

Musical Dimensions of Motion-Based Interindividual Behaviors

Situated in an Interactive Music Environment

Jin Hyun Kim and Marta Rizzonelli¹

Abstract: This chapter discusses the extent to which motion-based interindividual behaviors in an interactive music environment can be described as musical, focusing on an exploratory study using “Sentire,” a system that transforms proximity and movement into real-time musical feedback. We propose that coordinated and cooperative musical interaction consists not only of sounds, but also of body movements considered meaningful in a given cultural context. Through structured observation, we analyzed improvisations with Sentire, developing functional codes categorizing interindividual behaviors—non-interactive, mirroring, and residual interactive. This study allowed us to distinguish phases of interaction—exploratory and systematic. While mirroring occurs throughout, its characteristics vary by phase. In the systematic phase, more complex mirroring behaviors—based on an anticipatory agreement between participants and characterized as musical—emerged through continuous mutual adaptation.

Introduction

Motion-based interindividual behaviors are often observed in contexts of verbal interaction. Some of them are known as speech-accompanying non-verbal gestures that support or replace verbal expressions. Although processing such gestures does not always engage the neural language-processing regions, especially in the case that speech is absent,² co-speech gestures can support

1 These two authors contributed equally to this work.

2 See, for instance, Olessia Jouravlev, David Zheng, Zuzanna Balewski, Alvince Le Arnz Pongos, Zena Levan, Susan Goldin-Meadow, and Evelina Fedorenko, “Speech-Accompanying Gestures Are Not Processed by the Language-

language comprehension. Some motion-based interindividual behaviors do not serve to accompany or replace verbal expressions, but are considered meaningful in a given cultural context. Examples include behaviors that are exhibited in protoconversational communication taking place, for instance, in infant-caregiver interactions.³ Protoconversational communication is characterized by interindividual coordination that involves sound- and motion-based behaviors in sync/tune with one another as well as qualitatively attuned behaviors. These behaviors, in turn, provide a shared rhythmic foundation for turn-taking and a shared intonational foundation for prosodic matching and bodily adjustment. The developmental psychologists Stefan Malloch and Colwyn Trevarthen term such behaviors “communicative musicality”: the innate human abilities to move, remember and plan while interacting with others, which allow us to appreciate and produce an endless variety of events unfolding in time.⁴ “Communicative musicality” presupposes the concept of music, but is not limited to phenomena described with emic concepts of what is called “music” in Western culture. In addition, the concept of music that can be found in the definition of communicative musicality is too broad and therefore deserves thorough discussion. The concept of communicative musicality suggests that language and other human communicative practices—including musical practices—are based on a common innate ability to interact with others using body movements rather than verbal articulations that may bear representational and propositional meanings. The first author proposes a

Processing Mechanisms,” in *Neuropsychologia* 132 (2019), accessed 15 September 2022, <https://www.doi.org/10.1016/j.neuropsychologia.2019.107132>.

- 3 See, for instance, Stephen Malloch and Colwyn Trevarthen, “The Human Nature of Music,” in *Frontiers in Psychology* 9:1680 (2018), accessed 15 September 2022, <https://www.doi.org/10.3389/fpsyg.2018.01680>; Colwyn Trevarthen, “Musicality and the Intrinsic Motive Pulse: Evidence from Human Psychobiology and Infant Communication,” in *Musicae Scientiae* 3 (1999): 157–213, accessed 15 September 2022, <https://www.doi.org/10.1177/10298649000030S109>; Colwyn Trevarthen, “Learning about Ourselves, from Children: Why a Growing Human Brain Needs Interesting Companions?,” in *Research and Clinical Centre for Child Development, Annual Report 2002–2003*, 26 (2004): 9–44.
- 4 Cf. Stephen Malloch and Colwyn Trevarthen, “Musicality: Communicating the Vitality and Interests of Life,” in *Communicative Musicality: Exploring the Basis of Human Companionship*, ed. Stephen Malloch and Colwyn Trevarthen (Oxford: Oxford University Press, 2009), 1–11, 4.

minimal concept⁵ of musical practices that does justice to an innate ability for protoconversational communication by considering it a basis for musical practices, among other communicative practices:

Musical practices are coordinated and cooperative practices consisting of sounds and body movements that structure shared experiences. They are considered meaningful in a given cultural context in that act of understanding occurring in terms of interactive participation and embodied re-enactment, although they do not bear any representational semantics.

Musical practices are *coordinated* due to the regularity and predictability of sounds and body movements unfolding in time, which provide the means to coordinate and sustain interactions. Processes of coordination such as rhythmic adjustment (entrainment),⁶ alignment related to non-rhythmic properties (such as vertical body movement, intensity of body movement or sound, timbre and pitch of sound),⁷ and affective adjustment (such as attunement, empathy, etc.) are characteristics of behaviors that structure shared experiences.

Moreover, musical practices are *cooperative* as they involve shared intentionality—at least we-intentions⁸—to shape musical units using sounds and body movements. The individuals' intentions to participate in a collective act are derived from the overall goal of the collective behavior in such a way that each individual believes that they are doing their individual act as part of a collective

5 It is discussed as a broader concept of music (cf. Jin Hyun Kim, "Musicality of Coordinated Non-representational Forms of Vitality", in *Journal of Comparative Literature and Aesthetics, Special Issue – Contemplating Music across Cultures and Contexts: Philosophical Perspectives*, 46/1 (2023): 59–69).

6 Cf. Martin Clayton, Kelly Jakubowski, and Tuomas Eerola, "Interpersonal Entrainment in Indian Instrumental Music Performance: Synchronization and Movement Coordination Relate to Tempo, Dynamics, Metrical and Cadential Structure," in *Musicae Scientiae* 23/3 (2019): 304–31, accessed 15 September 2022, <https://www.doi.org/10.1177/1029864919844809> and Jin Hyun Kim, Andres Reifgerst, and Marta Rizzonelli, "Musical social entrainment," in *Music & Science* 2 (2019): 1–17, accessed 22 September 2022, <https://www.doi.org/10.1177/2059204319848991>.

