Interpersonal Synchrony : A Survey Of Evaluation Methods Across Disciplines

Emilie Delaherche, Mohamed Chetouani, Ammar Mahdhaoui, Catherine Saint-Georges, Sylvie Viaux, and David Cohen

Abstract—Synchrony refers to individuals' temporal coordination during social interactions. The analysis of this phenomenon is complex, requiring the perception and integration of multimodal communicative signals. The evaluation of synchrony has received multidisciplinary attention because of its role in early development, language learning and social connection. Originally studied by developmental psychologists, synchrony has now captured the interest of researchers in such fields as social signal processing, robotics and machine learning. This paper emphasizes the current questions asked by synchrony evaluation and the state-of-the-art related methods. First, we present definitions and functions of synchrony in youth and adulthood. Next, we review the non-computational and computational approaches of annotating, evaluating and modeling interactional synchrony. Finally, the current limitations and future research directions in the fields of developmental robotics, social robotics and clinical studies are discussed.

Index Terms—Synchrony evaluation, coordination, computational model.

✦

1 INTRODUCTION

1.1 Multidisciplinary need for synchrony assessment

The need for a multidisciplinary approach to assessing synchrony is evident at the interface of social signal processing, computational neurosciences, developmental psychology and child psychiatry. Synchrony is a complex

- E. Delaherche and M. Chetouani are with the Institute of Intelligent Systems and Robotics, University Pierre and Marie Curie, 75005 Paris, France (e-mail: emilie.delaherche@isir.upmc.fr; mohamed.chetouani@upmc.fr).

- A. Mahdhaoui was with the Institute of Intelligent Systems and Robotics, University Pierre and Marie Curie, 75005 Paris, France and is currently with Orange Labs R&D, France Telecom, Meylan Grenoble (e-mail:ammar.mahdhaoui@gmail.com).

- S. Viaux is with the Department of Child and Adolescent Psychiatry, Hôpital de la Pitié-Salpêtrière, University Pierre and Marie Curie, 75013 Paris, France (e-mail: sylvie.viaux@psl.aphp.fr).

- C. Saint-Georges and D. Cohen are with the Department of Child and Adolescent Psychiatry, Hôpital de la Pitié-Salpêtrière, University Pierre and Marie Curie, 75013 Paris, France, and also with the Institute of Intelligent Systems and Robotics, University Pierre and Marie Curie, 75005 Paris, France (e-mail:catherine.saintgeorges@gmail.com; david.cohen@psl.aphp.fr).

phenomenon requiring the perception and understanding of social and communicative signals and continuous adaptation. The implementation of interactive algorithms for complex tasks in human-machine interfaces requires a better understanding of human interaction regulation strategies, especially synchrony [1]. Rapport building, the smoothness of a social encounter and cooperation efficiency are closely linked to the ability to synchronize with a partner. The close link between synchrony and interaction quality bears promising perspectives for researchers building social interfaces, robots or Embodied Conversational Agents [2], [3], [4].

In addition, a lack of automatic tools for studying synchrony has limited the exploration of psychiatric conditions that affect social abilities, whether permanently (e.g., autism) or temporarily (e.g., major depression), in terms of interactive abilities. The study of interaction and intersubjectivity in infants is crucial, but currently, no commonly accepted method exists for detecting and assessing synchrony and dyssynchrony between interactive partners during early pathological development.

Recently, Meltzoff et al. described how research in developmental psychology may provide a good opportunity for enhancing computational models of such phenomena and vice-versa [5]. In particular, the mechanisms of social learning have interested researchers in the field of developmental robotics, in which the long-term goal is building robots that, like infants, learn through observation, imitation and synchronized exchanges.

1.2 Scope of this paper

Synchrony is difficult to define and delimit. Numerous terms have been used in the literature to describe the interdependence of dyadic partners' behaviors (mimicry, social resonance, coordination, synchrony, attunement, chameleon effect, etc.). Moreover, several concepts are closely related to synchrony or are prerequisites of synchrony, such as turn-taking and mutual attention. The first goal of this study is to clarify the concept of synchrony and its functions in both early infancy and adulthood.

While developmental studies have shown that babies have an early ability to detect disruption in interactional synchrony [6], a method for the objective evaluation of synchrony remains somewhat elusive despite being heavily targeted by researchers. In this study, we give a multidisciplinary overview of synchrony evaluation of human-human interactions. We will first present non-computational methods, which have been primarily developed by psychologists to evaluate interactional synchrony. We will then review recent advances in computational science efforts to capture pertinent information from behavior-coded databases and to directly model synchrony using low-level signals.

Finally, we will present the current limitations of computational methods and prospects for synchrony assessment research.

2 DEFINITIONS AND RELATED NOTIONS

2.1 Related theories and definitions

The concept of synchrony is complex, and the first step in its study is to define synchrony in relation to similar concepts. The study of synchrony is inextricably related to the study of communicative interaction and language. According to theories of dialog, conversation is a joint activity that requires coordination at two levels: content and process [7].

2.1.1 Content coordination

At the content level, conversational partners must coordinate what is being said and reach a common understanding. This common understanding is achieved by aligning the partners' situation models, "which are multi-dimensional representations containing information about space, time, causality, intentionality and currently relevant individuals" [8], [9]. According to Pickering and Garrod, this alignment is achieved via a nonconscious mechanism called "interactive alignment", by which partners align their representations at different linguistics levels at the same time. In a cooperative maze game, Garrod and Anderson observed that conversational partners trying to verbalize their position in the maze to a partner tended to use the pragmatical and lexical features of the utterances that they had just encountered [10]. This alignment serves communication efficiency: conversational partners tend to formulate their utterances to minimize the time and effort required for mutual understanding, minimizing their collaborative effort. Interactive

alignment has been observed in dialog at the lexical level (with different speakers repeating the same word to refer to particular objects [10], [11], [12]), at the syntactic level (using the same syntactic structure [13]) and in accent or speech rate [14].

The alignment of these last articulatory dimensions is related to Giles' Communication Accommodation Theory (CAT), evolved from the Speech Accommodation Theory (SAT), which addresses the tendency to unconsciously minimize or emphasize differences in speech, vocal patterns or gestures when interacting with a conversational partner [14]. CAT encompasses a variety of features, including accent [15], [16]; speech rate [17]; utterance length [18]; response latency [19]; pausing frequency and length [18], [19]; laughter [18] and postures [20]. CAT also focuses on the intercultural, interpersonal, psychological, social and contextual factors that modify communicative behaviors. CAT concerns both convergence and divergence, depending on whether the strategy is to minimize or maximize the speech pattern differences between the speaker and the conversational partner. Conversational partners often do not converge at all levels at any given time; they converge at some levels and diverge at others.

Behavior matching [21]; mirroring; mimicry [22], [23], [24]; congruence and the chameleon effect [25] are related to convergence. These concepts concern non-verbal communicative behaviors, such as postures, mannerisms or facial displays, and indicate similar, simultaneous behaviors by both social partners; the analyzed features are static and qualitative.

The automatic triggering of many social behaviors by the perception of action in others has also been studied in the neurophysiological literature: motor imitation arises from the firing of mirror neurons in macaque monkeys [26] and the premotor cortices, originally considered to be exclusively concerned with motor control, activate during the observation of actions, in the absence of any action execution, in humans [27]. The mirror neurons are an example of a more general mechanism: the neuronal structures involved when a mental state is experienced (internal representation of an action or sensation) are also used when perceiving others experiencing the same mental state. This mechanism also applies to emotion contagion: Lundqvist found that participants produced different EMG (Electromyography) patterns depending on the facial expressions that they observed in emotionally colored photographs [28]. For instance, participants showed increased muscular activity over the zygomaticus major (cheek) muscle region when facing happy facial expressions. Moreover, the motor mimicry was too subtle to be perceived visually.

2.1.2 Process coordination

At the process level, conversational partners are able to accurately predict the beginnings and endings of conversation phases, which are marked by syntax, morphology and intonation [7]. By accurately projecting the ending of the speaker's turn, the listener can begin his turn using

correct timing, allowing the conversational partners to achieve synchrony.

