn. Neural synchrony in mother–infant proto-conversation. Philosophical Transactions of the Royal Society B: Biological Sciences, 378(1875), 20210488. https://doi.org/10.1098/rstb.2021.0488

Niranjan, D., Toiviainen, P., Brattico, E., & Alluri, V. (2019). Dynamic functional connectivity in the musical brain. In Brain informatics (pp. 82–91). Springer. https://doi.org/10.1007/978-3-030-37078-7_9

Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2014). Motor simulation and the coordination of self and other in real-time joint action. Social Cognitive and Affective Neuroscience, 9(8), 1062–1068. https://doi.org/10.1093/scan/nst086

Nunez, P. L., & Srinivasan, R. (2006). Electric fields of the brain: The neurophysics of EEG (2nd ed.). Oxford University Press.

Patel, A. D. (2008). Music, language, and the brain. Oxford University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Frontiers in Psychology, 2, 142. https://doi.org/10.3389/fpsyg.2011.00142

Pelt, S., Heil, L., Kwisthout, J., Ondobaka, S., van Rooij, I., & Bekkering, H. (2016). Beta- and gamma-band activity reflect predictive coding in the processing of causal events. Social Cognitive and Affective Neuroscience, 11(6), 973–980. https://doi.org/10.1093/scan/nsw017

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect motor planning in the face of response uncertainty. Journal of Neuroscience, 35(44), 14738–14747. https://doi.org/10.1523/JNEUROSCI.0816-15.2015

Poeppel, D. (2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. Cognitive Neuropsychology, 29(1–2), 34–55. https://doi.org/10.1080/02643294.2012.710600

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? Trends in Cognitive Sciences, 10(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. Neuron, 72(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Ramírez-Moreno, M. A., Cruz-Garza, J. G., Acharya, A., Chatufale, G., Witt, W., Gelok, D., Reza, G., & Contreras-Vidal, J. L. (2023). Brain-to-brain communication during musical improvisation: A performance case study. F1000Research, 10, 732. https://doi.org/10.12688/f1000research.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. Nature Reviews Neuroscience, 20(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Rietmolen, N., Zhang, J., & Large, E. W. (2024). Neural oscillations in music and language: Toward a unifying framework of temporal prediction. Trends in Cognitive Sciences, 28(3), 211–225. https://doi.org/10.1016/j.tics.2023.11.005

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

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50(4), 696–735. https://doi.org/10.2307/412243

Sänger, J., Müller, V., & Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. Frontiers in Human Neuroscience, 7, 234. https://doi.org/10.3389/fnhum.2013.00234

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. Trends in Cognitive Sciences, 10(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Speer, S. P. H., Mwilambwe-Tshilobo, L., Tsoi, L., Burns, S. M., Falk, E. B., & Tamir, D. I. (2024). Hyperscanning shows friends explore and strangers converge in conversation. Nature Communications, 15(1), 7781. https://doi.org/10.1038/s41467-024-33562-3

Tachibana, A., Noah, J. A., Ono, Y., Taguchi, D., & Ueda, S. (2019). Prefrontal activation related to spontaneous creativity with rock music improvisation: A functional near-infrared spectroscopy study. Scientific Reports, 9, 16044. https://doi.org/10.1038/s41598-019-52348-6

Thomson, J. M., Leong, V., & Goswami, U. (2013). Auditory processing interventions and developmental dyslexia: A comparison of phonemic and rhythmic approaches. Reading and Writing, 26, 139–161. https://doi.org/10.1007/s11145-012-9359-6

Tiitinen, H., Salminen, N. H., Palomäki, K. J., Mäkinen, V. T., Alku, P., & May, P. J. C. (2006). Neuromagnetic recordings reveal the temporal dynamics of auditory spatial processing in the human cortex. Neuroscience Letters, 396(1), 17–22. https://doi.org/10.1016/j.neulet.2005.11.018

Turino, T. (2008). Music as social life : the politics of participation. London: University of Chicago Press.

Vuust, P., Heggli, O. A., Friston, K. J., & Kringelbach, M. L. (2022). Music in the brain. Nature Reviews Neuroscience, 23(5), 287–305. https://doi.org/10.1038/s41583-022-00578-5

Wan, X., Crüts, B., & Jensen, H. J. (2014). The causal inference of cortical neural networks during music improvisations. PLoS ONE, 9(12), e112776. https://doi.org/10.1371/journal.pone.0112776

Yuste, R. (2015). From the neuron doctrine to neural networks. Nature Reviews Neuroscience, 16(8), 487–497. https://doi.org/10.1038/nrn3962

Zatorre, R. J. (2001). Spectral and temporal processing in human auditory cortex. Cerebral Cortex, 11(10), 946–953. https://doi.org/10.1093/cercor/11.10.946

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. Nature Reviews Neuroscience, 8, 547–558. https://doi.org/10.1038/nrn2152

Zhang, T., Zhou, S., Bai, X., Zhou, F., Zhai, Y., Long, Y., & Lu, C. (2023). Neurocomputations on dual-brain signals underlie interpersonal prediction during a natural conversation. NeuroImage, 251, 119022. https://doi.org/10.1016/j.neuroimage.2023.120400

Zhou, G., Bourguignon, M., Parkkonen, L., & Hari, R. (2016). Neural signatures of hand kinematics in leaders vs. followers: A dual-MEG study. NeuroImage, 125, 731–738. https://doi.org/10.1016/j.neuroimage.2015.10.049

Conflict of interest

The authors declare no conflicts of interest

Acknowledgments

This research was supported by the ANID Millennium Science Initiative Program (ICS2019_024).

Author contributions

J-PR conceived the review, undertook all literature review, and wrote most of the article. JT conceived and wrote most of section 1.4 on caveats, and heavily revised the rest of the article's drafting. IC, MP and JK made major contributions to the article's drafting.