7 Cf. Jin Hyun Kim, "Musik als nicht-repräsentationales Embodiment. Philosophische und kognitionswissenschaftliche Perspektiven einer Neukonzeptualisierung von Musik," in *Musik und Körper. Interdisziplinäre Dialoge zum körperlichen Erleben und Verstehen von Musik*, ed. Lars Oberhaus and Christoph Stange (Bielefeld: transcript, 2017), 145–64.

8 Cf. John Searle, *The Construction of Social Reality* (New York, NY: Free Press, 1995); Kim, Reifgerst, and Rizzonelli, "Musical Social Entrainment."

act.⁹ Collective interindividual behavior involving we-intentions is oriented towards cooperation with a common goal, which makes it distinguishable from the summation of individual acts that do not involve cooperation.¹⁰ This type of interindividual behavior emerges if each individual is aware of the effect that is reached through interindividual behavior.¹¹

Understanding musical practices does not require meta-representational processes since a representational account of meanings of the sounds and body movement that musical practices consist of is unsatisfactory.¹² There are no meanings of musical expression that can be stated regardless of context or articulated by other expressions that have the same meaning. Terms like “true,” “satisfying,” “expressive,” etc., that articulate specific semantic relations between the expressions and the world, are not applicable to music. Instead, musical meanings are generated within the course of musical understanding, which occurs as interactive participation and embodied re-enactment.¹³ For instance, musical meanings are generated while producing sounds and body movements through interactive participation such as singing and dancing along, or through embodied re-enactment of sonorous-rhythmic structures such as dance movements imitating features of sonorous-rhythmic structures.¹⁴ Such activities of interactive participation and embodied re-enactment single out those aspects of musical practices that are relevant in a given cultural context.

Based on this minimal concept of musical practices, this chapter presents one of our exploratory studies and discusses the extent to which motion-based

9 Cf. John Searle, “Collective Intentions and Actions,” in *Intentions in Communication*, ed. Philip R. Cohen, Jerry Morgan, and Martha E. Pollack (Cambridge, MA: MIT Press, 1990), 401–15, 407.

10 Ibid., 406.

11 Cf. Searle, *The Construction of Social Reality*; Kim, Reifgerst, and Rizzonelli, “Musical Social Entrainment.”

12 Cf. Kim, “Musik als nicht-repräsentationales Embodiment”; Jin Hyun Kim and Matthias Vogel, “Nichtsprachliches Musikverstehen: Einleitung,” in *Nichtsprachliches Musikverstehen. Zur Neuperspektivierung der musikalischen Hermeneutik*, ed. Jin Hyun Kim and Matthias Vogel (Heidelberg and Wiesbaden: Metzler/Springer, forthcoming).

13 Cf. Kim and Vogel, “Nichtsprachliches Musikverstehen.”

14 Cf. Arnie Cox, *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking* (Bloomington: Indiana University Press, 2016); Matthias Vogel, “Nachvollzug und die Erfahrung musikalischen Sinns,” in *Musikalischer Sinn. Beiträge zu einer Philosophie der Musik*, ed. Alexander Becker and Matthias Vogel (Frankfurt am Main: Suhrkamp, 2007), 314–68.

interindividual behaviors that emerge while socially interacting in an interactive music environment could be described as musical.

Exploratory Study on Motion-Based Interindividual Behaviors Emerging in Free Musical Improvisation

In this section, we present our explorative study using structured observation, which was carried out on four participant couples who interacted using Sentire,¹⁵ an interactive music system consisting of a body-machine interface which detects movements and proximity between two persons and provides real-time music feedback. Subjects freely improvised their behavior through body movements accompanied by sound feedback. According to participants' cultural background, this kind of improvisation could be associated with a specific genre of dance improvisation or musical improvisation, yet it was not necessary for participants to apply skills in musical improvisation or dance improvisation. Rather, they were invited to act with their bodies in relation to another interactant. Sentire allows subjects to explore a dyadic interaction, starting with the broad common goal ("we-intentions") of interacting with each other using body movements, without a set of rules or a specific role for each interactant. Our research question asks whether it is possible to identify a relevant sequential order of behaviors through which we might be able to infer how interindividual synchrony emerges as a result of musical interaction in a broader sense. Based on our previous observational studies,¹⁶ our initial hypothesis holds that Sentire, which is designed to facilitate non-linguistic,¹⁷ musical interaction using body movements in the contexts of everyday life and therapy, fosters basic social interactive behaviors¹⁸ without necessarily involving a high-order capacity of mind-reading, as well as "a strong kind of

15 See <https://vimeo.com/317080128> (accessed 15 September 2022) for a video example of how Sentire can be used.

16 Marta Rizzonelli, Jin Hyun Kim, Pascal Staudt, and Marcello Lussana, "Fostering Social Interaction through Sound Feedback: Sentire," in *Organised Sound* 28/1 (2023): 97–109, accessed 28 October 2023, <https://www.doi.org/10.1017/S1355771822000024>.

17 By "non-linguistic" we mean that it is about musical interaction that does not serve as any kind of language-related interaction, not merely a matter of non-verbal interaction which might involve bodily gestures that can serve to accompany or replace language.

18 Cf. Ibid.

collective intentionality with [...] a shared semantic content and [...] specifically defined roles” for each actant.¹⁹

Study Design

This study was part of a larger experiment in which the usage of Sentire was compared with two control conditions (recorded sound and no sound) to look for a baseline effect of the Sentire system on the facilitation of interaction. This study was approved by the Ethics Committee of the Faculty of Humanities and Social Sciences Ethics Committee at Humboldt University of Berlin (approval no. HU-KSBF-EK_2020_0014). All participants provided written informed consent regarding study design, audio and video recordings, processing of sensitive data, and compliance with Covid-19 protective measures.