Bernieri et al. define synchrony as "the degree to which the behaviors in an interaction are non-random, patterned or synchronized in both form and timing" [21]. Synchrony refers to the temporal coordination between individuals. For Cappella, "Coordination is arguably the essential characteristic of every interpersonal interaction...Interpersonal communication requires the coordination of behavior" [29]. Synchrony is related to the adaptation of one individual to the rhythms and movements of the interaction partner [20] and the degree of congruence between the behavioral cycles of engagement and disengagement of two people. Newman and Newman add that synchrony suggests that the social partners move fluidly from one state to the next. In opposition to behavior matching, synchrony is a dynamic phenomenon [30]. Moreover, this synchronistic process may occur in different sensory modalities; for example, the intensity of an infant's physical behavior matches the intensity of the mother's vocal behavior [31].

In terms of interaction dynamics, the conditions for the emergence of synchrony include [32] (1) maintaining a prolonged engagement in mutual attention and turn-taking with both partners "tracking each other"; (2) temporal coordination, the matching of activity levels (body orientation, body movements, facial expressions), similarly to a dance; (3) contingency; (4) attunement, meaning that in infant-adult caregiver interactions, the caregiver senses the infant's state and adjusts accordingly. Related terms in the literature include contingency, mutual responsiveness, mutual adaptation reciprocity, mutuality, affect attunement, dyadic affect regulation, interactional synchrony, dyadic synchrony, and behavioral entrainment. The definitions of synchrony are often circular and certainly too vague to use in a computational method; the types of behaviors and patterns and the scale of the phenomenon are not specified.

2.2 Proposed definition of synchrony

For the authors, synchrony is the *dynamic* and reciprocal adaptation of the temporal structure of *behaviors* between *interactive* partners. Unlike mirroring or mimicry, synchrony is *dynamic* in the sense that the important element is the timing, rather than the nature of the behaviors. Taking the floor at the appropriate time and grasping an object being held out are matters of synchrony. As noted in [33], the distinction between synchrony and mirroring can be unclear; these phenomena are not disjunctive and can often be observed simultaneously. For instance, two people sitting with crossed legs or looking in the same direction are exhibiting either mirroring or the chameleon effect. This behavior becomes a matter of synchrony if they cross or uncross their legs at the same time or gaze in the same direction simultaneously.

These actions are a coordination of *behaviors* in the sense that they are multimodal (different modalities intervene at the same time). In this respect, synchrony differs from

alignment, mirroring or the chameleon effect in which the adaptation occurs in the same modality for the two partners. To grasp an object, a person must simultaneously follow the object visually and reach out his arm. Such actions are also intermodal (the coordination intervenes across modalities), such as nodding one's head to indicate agreement with what is being said. In this work, behavior refers to communicative verbal and non-verbal behavior (gestures, postures, facial displays, head gestures, etc.).

Finally, synchrony can occur in all *interactive* context: cooperative (playing a piece of music in duo) or not cooperative activities (fighting), linguistic (telephone conversations) or not linguistic interaction (catching a ball). We argue that synchrony entails interaction.

Given this definition, for each behavior produced by one partner, there is a limited window of time for the other partner to produce a coordinated behavior. Thus, when computing the coordination of two distinct behaviors, the size of the temporal window should be very limited and dependent on the duration of the participatory actions of each partner. If the activity of interest is playing catch, there is no need to study the catcher's coordination if the ball has fallen to the ground or if the next throw has begun. For natural conversation, the fluency of turn-taking is considered turn by turn. This fact does not mean that coordination cannot evolve during an interaction. Partners playing catch will likely become better coordinated with practice, and conversational partners may accommodate their speech style over the course of their encounter, which will help to smooth their turn transitions.

3 FUNCTIONS OF SYNCHRONY

3.1 Functions in early infancy

There is likely much that is not yet understood regarding the role of synchrony during early development. Some important functions have been highlighted. First, synchrony appears to be involved in the co-regulation of affective states [34], [35]; that is, a "process through which the mother and infant match each other's affective states within lags of seconds jointly moderating the level of positive arousal". Mothers tend to use this mechanism to maintain and regulate the exchanges with their infant during face-to-face interaction. Through these synchronized exchanges, the mother can smoothly move the infant from one state to another. In other words, synchrony facilitates the interaction, promotes openness between mother and child, and enhances the degree of presence in a gathering [36], [37], [38].

Second, synchrony seems to improve the infant's experience of effectance and social connection [32]. From the experience of dyadic synchrony, the infant gains a feeling that the interaction cycles are completed. When the interaction cycles are interrupted and then re-established, "the infant's sense of confidence in his ability to self-regulate and engage others effectively" is enhanced [32]. Newman and Newman describe this mechanism as follows: "Long before infants can use language to convey

feelings or needs, they experience the satisfaction of social connections through these cycles of communication. They do not rely on spoken language but on the many emotional cues that arise from rhythmic patterns of breathing, facial expressions, tone of voice, touch, and eye gaze. As the mother and baby move into renewed moments of coordination, their sense of pleasure increases, leaving a memory of such moments to guide future conversations" [30].

Third, given what has been said previously, synchrony should facilitate secure attachment. The increasing number of children with insecure attachment and behavioral symptoms after exposure during infancy to a mother with disruptive behaviors [39], depression [40], or social deprivation [41] substantiates the importance of synchrony for adequate emotional child development.

Fourth, synchrony also plays a role during language acquisition. Empirical evidence for probabilistic or statistical learning has matured in the fields of auditory and visual inputs [42]. In language acquisition, cultural factors are crucial for both oral and written languages; however, cultural influences on oral and written languages develop in radically different ways. Oral language develops "spontaneously" unless the child is deprived of language exposure. Saffran et al. [43] and Kuhl and colleagues [44], [45], [46] investigated the role of exposure to a given language by highlighting a statistical learning process and specialization in native languages (e.g., the magnet effect) [47]. Simple exposure, however, does not explain language learning. In both speech production and perception, the presence of an adult interacting with a child strongly influences learning [48].

Finally, imitation, which has been widely studied in developmental psychology, should be mentioned [49], [50]. Imitation can be defined as a motor or verbal act that is similar to a motor or verbal act previously initiated by a model. At first, imitation is a means to learn by observation and replication (observational learning). Imitation also helps the child construct a social code and replicate what he has observed in adequate situations. Next, imitation is a means to communicate for as long as the child cannot speak. The child learns to communicate at first with various forms of imitation, such as symbolic play, which occurs when the child begins to substitute one object for another and to represent this object in a fictive world (for instance, riding a chair as a horse), or postponed imitation (when the model of imitation is absent). Until approximately two years of age, the child does not speak and resorts to imitation to interact with his peers. Imitation tends to disappear with the acquisition of language.

3.2 Functions in adulthood

In adulthood, interactional synchrony has been shown to act as a facilitator to smooth social interactions, to achieve "coordination of expectancies among participants" [36]. Non-verbal synchrony also plays a role in building rapport among individuals [51], [24]. In their study of non-conscious mimicry, Chartrand et al. established a link between the degree of mimicry, the perception of interaction smoothness and the degree of liking between interaction partners [25]. Moreover, they showed that not all individuals share the same dispositions to imitate their partners and that empathic persons had a greater tendency to produce nonconscious mimicry.

A link has also been established between the degree of synchrony and how dyadic partners are perceived [52]. For instance, Lakens et al. manipulated movement rhythms of stick figures and asked judges to evaluate the perceived entitativity (i.e., the unity, the emergence of a social unit). He demonstrated a linear relationship between the differences of movement rhythms and the perception of entitativity [53].

Executing a task in synchrony seems to promote cooperation between individuals [54] and to enhance memory of interaction partner's utterances and face appearance [55]. More, in-phase coordination was shown to promote memory of interaction partner's utterances over self utterances [56]. Interestingly, Ramseyer et al. investigating non-verbal synchrony between patient and therapist during psychotherapy sessions evidenced that non-verbal synchrony was associated with therapy outcome and patient's view of the therapy process [57]. They found that synchrony was increased in sessions rated by patients as manifesting high relationship quality, and in patients experiencing high self-efficacy. Furthermore, higher non-verbal synchrony characterized psychotherapies with higher symptom reduction.