Examining four sessions conducted with Sentire, we carried out an exploratory analysis based on informal and structured observation, to investigate whether it is possible to find motion-based interindividual behaviors that show how synchrony emerges as a result of musical interaction—that is, cooperative interaction beyond mere coordination. To this end, based on previous analyses, we selected sessions in which synchrony had been detected consistently throughout the interaction.

Participants were two men and six women (mean age 25,88 years, s.d. $\pm$ 4,39) in four same-gender couples, in an effort to (try to) control for the possible confounding variable of sexual attraction.²⁰ Participants did not know each other before the study and were asked to interact with each other for eight minutes using body movements. Without being assigned any particular task, the instruction was to move freely and interact with the partner, given the limitations of avoiding touch and wearing a face mask due to Covid-19 regulations. During the interaction, an ambient pad sound intended to promote physical proximity was generated in real-time in response to participants’ movements and fluctuations in proximity. Audio and video were recorded for the duration of the interaction.

19 Kim, Reifgerst, and Rizzonelli, “Musical Social Entrainment.”

20 The choice of creating same-gender couples was made assuming that most participants were heterosexual. To verify this assumption, this information was collected in a brief questionnaire together with demographic data.

The technical setup consisted primarily of Sentire hardware and software.²¹ Each participant wore a bracelet that takes advantage of the human body's electric conductivity to generate a signal within the same electrical circuit (capacitive coupling). The signal's strength depends on the distance between the two participants ("proximity effect") and is used to produce a control signal which changes linearly with the distance of the two bodies. The control signal is mapped to specific parameters of an algorithmic sound synthesis environment (parameter mapping), which generate real-time, closed-loop auditory interaction between the two persons. Detection and sonification of the control signal are achieved by a custom software and a sound generator written in SuperCollider. In addition to the hardware and software components of the Sentire system, the technical setup included loudspeakers, three USB-cameras set in a triangular configuration (to guarantee an accurate capture of the full interaction space), and two computers (placed in a separate room) to control Sentire and the cameras.

A Behavioral Sequential Approach: Structured Observation

Structured observation is an established method in behavioral and social sciences. It consists in observing audio and/or video recordings and identifying specific behaviors by annotating them, i.e. assigning them a label (or code). Annotation can be carried out in different ways: simply counting the number of occurrences of a certain event, with the option of marking its length or not (timed or untimed event recording); or pre-segmenting the flow of behavior into brief, fixed time intervals and attributing a label either to each of them (interval recording) or only to some of them (selected-interval recording). Each approach has advantages and disadvantages, and the choice depends on research questions, available time and resources, and the complexity of the data.²² These approaches share an aim towards systematic observation which can be treated as measurements, so long as codes fulfill certain logical conditions:

-
- 21 Pascal Staudt, Anton Kogge, Marcello Lussana, Marta Rizzonelli, Benjamin Stahl, and Jin Hyun Kim, "A New Sensor Technology for the Sonification of Proximity and Touch in Closed-Loop Auditory Interaction," in *Proceedings of the 19th Sound and Music Computing Conference (SMC-2022)*, Saint-Étienne, June 8–11, 2022: 446–53, accessed 15 September 2022, <https://www.doi.org/10.5281/zenodo.6798242>.
 - 22 For a detailed explanation, see Roger Bakeman and Vicenç Quera, *Sequential Analysis and Observational Methods for the Behavioral Sciences* (New York: Cambridge University Press, 2011).

1) having a strict one-to-one correlation with the unit of behavior they are assigned to, which makes codes comparable to nominal scales used in quantitative studies,²³ 2) being exhaustive, meaning that every type of behavior must be attributable to a code, and 3) being mutually exclusive, meaning that only one code can be applicable for each behavior.²⁴ It is clear, therefore, that the development of the codes and their definitions (summarized in the so-called coding scheme) is the core process of the method.

To identify effective codes, structured observation must be preceded by one or more phases of informal observation, during which various examples of the behavior of interest are analyzed to determine relevant events, their order of magnitude (in terms of duration) and whether there are behaviors that have to be ignored. Informal observation is carried out by first describing the observed behaviors in a simple and objective manner, and then organizing these descriptions into hierarchically-ordered categories (the future codes). It is often necessary to carry out several phases of informal observation, starting from more general aspects and/or larger timescales, and proceeding step by step towards finer behavioral aspects, which usually unfold on a smaller timescale. During this iterative incremental process, empirical and theoretical work develop in parallel and inform each other, since, at every step, the empirical observation of specific phenomena is the basis on which theoretical questions, generalizations, or interpretations are built. Although it is therefore only after a repeated and refined use of informal observation that it becomes possible to perform structured observation in the strict sense (i.e., systematic application of behavioral codes according to a specific and often ad hoc developed coding scheme), an observational study using a behavioral sequential approach often consists of several steps of oscillation between informal and structured observation until final behavioral codes are developed. This involves a lengthy process of investigation, but allows scholars to address research questions closely related to real-world behaviors.

An additional aspect lending structured observation the property of a quasi-quantitative method is the ability to measure the annotation's reliability, through inter-observer agreement (IOA) between two independent observers.

23 Cf. Uwe Seifert and Jin Hyun Kim, "Towards a Conceptual Framework and an Empirical Methodology in Research on Artistic Human-Computer and Human-Robot Interaction," in *Human-Computer Interaction*, ed. Ioannis Pavlidis (Vienna: In-Tech, 2008), 177–94.

24 Cf. Colin Robson and Kieran McCartan, "Observational Methods," in *Real World Research*, ed. Colin Robson and Kieran McCartan (Chichester: Wiley, 2016), 318–45.

IOA is usually achieved by calculating Cohen's kappa coefficient²⁵ for at least twenty percent of the data.²⁶

In a previous analysis, we coded interaction based on empirical (i.e., directly and objectively observable) parameters. This observation allowed us to detect two main empirical categories that play a primary role in the interaction supported by Sentire:

- 1) whole-body movements delineated by steps (no steps, forward steps, backward steps, etc.);
- 2) arm movements, particularized by their directedness towards the partner (one or both arms directed towards the partner, arms in a rest position, etc.). Here, it has to be stressed that the importance of arms is strongly related to the use of bracelets to connect participants to the Sentire system.

After coding the video material with empirical categories, we observed several cases where the two participants displayed the same code (i.e., the two participants' codes were overlapping). Table 1 shows the codes for which we considered overlapping and non-overlapping segments, and how we grouped these empirical codes to develop higher-order functional categories. For codes related to whole-body movements, an overlap between participants (e.g., both display the code 'forward_step/s') indicates that behavioral matching has occurred, meaning that similar, and synchronous body configurations or movements are taking place. In this case, directedness towards the partner is not a criterion for the annotation of whole-body movements, and consequently, no information can be detected about the interaction.

For example, if both participants display the code "no_steps," they are exhibiting the same behavior with respect to steps (behavioral matching). This, however, does not provide any information about whether they are interacting or not: they may be looking at each other and moving their arms towards each other; they may equally be looking at the room and not interacting with each other at all. The codes simply do not detect interaction.

25 Jacob Cohen, "A Coefficient of Agreement for Nominal Scales," in *Educational and Psychological Measurement* 20/1 (1960): 37–46.

26 See, e.g., Melanie Pellecchia, Rinad S. Beidas, David S. Mandell, Carolyn C. Cannuscio, Carl J. Dunst, and Aubyn C. Stahmer, "Parent Empowerment and Coaching in Early Intervention: Study Protocol for a Feasibility Study," in *Pilot and Feasibility Studies* 6/22 (2020): 1–12.

Behaviorally matched whole-body movements	Interactive arm movements		
	Partner-directed		Non-partner-directed
	Behaviorally matched	Behaviorally non-matched	
no_steps overlapping forward_step/s overlapping backwards_step/s overlapping lateral_step/s overlapping full_circle overlapping turning_steps overlapping turning_no_steps overlapping	one_towards overlapping both_towards overlapping	one_towards without overlapping both_towards without overlapping	one_not_towards overlapping both_not_towards overlapping one_not_towards without overlapping both_not_towards without overlapping

Table 1: Behaviorally matched whole-body movements and interactive arm movements (see text for a detailed description). In the arm-movement codes, “one” and “both” refer to arms, not to participants. E.g., “one_towards” means: one arm towards the partner

Codes related to arm movements, however, incorporate directedness towards the partner, because every arm movement that can be empirically observed in dyadic interaction in the Sentire environment is coded depending on whether it is directed towards the partner or not. As a result, different functional categories can be inferred based on the presence of directedness and whether or not the two participants’ codes are overlapping. When participants display overlapping codes that indicate directedness (both “one_towards” or both “both_towards”), a behavioral matching is occurring in which both participants are directed towards each other, meaning interaction can be inferred (interactive, partner-oriented behavioral matching). When participants display codes that indicate directedness but do not overlap (only one participant shows “one_towards” or “both_towards”), no behavioral matching is occurring, but one of the two participants is exhibiting partner-related, and therefore possibly interactive, arm movements. Finally, when participants display codes that do not indicate directedness but do overlap

(both “one_not_towards” or both “both_not_towards”), or when participants’ codes neither indicate directedness nor overlap (only one participant shows “one_not_towards” or “both_not_towards”), no behavioral matching is occurring and no directedness towards partner is exhibited, but interaction may still be possible (e.g., due to mutual gaze or arm movements that are different but might complement each other).

Building on the empirical analysis just described, we developed a functional methodological approach, delineated in the following section. The sequential steps of the implementation of the method are indicated in square brackets.

- [1]]To answer the research question of whether synchrony emerges as a result of a specific behavioral pattern that is based on cooperative social interactions, we developed categories that can be understood as functional (i.e., that imply a certain degree of interpretation or inference). We organized the analysis in two steps: observation of the couple as a unit (A+B), followed by observation of each participant separately (A, B). In the annotation of A+B, “directedness” is the discerning criterion and its meaning needs to be specified: the iterative process of informal observation has made it evident that a person can be directed towards the partner with any body part, even with just a slight rotation of the head towards them. In fact, when two strangers interact only using body movements for the first time, and especially during the initial phase of the interaction, directedness often occurs only in a cautious, partial form. In our understanding, any form of directedness between participants, from the most cautious to the most complete, which includes gaze and whole-body, is an indicator of interaction. Tables 2.1 and 2.2 present and define functional categories for A+B and for A, B, respectively; the decision tree in figure 1 shows the logical relationship among the codes within and between the two steps of analysis.
- [a]]At first, functional categories were applied in concomitance with segmentation: every time a certain behavior was observed, it was necessary to identify a segment (i.e., the precise beginning and end of that behavior) before assigning it a specific code (category). While this procedure is convenient for empirical categories where the temporal limits of a behavior are usually recognizable due to a shared definition of that behavior (e.g., step forward), identifying the limits of a functional category is much more prone to subjective judgment and can easily lead to disagreement between observers.
- [b]]Therefore, we applied the strategy of interval recording: the behavioral flow is pre-segmented in blocks of a specific duration (in our case, ten seconds) and each block must be annotated with one code.