Bouhuys et al. found that a lack of coordination in non-verbal behaviors constitutes a risk factor for depression recurrence [58]. In addition, Bird et al., comparing adults with Autism Spectrum Disorders (ASD) to controls in an imitation task, found that ASD individuals responded faster to robotic hands, whereas the comparison group responded faster to human hands [59].

4 NON-COMPUTATIONAL METHODS OF SYNCHRONY ASSESSMENT

In the earliest days of synchrony research, instances of synchrony were directly perceived in the data by trained observers. Several methods have been proposed to evaluate interactional synchrony, ranging from behavior micro-analysis to global perception of synchrony.

Behavioral coding methods propose evaluating the behavior of each interactional partner on a local scale. These methods require the use of computer-based coding (e.g., Observer ® or Anvil [60]) and trained raters. Various category and time scales can be used for coding. In [29], Cappella synthesized the three crucial questions to be addressed when conducting an interaction study: "what to observe (coding), how to represent observations (data representations) and when and how frequently to make the observations (time)". For instance, Condon and Sander as well as Cappella [31], [61] proposed analyzing micro-units of behavior. They annotated the speech segments and the direction of movement of different body parts (head,

eyes, mouth, elbows, trunk, shoulder, wrists and fingers). At a higher scale, some grids directly analyze interactive behaviors (smiles, gazes, illustrative gestures, adaptors, head gestures (nods, shakes)) or functional states (alertness, orientation between the partners, communicative expression, emotion, body contact and postural tension) [62], [63]. Generally, a measure of synchrony is deduced from the covariation of the annotated behaviors. The codes can be either continuous (speed of a gesture) or categorical (type of gesture). This type of grid has been widely used for coding home movies. Coding home movies is particularly complex because of the naturalistic setting and the varying quality of the films. This approach has been largely used in the field of autism to improve our understanding of the early developmental course of children who will be eventually diagnosed with ASD [64]. This knowledge is needed to better understand the complex pathogenic phenomena of autism and to improve the early screening and management of autism.

Behavioral coding methods are time-consuming and tedious with regard to the training of observers, the number of behaviors coded and the duration of the video files to be coded, particularly for longitudinal studies. Cappella [61] and Bernieri et al. [65] proposed an alternative to behavior micro-analysis: the judgment method. In their studies, they investigated the use of human raters to evaluate video clips of infants interacting with their mothers. Raters judge for simultaneous movement, tempo similarity and coordination and smoothness on a longer time scale using a Likert scale. Cappella showed that untrained judges were consistent with one another and reliably judged the synchrony between partners [61]. Nevertheless, considering the complexity of the underlying phenomenon, the reliability of the coders' judgment can be problematic. This question can be partially circumvented by the use of a scale of several items to test a given construct and several judges.

Another method is the Coding Interactive Behavior (CIB) [66], a well validated system for coding mother-infant interactions requiring trained observers [34], [67]. The CIB is a global rating system of parent-child interaction that contains both microlevel codes and global rating scales. Codes are averaged into six composites (maternal sensitivity, mother intrusiveness, limits, mother-infant positive affect, infant involvement, and negative dyadic status) that are theoretically derived and address several aspects of the early mother-infant relationship, showing acceptable to high internal consistency [66]. The coding of feeding interactions has been shown to differentiate infants diagnosed with psychiatric disorders in infancy [68], [69] or prematurity [70]. The CIB was also used in studies assessing the effect on early childhood development and interactive behavior of various maternal/parental styles, such as breastfeeding [71], skin-to-skin contact (kangaroo care) [72] and parent or child gender [73].

Non-computational methods suffer serious drawbacks. Within the tedious task of coding, segmenting and annotating behaviors can be confusing: when does a behavior start, when does it end, how should it be labeled? Often, the annotator makes trade-off because no label accurately describes what he observes. The judges' reliability in assessing such a subjective and complex construct is also questionable, and no general framework for synchrony assessment has been accepted to date. A method was recently proposed to convert the judgments of multiple annotators in a study on dominance into a machine-learning framework [74]. Finally, conversational partners are often studied individually when coding. Thus, it is particularly difficult to recreate the dynamic and interpersonal aspects of social interaction manually and after coding. Nonetheless, annotation and judgment methods are essential in proposing automatic systems for synchrony assessment and testing their performance. Currently, no automatic systems modeling synchrony using real interaction data are free from annotation. Annotation is mainly used in two different manners. First, annotation is used to train automatic systems to model and learn communication dynamics (see section 6). These studies often rely on behavioral coded databases. Second, another set of studies intends to measure the degree of synchrony between dyadic partners with unsupervised methods (see section 5). In these studies, the measure of synchrony is not validated per se but is judged by its ability to predict an outcome variable that has been manually annotated, often using judgment methods. The outcome variable can be friendship [75], conflicting situations [75], success in psychotherapy [57], etc.

5 FULLY AUTOMATIC MEASURES OF MOVEMENT SYNCHRONY

To avoid tedious coding, automatic techniques can be used to capture pertinent social signals and assess movement synchrony in human-human interactions. The studies reviewed in this section aim to measure the degree of similarity between the dynamics of the non-verbal behaviors of dyadic partners. The goals of these studies are generally divisible into two categories: to compare the degree of synchrony under different conditions (e.g., with or without visual feedback) and to study the correlation between the degree of synchrony and an outcome variable (e.g., friendship, relationship quality). Consequently, these methods are mostly unsupervised in the sense that the measure of synchrony is not validated per se; rather, the ability of the measure to predict the outcome variable or to discriminate the different conditions is important. The methods described in this section were applied to adult-adult and child-adult interactions (Tables 1, 2, 3). In this section we describe in detail the traditional steps of a computational model to assess synchrony (Fig. 1).

Although our focus is mostly on the study of movement synchrony, many of the methods, issues and findings are similar to the study of such subjects as entrainment or adaptation in spoken language interaction. For instance, Levitan et al. [76] studied global and local measures of entrainment in backchannel-preceding cues based on audio features (intonation, voice quality, pitch, intensity,

duration) and its association with dialog coordination and task success. Benus et al. [77] studied the link between the alignment of turn-taking behavior and the achievement of pragmatic goals. They quantitatively measure the rhythm entrainment between speakers as the latency of the first pitch accent after a turn exchange divided by the rate of pitch accents in the utterance preceding the turn exchange. Finally, [78] proposed using a machine learning algorithm to predict the emotional coloring (valence, activation, power) of an utterance based on the emotional coloring of the previous utterance.
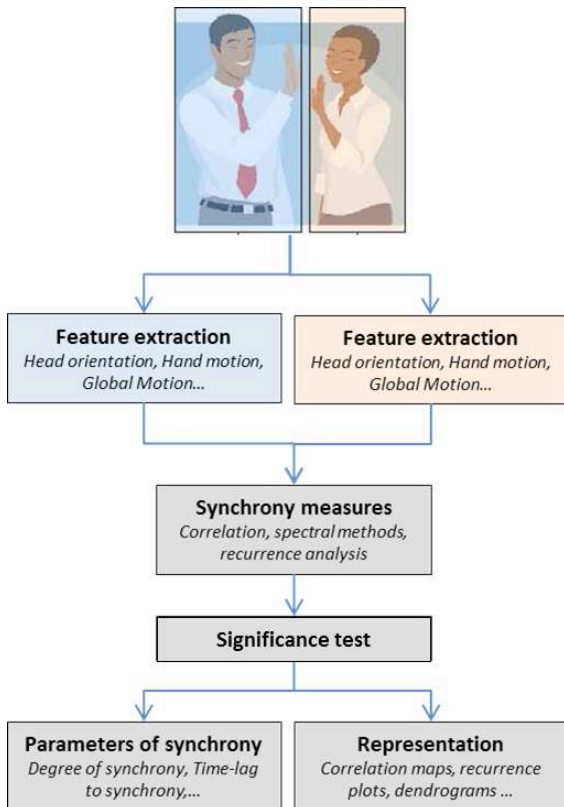

Fig. 1. Synopsis of a synchrony computational model

5.1 Features

The first step in computing synchrony is to extract the relevant features of the dyad's motion. With the exception of Delaherche and Chetouani, who tried to model the coordination between movement features and prosodic features of speech (pitch, energy, pause and vocalic energy) [79], previous studies have focused on unimodal features. We can distinguish between studies focusing on the movement of a single body part and those capturing the overall movement of the dyad. Several acquisition techniques are prominent in the literature: motion-tracking devices, image-processing techniques (tracking algorithms, image differencing) and physiological sensors. Studies on a single body part usually use dedicated motion tracking devices (speaker tongue position [80], finger motion [81], eye movement [82], hand motion

[83], leg motion [84]). Several studies have focused on the coordination between the postural movements of the participants [85], [86], [87], as postural movements can be mediated by a common tempo of verbal interaction.