This avoids not only disagreement regarding beginning and end of a behavior, but also excessive or insufficient annotation (i.e., a tendency to detect too many or too few events). Of course, this segmentation being arbitrary, it may well happen that different behaviors occur in the same block: in this case, only the behavior that covers the majority of the block is annotated, whereas any others are ignored.

A + B	
Near-simultaneous, similar or opposite behavior	Movements Both subjects are directed towards each other (through gaze and/or whole body/body part) Movements are near-simultaneous (not necessarily exactly simultaneous, but with temporal overlap) Movements are either similar or opposite
Turn-taking	Movements Both subjects are directed towards each other (through gaze and/or whole body/body part) Movements do not overlap in time
Residual interactive category	EITHER There are no movements, but both subjects are directed towards each other through gaze OR There are movements, and both subjects are directed towards each other (through gaze and/or whole body/body part), but movements are neither near-simultaneous nor turn-taking
No interaction	EITHER Neither of the two subjects is directed towards the other (whether through gaze or whole body/body part) OR Only one person is directed towards the other

Not assignable	Movements or no movements It is not possible to determine whether subjects are directed towards each other (through gaze and/or whole body/body part)
----------------	--

Table 2.1: Coding scheme of functional categories for the observation of the couple as a unit (A+B)

A, B	
Does nothing and waits but watches partner	The observed person does not perform any movement or other action (passive behavior or posture) The observed person watches the partner (and is therefore interactive)
Becomes initiative within the interactive context	The observed person performs movements or other actions (active behavior, including moving away from partner) The observed person takes initiative, performing a movement first (or at the same time as the other person) For all sections where it is very difficult to determine who starts and who follows, we choose for both participants the code becomes initiative within the interactive context.
Follows partner	The observed person performs movements or other actions (active behavior) The observed person does not take initiative but simply follows the partner
Is not directed to partner (neither gaze nor movements)	The observed person does not watch the partner (and is therefore not interactive) The observed person performs one of the following actions: touches the cable* looks around touches her/his own clothes, glasses, etc. * For each situation where a person touches the cable, we need to assess whether touching the cable is really the primary focus (is not directed to partner) or the touch is just a secondary focus (and in this case we choose between becomes initiative within the interactive context and follows partner)

- Not assignable	The first and the last categories cannot be distinguished because directedness is not clearly observable
------------------	--

Table 2.2: Coding scheme of functional categories for the observation of each participant separately (A, B)

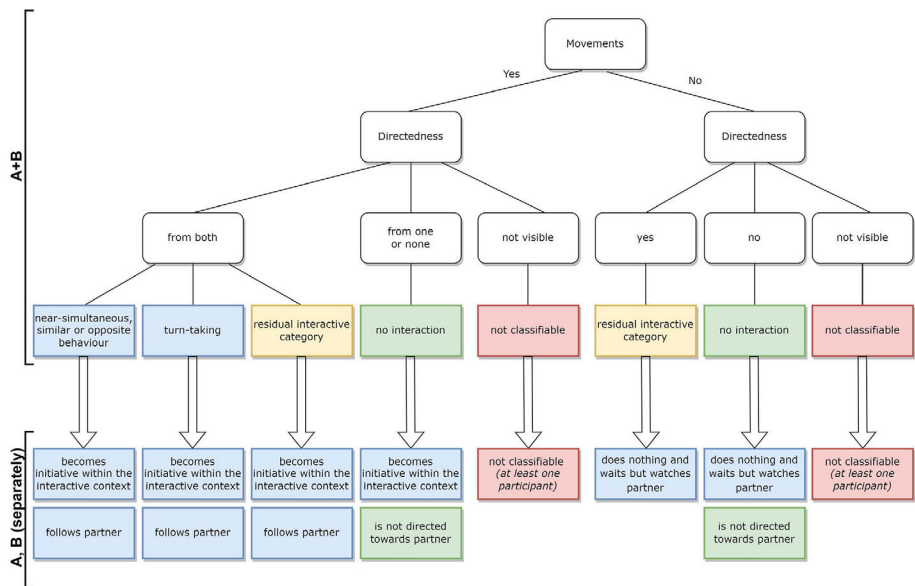


Figure 1: Decision tree showing the logical path for the attribution of codes, to the couple as a unit (A+B) and to each participant separately (A, B), as well as the logical relationship between these two levels. In each column of “A, B (separately),” either a combination of the two existing categories or the duplication of one category is possible. © Marta Rizzonelli (02/2024)

[2] At this point, a systematic analysis was undertaken to ascertain whether interactive behaviors tended to occur in a specific sequence. Whenever an interactive behavior (“near-simultaneous, similar or opposite behavior,” “turn-taking,” “residual interactive category”) occurred in A+B, the three intervals (thirty seconds) preceding that behavior were observed for A, B, and A+B. The goal was to identify sequences of individual (A, B) or collective (A+B) behavior that repeatedly led to interactive behavior. This procedure allowed us to identify the specific roles of each actant that preceded interactive behaviors. However, the patterns that we identified

both for individual (A, B) and for collective (A+B) behavior did not prove very informative, as in most cases the sequence leading to interactive behavior “x” consisted itself of behavior/s “x.”

- [3]]For this reason, an alternative strategy was undertaken: after two independent observers had annotated each session using the codes shown in tables 2.1 and 2.2, all segments of the A+B level where observers had agreed on the code “near-simultaneous, similar or opposite behavior” were further annotated to identify mirroring. This is based on our informal observation, carried out parallel to [1] and [2], indicating that in many cases observable interactive behaviors are characterized as mirroring. The concept of mirroring as used in the present study makes reference to Donald’s²⁷ distinction between mimicry and imitation. Mimicry is a type of copying, involving a replication of the observed movement as closely as possible—either congruently (right-right or left-left) or inversely (right-left). In contrast, imitation is characterized by the fact that exact replication is not required. We use the term “mirroring” as a synonym for Donald’s “mimicry.”

- [a]]To facilitate annotation, at first we distinguished three types of mirroring based on its duration: “interaction without mirroring,” “partial mirroring” (shorter than five seconds), “near-complete mirroring” (longer than five seconds).
- [b]]After annotating mirroring, a comparison was carried out between the segments where “near-complete mirroring” had been annotated with IOA and the segments for which empirical codes overlapped (as described at the beginning of this section). This was conducted to verify whether there was a relationship between mirroring coded for the couple as a unit (A+B) and overlaps of behavior coded for each participant individually (A, B). Since informal observations revealed that there is a phase of the interaction (mostly in the second half of each session) in which mirroring occurred in a smoother, more repetitive, choreomusically and continuously varied way, we undertook a more detailed analysis, taking into account not only the duration of the mirroring episodes, but also movement fluidity, directedness towards the partner and IOA regarding the annotation of the two participants separately (A, B).

27 Cf. Merlin Donald, *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition* (Cambridge, MA, and London: Harvard University Press, 1991).

- [4] In parallel with the structured observation of functional categories, informal observation was conducted iteratively to gain insight into how interaction developed on a larger scale (i.e., not based on a ten-second fragmentation of the behavioral flow but considering macro-phases of the whole interaction).

Results

In the following, results will be described reflecting the numbered steps outlined in the previous section.

- [1] Even in the first phase of informal observation, it was necessary to refine the codes. In fact, insufficient IOA compelled us to formulate more specific definitions based on a list of necessary conditions for each code (as shown in tables 2.1 and 2.2). Moreover, one of the codes that had originally been proposed, “withdraws from the interaction,” was rejected because it was not possible to systematically distinguish it from the code “is not directed towards partner.”
- [a] We applied a strategy of starting only with the codes related to non-interaction (e.g., “is not directed towards partner” for A and B and “no interaction” for A+B), to verify if sufficient IOA could be achieved, at least with respect to these categories. This was not the case, because IOA in ELAN is calculated by means of the modified Cohen’s kappa coefficient,²⁸ for which segmentation and overlap are crucial. Additionally, even though independent observers generally agreed about where to annotate non-interactive events, they disagreed significantly about how to segment them.
- [b] Implementing interval recording with ten-second blocks allowed us to avoid disagreement on segmentation. To comply with this strategy, however, it was necessary to reject another initially proposed code, “uses communicative gestures,” because these kinds of gestures—understood as movements intended to convey a message to the partner, such as a shoulder shrug or head shake—are always much shorter than ten seconds. Even for interval recording, we applied the strategy of

28 Cf. Henning Holle and Robert Rein, “EasyDIAG: A tool for easy determination of interrater agreement,” in *Behavior Research Methods* 47/3 (2014): 1–25, accessed 10 January 2023, <https://www.doi.org/10.3758/s13428-014-0506-7>.

starting with only one code (“near-simultaneous, similar or opposite behavior”), achieving agreement for around seventy-five percent of the annotations in the first two tests. The analysis of the complete coding of the four couples considered in the present study yielded an average agreement of around seventy-three percent.

- [2]]Analysis of the sequences leading to interactive behaviors showed that the most common pattern preceding all interactive behaviors taken into account (“near-simultaneous, similar or opposite behavior,” “turn-taking,” “residual interactive category”), was a thirty-second block of either “becomes initiator within the interactive context” for individual participants (A, B) or “near-simultaneous, similar or opposite behavior” for the couple as a unit (A+B). It was not possible, therefore, to detect an informative pattern of behaviors leading to interaction.
- [3]]
- [a]]The annotation of mirroring categories based on duration yielded an average agreement of around eighty-eight percent.
- [b]]The comparison between “near-complete mirroring” annotated with IOA and overlaps of individual behaviors—as used in a previous analysis—showed that there was a systematic correspondence between the former and the latter. In other words, all segments where mirroring was long and clearly observable were also segments where both participants were exhibiting the same (overlapping) behavior, either with their whole body or with their arms, or with both (see figure 2). This confirms the relevance of the overlapping empirical codes and proves the logical continuity between empirical and functional categories. Moreover, a further analysis of mirroring revealed that A and B’s roles tended to become more evident (thus generating higher IOA) when interaction tended to become more systematic, generally in the second half of each session.
- [4]]Informal observation conducted iteratively on a timescale larger than ten seconds led to the identification of two main categories: “exploratory phase,” a stage of interaction with the partner in which gestures and movements are tested, tried, and explored, often with pauses, uncertainties, and a general lack of confidence towards a) the system, b) a joint affordance offered by the Sentire environment, and c) a situation convention; and “systematic interaction phase,” a more structured stage in which a) interaction develops far more smoothly and with fewer or no interruptions, and b) participants are fully directed towards each other with eye contact, and show repeated and choreomusically varied similar behaviors, including more mirroring and residual interactive behaviors

than during the exploratory phase. The superimposition of this informal macro-analysis and the final step of the previously described structured micro-analysis (i.e., the annotation of mirroring, [3]) is intended to clarify whether and to what extent mirroring episodes appear differently in exploratory versus systematic interaction phases. One of our findings was that in the systematic interaction phases, mirroring occurs more frequently and in a higher-quality form (i.e., more synchronous, more fluid and more partner-directed).