Numerous studies focus on head motion, which can convey emotion, acknowledgement or active participation in an interaction. Head motion is captured using either a motion-tracking device [88], [89], [90] or a video-based tracking algorithm [91], [92], [93], [94], [95], [96]. Many studies capture the global movements of the participants [97], [75], [98], [79], [33], [99], [100], [101], [57], [102], [103], [104]. Except for Boker and Rotondo [98], who used a motion-tracking device, these studies use a video-based algorithm to evaluate the dyad's motion. Other studies have also focused on the motion of an apparatus being actuated by the participants (swinging pendulum [105], [106], [107], [108], [109], [108] or rocking chair [110]).

5.2 Measures

5.2.1 Correlation

Correlation is certainly the most commonly used method to evaluate interactional movement synchrony. After extracting the movement time series of the interactional partners, a time-lagged cross-correlation is applied between the two time series using short windows of interaction. Several studies also use a peak picking algorithm to estimate the time-lag of the predictive association between two time series (i.e., the peak cross-correlation that is closest to a lag of zero) [88], [89], [75].

A critical question is the choice of the length of the windows of interaction. In the studies reviewed, the length of the window varies from 1 s to 10 min with time-lags of 0 to 5 s. Boker et al. raise the question of the time series stationarity [89]. They compare the cross-correlation between (a) the movements of two dancers synchronized with each other and the rhythm of the music and (b) the head movements of two persons conversing. In (a), there is a stable pattern of synchronization during the entire interaction; the data follow the assumptions for a stationary process and the cross-correlation calculated on the whole interaction is high. In (b), there might be a "high degree of association on short scales, but due to nonstationarity, overall there might be only low values of correlation". Thus, a weak correlation between time series could indicate either little coordination in the conversation or nonstationary short-term coordination.

An additional issue is linked to the representation or meta-parameters extracted from the cross-correlation coefficients. A color-coded correlation map is the most common way to represent cross-correlation coefficients [80], [88], [89], [98], [33], [99], [100], [111], [104]. Time is represented on one axis, and the different time-lags are represented on the other. The correlation strength is represented by different color shades. Correlation maps have the advantage of showing a global snapshot of an interaction. Sequences of high synchrony are easy to identify, and the difference between two dyads can be grasped

immediately. When the time-lag between partners is measured using a peak-picking algorithm, researchers usually plot the evolution of the time-lag over time [88], [89]. Such plots reveal leading-following relationships between the partners and any dominance traits. These representations are particularly useful when only one signal is being studied. When partners are characterized by several features, the relationship must be represented across the different features. Delaherche and Chetouani proposed using dendrograms to characterize the hierarchy in feature similarity [79]: a clustering algorithm was applied to understand how local synchrony was established across all features and to represent their similarity hierarchy. Dendrograms are tree diagrams illustrating the hierarchical relationship between data, often used to represent taxonomy or a hierarchical clustering in biology. U-shaped lines connect features according to their similarity. The height of each U-shaped line represents the distance between the two connected features. The construction of a dendrogram relies on a similarity symmetric matrix containing the distance between every possible pair of features. Dendrograms have two main advantages over the previously described techniques: they characterize the coordination across more than two features (from different modalities) and offer a snapshot of a given window of interaction.

Apart from the representation of synchrony, there is also a need to aggregate the measures from the cross-correlation matrices in synthetic parameters to quantitatively compare different dyads or study the relation between the existence of synchrony and an outcome variable (e.g., smoothness of interaction). Two meta-parameters are traditionally assessed to characterize synchrony between partners. First is the degree of synchrony, the percentage of synchronous sequences. When a peak-picking algorithm is applied, the mean and variance of the correlation coefficient's peak value indicate the strength and the variability, respectively, of the coordination during the interaction. Second, the orientation of synchrony indicates who is driving the interaction. The orientation is measured using the time-lag between synchronous windows of the video [88], [89]. A positive lag between partner 1's features and partner 2's features indicates partner 1 is leading the interaction, while a negative lag indicates partner 2 leading the interaction. A zero lag between each partner features indicates mutual synchrony.

5.2.2 Recurrence analysis

Recurrence analysis was inspired by the theory of coupled dynamical systems (see Table 2), providing graphical representations of the dynamics of coupled systems. Recurrence analysis assesses the points in time that two systems visit similar states, called "recurrence points". These points represent the points in time that the two systems show similar patterns of change or movement. Consider, for example, two time series of numeric measurements. First, time-delayed vectors v of points m are constructed from the time series, where m represents the embedding dimension and t the delay between sequential

time-points. Every vector from the first time-series is compared with every vector from the second time-series using a distance measure (e.g., Euclidean Distance). A cross-recurrence matrix is created at this stage. A threshold on the distance between vectors is fixed to decide whether two vectors are similar. A timepoint $(i; j)$ on the cross-recurrence matrix is set to 1 if the vectors i and j are similar and 0 otherwise. The cross-recurrence plot is the two-dimensional representation of the cross-recurrence matrix.

Webber and Zbilut proposed several parameters to illustrate the coordination structure between both systems [112]. The first measure (%REC) is the percentage of recurrent points on the plot. Ranging from 0% to 100%, this measure indicates the degree to which both systems tend to visit similar states. Diagonal structures represent periods in one time series that show a similar trajectory as those in another time series at a different time. Stochastic behavior tends to produce very short diagonals, whereas deterministic behavior produces longer diagonals. Thus, the rate of recurrence points forming diagonal lines is informative of the determinism of the interaction between the two time series. The average length of the diagonal line represents the time that both systems stay attuned. Finally, by computing a histogram of the length of all diagonals, the authors deduce the entropy of the cross-recurrence plot. Entropy reflects the complexity of the deterministic structure in the system.

5.2.3 Spectral methods

Spectral methods constitute an interesting alternative to temporal methods when dealing with rhythmic tasks. Spectral methods measure the evolution of the relative phase between the two partners as an indication of a stable time-lag between them. For instance, Oullier et al. [81] and Richarson et al. [105], [110] proposed plotting the histogram of the relative phases across the whole interaction. The stability of the interpersonal coordination between the dyadic partners was indicated by the degree of flatness of the phase distribution. Spectral methods also measure the overlap between the movement frequencies of the partners, called cross-spectral coherence [82], [110], [79] or power spectrum overlap [81], [113]. This quality is measured as the area of intersection between each participant's normalized spectral plots and indicates the strength of the frequency entrainment between the two partners.

5.3 Significance test : pseudo-synchrony

A critical question when attempting to detect dependence relationships between features is where the boundary between scores indicating significant and insignificant synchrony should be. Ashenfelter et al. summarize the issue [88]: "...it is reasonable to ask whether these coefficients are due to the coordination between people as they act in a mutual perception-action cycle or if these values might be due to the overall context of the conversation."

TABLE 1

Studies on computational assessment of synchrony - Correlation

Study	Model	Setting	Features	Participants[1]
[75]	Windowed cross-lagged regression	Friends and non-friends dyads playing a computer game in neutral/conflict situations	Global Motion	N=13 (C[2]-C)
[88]	Correlation	Job interview role-play sessions in mixed/sex and same/sex interactions	Head Motion	N=128 (A-A)
[80]	Correlation	Attempt to synchronize repeated productions of a one or two-word sequences	Speaker tongue position	N=1 (A-A)
[89]	Correlation	Free interacting sessions in noisy/quiet environments	Head and hand motion	N=4 (A-A)
[98]	Correlation	Pairs of individuals imitating each other's movements in dance in various leading/following instructions	Velocity	N=6 (A-A)
[91], [92], [93]	Correlation	Free interacting sessions, meetings	Head and body motion	N=1[3](A-A)
[79]	Correlation Dendrogram	Construction task between a demonstrator and a experimenter separated with a folding screen	Motion Energy Image Motion History Image Hands trajectory Pitch Energy Pause Vocalic energy	N=7 (A-A)
[100]	Correlation	Conversation	Global Motion	N=1 (A-A)
[33]	Correlation	Therapy sessions	Global Motion	N=2 (A-A)
[99]	Correlation	Therapy sessions	Global Motion	N=50 (A-A)
[101], [57]	Correlation	Therapy sessions	Global Motion	N=70 (A-A)
[102]	Correlation	Role-playing interview counseling sessions	Global Motion	N=4 (A-A)
[103]	Correlation	Educational counseling sessions (N=2) Psychotherapeutic counseling sessions (N=4)	Global Motion	N=6 (A-A)
[111], [104]	Correlation	Face-to-face discussions and conversations	Global Motion	N=40 (A-A)

[1] Number of dyads

[2] A=Adult,C=Child,M=Mother

[3] Group of 4 persons

Consequently, a baseline is needed to compare the scores and determine the significance of the measure. Bernieri et al. originally proposed a rating method ("the pseudo-synchrony experimental paradigm") to evaluate the interactional synchrony that occurred in a dyadic interaction [21]. The method consists of synthesizing surrogate data (pseudo-interactions): video images of dyadic partners are isolated and re-combined in a random order. Judges then rate the original and pseudo-interactions videos. Pseudo-interaction scores constitute a baseline to judge the scores obtained in the original interaction. The idea of generating surrogate data and comparing the synchrony scores on the genuine and surrogate datasets has been extended to automatic computation of interactional synchrony. First, features are extracted for each dyadic partner. The temporal structure of the first partner's time series is destroyed and re-associated with the second partner's original time series. Synchrony scores are assessed using the original and surrogate datasets. The synchrony scores on the surrogate dataset constitute a baseline for judging for the dyad's coordination [82], [88], [111].

Ramseyer and Tschacher go beyond the comparison with a single surrogate dataset [101]. They build N (N=100) surrogate datasets and estimate the distribution of the surrogate synchrony scores. A statistical test is performed to test the hypothesis that the genuine synchrony score stems from the same distribution as the surrogate synchrony scores. Interactional synchrony scores are considered significant if genuine synchrony scores are above two standard deviations of the pseudo-synchrony scores (one-sided z-test, $p < 0.05$).

Various methods have been used to generate the surrogate datasets (offsetting one time series of a large time-lag [88], time-shuffling n-second-length windows of one time series [101], [79], [82], associating mismatched partners who did not interact with one another [82]). [101] provides for a methodological description of resampling methods for synchrony assessment.

TABLE 2

Studies on computational assessment of synchrony - Recurrence analysis

Study	Model	Setting	Features	Participants[1]
[82]	Recurrence analysis	1) A first set of participants talk about a visual scene they are looking at. Audio recordings of their speech are played to a second set of participants looking at the same display. 2) Same display as (1). During the presentation to the second set of participants, the pictures turns from dimmed to full color in a synchronized/random condition	Eyes movements	1)N=36 (A[2]-A) 2)N=36 (A-A)
[85], [86]	Recurrence analysis	Pair of participants standing and performing a puzzle interpersonal task with several variables manipulated (facing toward or away from each other, conversing with each other or a confederate)	Postural movements	N=13 (A-A)
[87]	Recurrence analysis	Pair of participants standing and producing words in synchrony or in alternation, as the experimenters varied speaking rate (1) and word similarity (2)	Postural movements	1)N=36 (A-A) 2)N=17 (A-A)
[94]	Recurrence analysis	1) Pair of violin players acting four emotions (Joy, Anger, Sadness and pleasure) during their music performance in different conditions (with or without visual contact) 2) Quartet of violin players acting in three different ways (functional, regular, over-expressive)	1) Head trajectory and velocity 2) Head trajectory and velocity, biometric data (heart rate, breath, ocular movements, face muscles), music beat	1)N=2 (A-A) 2)N=1 (A-A)
[95], [96]	Recurrence analysis	On-stage live musical performance for emotional entrainment analysis	Head motion	N=4
[97]	Recurrence analysis	Dataset containing random or periodic motions	Silhouette motion	N/A

[1] Number of dyads

[2] A=Adult,C=Child,M=Mother

These methods are subject to three main criticisms in the context of studying naturalistic interaction data. First, the evaluation and interpretation of these methods are particularly delicate. On the basis of the discriminative or predictive power of the measure, the authors cannot really know if it is real synchrony that is measured or just a co-occurrence of events without meaning, particularly for features as global as Motion Energy. Moreover, the measures provided by these methods are mostly global and do not shed light on what happened locally during the interaction; they do not provide a local model of the communication dynamics. Second, the importance of speech and multimodality is often concealed in these methods. Third, these methods are suitable for analyzing a database but do not provide direct insights on how to equip a machine with such coordination skills.

6 MODELING COMMUNICATION DYNAMICS

Given these criticisms, many in the field adopted the alternative practice of modeling the timing and occurrence of higher-level behavioral events such as smiles, head gestures, gazes and speaker changes. These behavioral events can be either extracted from a human-annotated database or predicted from low-level signals automatically extracted from data. These methods arise from a great interest in identifying the dynamical patterns of interaction. Instead of studying each partner's behavior separately, the central idea is to characterize recurrent interpersonal behaviors. As Fogel explains, "traditional discrete state analytical tools are useful so long as researchers can couple them with other measures of the creative dynamics of social communication process" [116]. Another concern is obtaining knowledge from human-human communication dynamics to improve dialog systems, agent or robot interaction skills.

TABLE 3

Studies on computational assessment of synchrony - Spectral methods

Study	Model	Setting	Features	Participants[1]
[81]	Relative phase and frequency overlap	Pair of participants sitting in front of each other, executing rhythmic finger movements at their own pace	Fingers motion	N=6 (A[2]-A)
[106]	Relative phase	Pair of participants sitting side-by-side, swinging hand-held pendulum with several variables manipulated (frequency competition between oscillators, frequency of oscillation, in-phase/anti-phase coordination)	Motion of the pendulum	N=3 (A-A)
[107]	Relative phase	Pair of participants sitting side-by-side, swinging hand-held pendulum with several variables manipulated (frequency competition between oscillators, frequency of oscillation, in-phase/anti-phase coordination)	Motion of the pendulum	N=5 (A-A)
[108]	Relative phase	Pair of participants sitting side-by-side, swinging hand-held pendulum with several variables manipulated (social competence, frequency competition between oscillators, frequency of oscillation)	Motion of the pendulum	N=9 (A-A)
[109]	Relative phase, Cross-spectral coherence	Pair of participants sitting side-by-side, swinging hand-held pendulum with several variables manipulated (visual/non-visual, frequency competition between oscillators, with/without methodological controls on the respect of the experimental conditions)	Motion of the pendulum	N=10 (A-A)
[84]	Relative phase	Pair of participants coordinating the oscillation of their legs at a different tempos, in in-phase/anti-phase conditions	Leg motion	N=6+1+1 (A-A)
[114]	Relative phase	Pair of participants sitting side-by-side, swinging hand-held pendulum with several variables manipulated (coordinated or uncoupled, frequency competition between oscillators)	Motion of the pendulum	N=3 (A-A)
[83]	Cross-spectral coherence	Pair of participants sitting in front of each other, sorting card decks in suit piles with several variables manipulated (social facilitation, shared piles and control)	Hand motion	N=15 (A-A)
[105]	Cross-spectral coherence, phase distribution	Pair of participants sitting side-by-side, swinging hand-held pendulum and performing a puzzle interpersonal task with several variables manipulated (visual, visual+verbal, verbal)	Motion of the pendulum	N=12+9 (A-A)
[110]	Cross-spectral coherence, Phase mode, relative phase shift	1)Pair of participants sitting side-by-side in rocking chairs asked to rock in inphase/antiphase conditions with focal of peripheral vision of their partner 2)Pair of participants sitting side-by-side in rocking chairs asked to rock at their own preferred frequency with focal of peripheral vision of their partner	Motion of the chair	1)N=12 (A-A) 2)N=8 (A-A)
[115]	Wavelet transform Relative phase	Pair of participants sitting across of each other swinging their forearm with the intention to coordinate or not to coordinate	Forearm motion	N=6 (A-A)