Figure 2: “Fast vollständig” is German for “near-complete.” For all segments where “near-complete mirroring” was annotated, there was an overlap of empirical codes as well. Session numbers are displayed above the examples. © Marta Rizzonelli (02/2024)

Building on this last point [4], several structured observations can be made. The exploratory phase (or phases, as there can be more than one, for example with increasing directedness towards the partner) demonstrates two general features of mirroring behaviors: 1) both “partial mirroring” and “near-complete mirroring” tend to occur less frequently (around thirty percent) than in the subsequent systematic exploration phase (around fifty percent), and when they

do occur, they are generally of brief duration (five, ten or twenty seconds); 2) in the individual analysis (A, B), “near-complete mirroring” generally corresponds to a leader-follower dynamic.

The systematic exploration phase demonstrates five general features: 1) as already mentioned, both “partial mirroring” and “near-complete mirroring” occur more often (around fifty percent); 2) participants are fully directed towards each other with eye contact; 3) their roles are generally unambiguous, as shown by IOA related to the annotation of individual participants (A, B), which corresponds to seventy-nine percent on average. 4) Moreover, within the systematic interaction phase there are two sub-phases: first sub-phase, with a general oscillation between mirroring behaviors (more common) and residual interactive behaviors, and with several cases of clear leader-follower dynamic (around twenty percent on average); second sub-phase, just before the end of the session, with a block of mirroring, either the longest or the second longest of the session, in which both participants are proactive (rather than exhibiting a leader-follower dynamic) and tend towards perfect mirroring.²⁹ 5) Finally, systematic interaction exhibits more movement fluidity compared to the exploratory phase (i.e., movements occur continuously and almost without breaks between the two interactants). This last aspect, however, was observed exclusively on an informal level; systematic evidence would need to be provided by a further study with structured observation.

Systematizing the observations made on exploratory and systematic interaction phases can be achieved by classifying interindividual synchrony in four successive levels of increasing complexity. The first level corresponds to interindividual synchrony that is understood as behaviorally matched (interindividual) synchrony, and does not rely on cooperative social interaction. The simple overlap (or match) of two interactants’ behaviors that we observed through empirical categories in previous studies does not *per se* indicate cooperative social interaction. A second level is what we observed in the exploratory phase/s: here, the interactants’ movements do not flow seamlessly from one to the other, but are interrupted by brief pauses and often performed in a fragmentary fashion. However, mirroring behaviors occur repeatedly (around thirty percent of the observed time, as reported above), either in a partial or in a near-complete form; most importantly, interactants are directed towards each other, even if only in a partial way (i.e., with fragmentary eye

29 Under “perfect mirroring” we understand here a perfect copy of the partner’s behavior; this differs from “near-complete mirroring,” which is identified based on its duration (longer than five seconds), as clarified in section 2.2.

contact). This directedness (albeit partial) is what allows us to infer interaction in a functional sense.

The third, more complex level of interindividual synchrony can usually be observed in the first part of the systematic interaction phase. Movement fluidity increases and the interactants' roles become less ambiguous—that is, it becomes clearer whether a person takes initiative or follows their partner. Mirroring behaviors tend to occur more often than in the previous level, with a slight tendency towards (near-) complete forms of mirroring. Forms of interaction other than mirroring, which we call residual, often occur between or prior to mirroring behaviors and contribute significantly to the interactional flow. More importantly, directedness becomes more comprehensive, involving eye contact and often multiple body parts. Altogether, this form of interindividual synchrony exhibits fluidity, repetitive and choreomusically varied mirroring and directedness, indicating a high level of cooperative social interaction. Finally, the fourth level of interaction, corresponding to the last part of the systematic interaction phase, displays usually long-lasting, symmetric mirroring behaviors where both interactants simultaneously begin to take initiative. Here, a situation convention seems to have been established, allowing interaction to flow according to tacit rules that were not available in the previous phases of the session.

Discussion

The results show that interactions mediated by Sentire involve an exploratory phase/s and a systematic phase, which both include synchronous mirroring behaviors. However, mirroring behaviors emerge through different processes. A low-level of interindividual synchrony could be identified in the exploratory phase/s, consisting of both mirroring and non-mirroring behaviors which occur even when two persons are not fully directed towards each other (i.e., they do not make eye contact). Here, action-based mirroring goes beyond the mirroring that occurs—for instance, when two persons are merely exposed to the same auditory or visual milieu—because in this case participants are explicitly invited to interact with each other in a shared environment and social context. However, action-based mirroring occurring in the exploratory phase/s is not yet based on a “situation convention”—that is, “a socially emergent anticipatory agreement”³⁰—but on the simultaneous affordance that allows two persons to

30 Robert Mirski and Mark H. Bickhard, “Conventional Minds: An Interactivist Perspective on Social Cognition and Its Enculturation,” in *New Ideas in*

behave in relation to the common features of the environment that they perceive, while being neither fully directed towards nor closely related to each other. Interactive behaviors including mirroring and non-mirroring behaviors that are observed in this level consist of fragmentary single movements, or a series of movements which occur either only once or several times but with breaks, without being fully directed towards the partner.

On the contrary, mirroring behaviors in a higher-level of interindividual synchrony, which could be identified in the first part of the systematic interaction phase, are observed when interactants are fully directed towards each other, making eye contact. Those behaviors are often followed by residual (non-mirroring) interactive behaviors—or, in few cases, by non-interactive behaviors which quickly become interactive (mirroring or residual) again. Interactants' behaviors in this level are based on their previous interactions during the exploratory phase; these interactions allow interactants to perceive a joint affordance offered by the Sentine environment. This joint affordance serves as a basis for interactants' collaboration, although their behaviors might be different; the Sentine environment provides some degree of freedom as to how each interactant can behave. Nevertheless, each behavior contributes to a collective act, involving a joint commitment. Synchronous mirroring behaviors that emerge within that level of interindividual synchrony could be interpreted as a process of mutual adaptation giving rise to a situation convention. In this process, interactive behaviors characterized as partial mirroring or residual interactive behaviors tend towards leader-follower dynamics, whereas in interactive behaviors categorized as near-complete mirroring, both actants tend to be simultaneously proactive. This allows us to pose a research question for a further observational study: whether near-complete mirroring behaviors where both actants become simultaneously proactive emerge as a result of established anticipatory agreement during an improvisatory interaction in the Sentine environment.

Remarkably, every session ends with relatively long segments in which long-lasting, near-complete mirroring behaviors are simultaneously initiated by both interactants, which we characterize as the highest level of interindividual synchrony. Those mirroring behaviors differ not only from fragmentary synchronous mirroring behaviors observed in the very beginning of improvisation, which are not based on an established situation convention at all, but also from those observed in the exploratory and first systematic interaction phases, which involve a process of establishing a

situation convention while interactants incrementally choose their behavior (influenced by others' decisions), and mutually adapt their behaviors. These interactions give rise to a socially emergent anticipatory agreement, which serves as a basis for joint decision making.³¹ Near-complete, symmetric mirroring behaviors that are displayed towards the end of each session can be interpreted as those that emerge as soon as a situation convention concerning a dyadic interaction within the Sentire environment is established.

Taking into account the role of past interactions in the emergence of an anticipatory agreement, we may state that our behavioral sequential approach to improvisatory musical interaction allows us to identify a sequential order of behaviors. Some limitations of the present study need to be addressed. Due to the time-consuming, iterative process of structured observation, it was only possible to analyze a few sessions. The analysis led to several specific research questions, some of which would have required a more thorough investigation and the development of further steps of structured observation. Despite these limitations, our in-depth analysis of sequential behaviors on a moment-to-moment basis showed that some features of the interindividual synchronous behaviors observed during each session are distinguishable from one another, and result from a degree of continuous mutual adaptation which unfolds parallel to musical improvisation.

Contrary to some scholars' claim that synchrony or mirroring alone is insufficient for understanding the joint action that occurs when interactants are dynamically coupled and follow a common goal,³² the present study showed that both residual interactive behaviors and synchronous mirroring behaviors can be characterized as dynamically coupled rather than simply aligned, depending on the process of mutual adaptation. As a result, we can rethink the roles of basic social behaviors that have long been disregarded in social sciences and social cognition (albeit more recently acknowledged in an increasing

31 Cf. Uri Hasson, Asif A. Ghazanfar, Bruno Galantucci, Simon Garrod, and Christian Keysers, "Brain-to-brain Coupling: a Mechanism for Creating and Sharing a Social World," in *Trends in Cognitive Sciences* 16/2, 114.121 (2012): 114–21, accessed 15 September 2022, <https://www.doi.org/10.1016/j.tics.2011.12.007>.

32 Cf. Riitta Hari, Tommi Himberg, Lauri Nummenmaa, Matti Hämäläinen, and Lauri Parkkonen, "Synchrony of Brains and Bodies During Implicit Interpersonal Interaction," in *Trends in Cognitive Sciences* 17/3 (2013): 105–6, accessed 15 September 2022, <https://www.doi.org/10.1016/j.tics.2013.01.003> and Uri Hasson and Chris D. Frith, "Mirroring and Beyond: Coupled Dynamics as a Generalized Framework for Modelling Social Interactions," in *Philosophical Transactions of the Royal Society B: Biological Sciences* 371:1693, 20150366 (2016), accessed 15 September 2022, <https://www.doi.org/10.1098/rstb.2015.0366>.

number of studies on infant-caregiver interaction and basic interindividual behaviors underlying social interaction). The present study shows that synchronous mirroring behaviors not only serve as a prerequisite for a higher-order form of social interaction, but emerge as central interindividual behaviors during an interaction involving joint commitments and we-intentions.

This result could provide a fruitful perspective for future research on social cognition, in which non-linguistic forms of interaction that neither bear representational semantics nor involve given social norms could be better integrated into empirical studies on social cognition. A question to be addressed would be whether and to what extent interindividual synchrony not only underlies cooperative social interactions, but also emerges as a systematic form of behavior while mutual adaptation occurs during interactions. In particular, musically oriented interindividual behaviors that can be investigated in the context of cooperative social interaction could help assess whether a sequence of interindividual behaviors that emerges towards the end of that interaction could be seen as a result of the mutual adaptation that occurs during different exploratory and systematic interaction phases, establishing a situation convention. This is due to the fact that a musically meaningful unit of behavior that does not bear representational semantics but is considered to make sense in a given musical context emerges through the relation of each behavioral event to another, and cannot therefore be reduced to a single event.³³ A linguistic-semantic unit, on the other hand, which has a representational reference to the world, can be assigned to a discrete linguistic event (albeit also related to other linguistic events).

Of particular note is the fact that interactions previously observed in free musical improvisation could be limited to those that only take place in a given context of improvisation, since past interactions going beyond that context cannot be presupposed. The field of free musical improvisation is therefore appropriate for investigating how interindividual synchrony emerges due to a dynamic and mutual process of anticipation and intermediate-term memory, and not necessarily relying on explicit knowledge that is available outside of an interactional context and long-term memory. This would allow scholars to investigate whether cultural norms could be established through musical

33 Cf. Jin Hyun Kim, "From the Body Image to the Body Schema, From the Proximal to the Distal: Embodied Musical Activity Toward Learning Instrumental Musical Skills," in *Frontiers in Psychology* 11:101 (2020), accessed 15 September 2022, <https://www.doi.org/10.3389/fpsyg.2020.00101>.

interindividual interaction involving synchronous behaviors that could be conceived of as synergetic interactive behaviors.³⁴

34 E.g., Riccardo Fusaroli and Kristian Tylén, “Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance,” in *Cognitive Science* 40 (2016): 145–71, accessed 15 September 2022, <https://www.doi.org/10.1111/cogs.12251> and Hasson and Frith, “Mirroring and Beyond.”