[1] Number of dyads

[2] A=Adult,C=Child,M=Mother

6.1 Modeling social interaction as sequences of behaviors

Machine learning methods offer an interesting framework for the exploration of interactive behaviors. A key challenge is proposing models with the content and temporal structure of dyadic interactions. Various sequential learning models, such as Hidden Markov Models (HMMs) or Conditional Random Fields (CRFs), are usually used to characterize the temporal structure of social interactions. Messinger et al. employ related techniques for the understanding of communicative development, which is characterized by mutual influences during interaction: infants and parents influence and respond to one another during communication [117]. The authors focus on some specific social signals (e.g., smiling) and propose statistical approaches for the characterization of this signal on the response of the partner, who can be the parent or the infant. Probability distributions of transitions between states of behaviors previously annotated are estimated by maximum likelihood approach: $p(i_a, m_a, i_{a-1}, m_{a-1})$). These probabilities are used to characterize the dynamic of the early dyadic interaction using high-level labeled information, such as smiles. The comparison between the interactive situations among in infant development is accomplished using a similarity metric (Battacharyya coefficient).

Modeling temporal structure is not always sufficient. Magnusson underlines the complexity of the detection of such interactive patterns using methods that simply study the sequence of events [118]. Interactive patterns are (1) hierarchical, a pattern is often formed of different levels of sub-patterns, and (2) variable, the number and types of behavior can vary greatly from one instance to another of the same pattern. He proposed an algorithm to characterize the complex hidden and repetitive temporal structure of verbal and non-verbal behavior (T-pattern). A T-pattern is defined as a repeated occurrence of a sequence of events that are separated by a "relatively invariant" time interval, the critical interval. The invariance of the time interval is evaluated by comparing the probability of observing this time interval between two successive events with the probability of observing the same time interval in a randomly distributed sequence of events. The strength of the algorithm is that it allows the insertion or omission of events according to the instance of the pattern. Moreover, the algorithm provides a hierarchical tree diagram to model the existence of patterns and subpatterns. Running the algorithm on randomized time series allows an analysis of whether some of the patterns are expected to occur by chance. Magnusson also suggests the danger of simply focusing on two successive events, which could lead to a misinterpretation of the causality of events, while the n prior behaviors of the T-pattern may also have an influence. T-pattern methodology and the Theme software have been applied to the analysis of social interaction with autistic children [119], movement synchrony in interpersonal attraction [120], [121] and symmetric behaviors in social interaction [122].

Dynamic verbal and non-verbal communications usually involve multimodal signals more or less simultaneously produced. Investigations of communication dynamics should address this phenomenon; integrated models offer relevant solutions. In [123], [124], an integrative approach is proposed to explicitly consider the interaction synchrony of behaviors. The model is applied to the characterization of parent-infant interactions for differential diagnosis: autism (AD), intellectual disability (ID) and typical development (TD). As in [117], the authors estimate transitions between behaviors of the infant and the parent by analyzing behaviors co-occurring in a 3s window. Assuming a Markovian process, they used a maximum likelihood estimation to estimate the probability of each interactive pattern, resulting in bi-gram models characterizing the temporal structure. To study these interactive patterns from an integrative perspective, the authors proposed employing a more global model using non-negative matrix factorization (NMF) [125], coupled with statistical representation, namely tf-idf (term frequency-inverse document frequency), to transform the scene annotations (bi-gram) into a representation suitable for the learning algorithm and the clustering task (NMF). A global non-negative matrix grouping all interactive behaviors (bi-gram) is constructed and decomposed into a few interactive behaviors groups. NMF is an unsupervised feature extraction method involving the decomposition of the non-negative matrix into two non-negative matrices. The non-negativity constraints are relevant for the analysis of human behaviors as they allow only additive, not subtractive, combinations. Because of the mathematical properties of NMF, the analysis accommodates an integrative perspective by providing clusters of interactive behaviors. In addition, to understand the development similarity of TD infants compared with AD and ID infants, the authors used Normalized Mutual Information (NMI), as proposed by [126]; the NMI between two different clustering solutions measures their agreement. Interestingly, the NMI profiles fitted the clinical hypothesis closely, showing a pervasive development in AD and a delayed development in ID [123].

6.2 Prediction of communication dynamics: turn-taking and backchannels

As discussed in section 2, synchrony is related to the continuous adaptation of behaviors between interaction partners. Several teams are interested "in developing predictive models of communication dynamics that integrate previous and current actions from all interlocutors to anticipate the most likely next actions of one or all interlocutors" [127].

The prediction of turn-taking has been largely studied in the perspective of building fluent dialog systems. The purpose of the prediction is to accurately predict the timing between speaker transitions and the upcoming type of utterance (speaker holding the floor, speaker changes) as it occurs in human-human interactions. For instance, [128] proposed predicting whether a speaker change will occur

or not at the end of an utterance given prosodic, spectral features and the duration of the previous talkspurt. [129] proposed combining features from several modalities to predict the end of a turn. Another option to improve the performance of turn-taking predictors is to look at the behavior of both partners, instead of focusing only on the speaker. For instance, [130] proposed predicting interruptions in dialog. They used Hidden Conditional Random Fields and compared three sets of features: (1) interrupter gestural features (mouth, eyebrow and head), (2) interruptee prosodic features, and (3) an optimized combination of both first sets. They showed that the set combining features from both partners outperformed the performance of individual sets. [131] proposed several guidelines for predicting turn-taking in a dialog system. They propose that (1) predictions should be made constantly rather than at certain time points, (2) predictions should be made for several points in the future and not only for the next instant, (3) predictions should be made for the user and the system, instead of predicting the user's behavior and then determining the behavior of the system with some additional reasoning. This framework seems promising in the sense that it could handle any turn-taking pattern. Moreover, the authors propose going beyond turn-taking and predicting "turn-shaping", the prosodic parameters of the next turn (pitch or speaking rate).

Back-channel behavior is intrinsically linked to turn-taking and includes continuers ("hum", "aha") and regulatory gestures (head nods, shakes, laughter). Back-channel behavior assures the speaker that the listener is paying attention and is in the same state in the conversation [132]. Several teams have investigated how the speaker behavior triggered listeners' back-channels. For instance, Cathcart et al. proposed predicting back-channels continuers using part-of-speech tags and pause duration [133]. Gravano et al. studied how intonation, intensity level, pitch level, voice quality and inter-pausal unit duration yielded to back-channel continuers [134]. Morency et al. proposed studying which speaker feature (prosody, pause, spoken words, eye gaze) is important to predict the occurrence and timing of listener's head nods [135]. They used sequential probabilistic modeling (Hidden Markov Model and Conditional Random Field) to learn the dynamics from a human-human interaction database. An important aspect of their model is the ability to consider the joint influence of several features to trigger a back-channel.

7 OPEN QUESTIONS AND PROSPECTS

7.1 Open questions

Several questions regarding the dimension and perception of synchrony remain to be explored. These questions are fundamental to the development of an automatic model to assess synchrony. A first question relates to the timescale of synchrony: second, minute, all interaction. Is it appropriate to break behavior into small units? Is it possible to operationalize synchrony and/or measure occurrences of synchrony? A second question concerns the dimension of synchrony: is synchrony an all-or-none condition (synchronous vs. non-synchronous)? In other words, can dyadic interaction can approach or move away from synchrony [32]? Is synchrony a continuous or a discrete notion? Six-month-olds can detect when a modality is synchronous and another modality is not [136]. Thus, with regard to the unimodal versus inter-modal question, synchrony appears to be experienced as an all-or-none condition. Out of the question of modality and when addressing naturalistic interaction, much of the current information suggests that synchrony is a continuous notion. Various sources indicate that synchrony varies over the course of interaction, being stronger at the beginning and the ending of an exchange [36] or at moments of particular engagement [93]. Feldman operationalizes synchrony as the degree to which the partners change their affective behavior in reference to one another and obtains a number ranging between 0 and 1 [34]. When addressing the matter of movement synchrony and its relation to perceived entitativity, Lakens observed that objective differences in movement rhythms were linearly related to ratings of perceived entitativity [53]. Finally, a recent study showed that the perception of coordination was more unanimous when coordination was very high or very low. However, judges were not reliable when judging dyads with "medium" coordination [137].

The question of the corpus is also crucial. Until Sun et al.'s [138] recent contribution of their mimicry database, no publicly available annotated corpus were dedicated to the detection of synchrony. We can hope that this effort will benefit the field, aiding engineers in their work to develop new algorithms, skipping the data collection and annotation phases.

7.2 Prospects in developmental robotics

In the last decade, researchers in the field of robotics, signal processing and artificial intelligence have taken a growing interest in developmental phenomena, such as parent-infant synchrony, language acquisition and joint attention [139]. For instance, the goal of developmental robotics is to enable robots and other artificial systems to autonomously develop skills for any particular environment rather than programming them to solve particular goals for a specific environment. This approach was inspired by human-infant interaction to design robots.

Cangelosi et al. [140] identified a sequence of milestones for future research in the field. In the social learning section, the first target involves studying and implementing non-verbal social cues for language and skill learning. By decoding the appropriate non-verbal signals, robots can achieve joint attention with the partner, orient their gaze toward the partner's focus of attention, and mirror the partner, among other things. Through these synchronized exchanges, the robot can acquire language by associating its focus of attention with information extracted from the partner's speech.

Prepin and Gaussier [141] proposed a robotic architecture (ADRIANA) able to measure the degree of synchrony with a human and adapt its behavior accordingly. The robot is equipped with two arms and two possible positions (lowered or raised). Participants were asked to "make the robot learn to move the arms, which is on the same side as the one they move". At the beginning of the experiment, the robot raises or lowers one of its arms, randomly left or right, when the human does. At each trial, the time-delay between the robot and the human is predicted according to the past observations. Synchrony is assessed as the error between the predicted delay and the real delay. Thus, synchrony is used as a reinforcement signal to learn right-left associations. They found that the learning process converged for all participants.

7.3 Prospects in social robots and embodied conversational agents

Gratch et al. evaluated the importance of contingency, a prerequisite of synchrony, on various settings involving human-human and human-virtual agent interactions [142]. They compared the participants' feeling of rapport when facing a responsive or non-contingent virtual agent listener. The responsive behavior corresponded to non-verbal productions (nodding, shaking, mirroring) when specific vocal and motion behaviors from the human were recognized. In the non-contingent behavior, the human was presented with a virtual agent with a pre-recorded behavior sequence. The researchers found evidence that the contingency of agent feedbacks influenced the behavior of the human participant and was involved in the creation of virtual rapport.

Similarly, in the field of social robotics, equipping a machine with social abilities, such as synchrony and turn-taking, represents a great challenge in the design of socially accepted interfaces [2]. Michalowski et al. designed the robot, Keepon, to engage in synchronous interactions with children [143]. Keepon is programmed to dance by performing periodic movements that smoothly and dynamically change tempo according to perceived rhythms. The rhythms can be extracted from various sensors according to the conditions being tested (vision, audio, pressure sensors, accelerometers, etc.). They studied the effect of synchronized movements on engagement under several conditions with the robot following the rhythm of a song or the movement of the child.

Kozima et al. also performed a longitudinal observation of autistic children interacting with Keepon [144]. Keepon was introduced as a toy in a daycare center for children with pervasive developmental disorders over a three-year period. They observed the emergence of interactions between the children and the robot. Some children engaged in imitation play with the robot. For other children, the robot acted as a pivot for "sharing and exchanging pleasure and surprise with the caregivers". The authors claimed that the predictable and simple behavior of the robot facilitated the emergence of social behaviors.

7.4 Prospects in developmental and clinical studies

In the field of child psychiatry, many potential advantages to using interactive robots in clinical settings with individuals with ASD have been proposed. These advantages include the intrinsic appeal of technology to individuals on the spectrum, the ability of robots to produce simple and isolated social behaviors repetitively, and the fact that they can adapt to provide individualized treatment [145]. However, despite media interest, research in this area has been only exploratory, aiming at evaluating preference for machine-like characteristics or using a robot to elicit behaviors, practice a skill and provide feedback (for a review see [145]).

Tartaro et al. proposed to design virtual peers to help children acquire communicative skills [146]. They studied the production of contingent discourse of children with ASD in a collaborative task with a virtual peer. The virtual peer was controlled with the Wizard Of Oz methodology and incorporated "facilitating features" such as yes/no questions or conceptually-simple questions, to elicit responses from the child. They observed that compared to an interaction with a human peer, ASD children produced more contingent responses with the virtual peer. More, over the course of interaction the production of contingent responses increased.

However, studies on ASD have not yet considered the key role of social-communicative interaction in mediating interest and infants' gaze following of a robot. In a very elegant study including several conditions, Meltzoff et al. [5] reported that 18-month-olds' observation of a social robot interacting with an adult and imitation at a distance changed the infants' interpretations of what the robot was. He was no longer seen as toy with random physical movement but as a psychological agent that could see.

At last, communicative development is characterized by mutual influences and co-regulations of social signals and affective states between parent and infant during interaction. Thus, probabilistic models and machine learning techniques offer an interesting framework to model such interactive behaviors [147], [45], [148], [149].

8 CONCLUSION

The current essay attempted to show that the assessment of interactional synchrony bears challenging questions at the crossover of several research domains. Psychologists' coding methods and non-computational evaluation tools are essential for engineers to identify the relevant signals, validate machine-learning techniques to automatically detect occurrence of synchrony and model interactive patterns. New socially adapted interfaces could emerge from a better analysis of these social mechanisms. In return, psychologists could benefit from computational methods developed to study synchrony. Such methods could provide automatic and objective tools to study interactive abilities in several psychiatric conditions, such as depression and autism. Although few studies are currently available in this specific field, they appear to

be very promising (couple therapy [150], success in psychotherapy [57], mother-infant interaction [151]). Another great potential lies in the opportunity to build robots or virtual agents with interactive abilities. Indeed, such a setting allows control of the variables and testing of different settings and behaviors. Such manipulations can benefit both engineers and clinicians and can lead to a better understanding of the underlying mechanisms of social interactions.

ACKNOWLEDGMENTS

REFERENCES

[1] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," Image and Vision Computing, vol. 27, no. 12, pp. 1743–1759, November 2009.

[2] K. Prepin and C. Pelachaud, "Shared understanding and synchrony emergence: Synchrony as an indice of the exchange of meaning between dialog partners." in ICAART2011 International Conference on Agent and Artificial Intelligence, vol. 2, Jan 2011, pp. 25–30.

[3] K. Prepin and A. Revel, "Human-machine interaction as a model of machine-machine interaction : how to make machines interact as humans do," Advanced Robotics, vol. 21, no. 15, December 2007.

[4] J. Cassell, T. Bickmore, L. Campbell, H. Vilhj‡lmsson, and H. Yan, "Conversation as a system framework: Designing embodied conversational agents," in Embodied Conversational Agents. MIT Press, 2000, pp. 29–63.

[5] A. N. Meltzoff, R. Brooks, A. P. Shon, and R. P. N. Rao, ""Social" robots are psychological agents for infants: A test of gaze following," Neural Networks, vol. 23, no. 8-9, pp. 966–972, 2010.

[6] J. Nadel, I. Carchon, C. Kervella, D. Marcelli, and D. Réserbat-Plantey, "Expectancies for social contingency in 2-month-olds," Developmental Science, vol. 2, no. 2, pp. 164–173, 1999.

[7] H. H. Clark, Using Language, B. G. Blount, Ed. Cambridge University Press, 1996, vol. 23, no. 4.

[8] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," Behavioral and Brain Sciences, vol. 27, no. 02, p. 169–190, 2004.

[9] S. Garrod and M. J. Pickering, "Why is conversation so easy?" Trends in Cognitive Sciences, vol. 8, no. 1, pp. 8–11, 2004.

[10] S. Garrod and A. Anderson, "Saying what you mean in dialogue: a study in conceptual and semantic co-ordination." Cognition, vol. 27, no. 2, pp. 181–218, 1987.

[11] S. Garrod and A. Clark, "The development of dialogue coordination skills in schoolchildren," Language & Cognitive Processes, vol. 8, no. 1, pp. 101–126, 1993.

[12] S. E. Brennan and H. H. Clark, "Conceptual pacts and lexical choice in conversation," Journal of Experimental Psychology Learning Memory and Cognition, vol. 22, no. 6, pp. 1482–1493, 1996.

[13] H. P. Branigan, M. J. Pickering, and A. A. Cleland, "Syntactic co-ordination in dialogue," Cognition, vol. 75, no. 2, pp. B13–B25, May 2000.

[14] H. Giles, J. Coupland, and N. Coupland, Accommodation theory: Communication, context, and consequence. Cambridge University Press, 1991, no. 1984, pp. 1–68.

[15] H. Giles, "Accent mobility: A model and some data," Anthropological Linguistics, vol. 15, no. 2, pp. 87–105, 1973.

[16] R. Y. Bourhis and H. Giles, The language of intergroup distinctiveness. Academic Press, 1977, pp. 119–134.

[17] R. L. Street, "Speech convergence and speech evaluation in fact-finding interviews," Human Communication Research, vol. 11, no. 2, pp. 139–169, 1984.

[18] F. R. Bilous and R. M. Krauss, "Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads," Language and Communication, vol. 8, no. 3/4, pp. 183–194, 1988.

[19] J. N. Cappella and S. Planalp, "Talk and silence sequences in informal conversations iii: Interspeaker influence," Human Communication Research, vol. 7, no. 2, pp. 117–132, 1981.

[20] W. Condon and W. Ogston, "A segmentation of behavior," Journal of Psychiatric Research, vol. 5, pp. 221–235, 1967.

[21] F. Bernieri and R. Rosenthal, Interpersonal coordination: Behavior matching and interactional synchrony. Fundamentals of nonverbal behavior. Cambridge University Press, 1991.

[22] J. L. Lakin and T. L. Chartrand, "Using nonconscious behavioral mimicry to create affiliation and rapport," Psychological Science, vol. 14, no. 4, pp. 334–339, 2003.

[23] T. L. Chartrand and A. N. Dalton, "Mimicry: Its ubiquity, importance and functionality," Oxford handbook of human action, vol. 2, pp. 458–483, 2009.

[24] X. Sun and A. Nijholt, "Multimodal embodied mimicry in interaction," in Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues, ser. Lecture Notes in Computer Science, A. Esposito, A. Vinciarelli, K. Vicsi, C. Pelachaud, and A. Nijholt, Eds. Springer Berlin Heidelberg, 2011, vol. 6800, pp. 147–153.

[25] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception-behavior link and social interaction," Journal of personality and social psychology, vol. 76, no. 6, pp. 893–910, 06 1999.

[26] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, "Premotor cortex and the recognition of motor actions." Brain Research, vol. 3, no. 2, pp. 131–141, 1996.

[27] T. Chaminade and J. Decety, "A common framework for perception and action: Neuroimaging evidence," Behavioral and Brain Sciences, vol. 24, no. 5, pp. 879–, 2001.

[28] L. O. Lundqvist, "Facial emg reactions to facial expressions: a

case of facial emotional contagion?" Scandinavian Journal of Psychology, vol. 36, no. 2, pp. 130–141, 1995.

[29] C. J., Coding Mutual Adaptation in Dyadic Nonverbal Interaction. Lawrence Erlbaum, 2005, pp. 383–392.

[30] B. Newman and P. Newman, Development through life: A psychosocial approach. Cengage/Wadsworth, 2009, pp. 171–175.

[31] W. Condon and L. Sander, "Neonate movement is synchronized with adult speech: interactional participation and language acquisition," Science, vol. 183, pp. 99–101, 1974.

[32] A. Harrist and R. Waugh, "Dyadic synchrony: Its structure and function in children's development," Developmental Review, vol. 22, no. 4, pp. 555–592, 2002.

[33] F. Ramseyer and W. Tschacher, "Synchrony: A core concept for a constructivist approach to psychotherapy," Constructivism. The Human Sciences., vol. 11, pp. 150–171, 2006.

[34] R. Feldman, "Infant-mother and infant-father synchrony : the coregulation of positive arousal," Infant Mental Health Journal, vol. 24, no. 1, pp. 1–23, 2003.

[35] D. Stern, A. Lazartigues, and D. Pérard, Le monde interpersonnel du nourrisson: une perspective psychanalytique et développementale, ser. Le Fil rouge. Section 2, Psychanalyse et psychiatrie de l'enfant. Presses universitaires de France, 2003.

[36] A. Kendon, "Movement coordination in social interaction: some examples described," Acta Psychologica, vol. 32, pp. 100–125, 1970.

[37] C. Trevarthen and K. J. Aitken, "Infant intersubjectivity: research, theory, and clinical applications." The Journal of Child Psychology and Psychiatry and Allied Disciplines, vol. 42, no. 1, pp. 3–48, 2001.

[38] C. Trevarthen, Communication and cooperation in early infancy: A description of primary intersubjectivity. Cambridge University Press, 1979, pp. 321–347.

[39] D. S. Schechter and E. Willheim, "Disturbances of attachment and parental psychopathology in early childhood," Child and Adolescent Psychiatric Clinics of North America, vol. 18, no. 3, pp. 665 – 686, 2009.

[40] J. Kim-Cohen, G. A. Light, D. L. Braff, C. L. M. Caton, R. E. Drake, D. S. Hasin, P. E. Shrout, S. Samet, W. B. Schanzer, T. E. Moffitt, A. Taylor, S. J. Pawlby, A. Caspi, J. M. Hettema, C. A. Prescott, J. M. Myers, M. C. Neale, K. S. Kendler, M. R. Liebowitz, A. J. Gelenberg, and D. Munjack, "Maternal depression and children's antisocial behavior: Nature and nurture effects," Archives of General Psychiatry, vol. 62, pp. 173–181, 2005.

[41] M. Rutter and T. G. O'Connor, "Are there biological programming effects for psychological development? Findings from a study of Romanian adoptees." Developmental Psychology, vol. 40, no. 1, pp. 81–94, 2004.

[42] P. Perruchet and S. Pacton, "Implicit learning and statistical learning: one phenomenon, two approaches," Trends in Cognitive Sciences, vol. 10, no. 5, pp. 233 – 238, 2006.

[43] J. R. Saffran, R. N. Aslin, and E. L. Newport, "Statistical learning by 8-month-old infants," Science, vol. 274, pp. 1927–1928, December 1996.

[44] P. K. Kuhl, "Is speech learning "gated" by the social brain?" Developmental Science, vol. 10, no. 1, pp. 110–120, 2007.

[45] ——, "Early language acquisition: cracking the speech code," Nat Rev Neurosci, vol. 5, no. 11, pp. 831–843, Nov. 2004.

[46] P. K. Kuhl, J. E. Andruski, I. A. Chistovich, L. A. Chistovich, E. V. Kozhevnikova, V. L. Ryskina, E. I. Stolyarova, U. Sund-berg, and F. Lacerda, "Cross-language analysis of phonetic units in language addressed to infants," Science, vol. 277, pp. 684–686, 1997.

[47] P. K. Kuhl, "A new view of language acquisition," Proceedings of The National Academy of Sciences, vol. 97, pp. 11 850–11 857, 2000.

[48] M. H. Goldstein, A. P. King, and M. J. West, "Social interaction shapes babbling: Testing parallels between birdsong and speech," Proceedings of the National Academy of Sciences of the United States of America, vol. 100, no. 13, pp. 8030–8035, 2003.

[49] J. Piaget, La formation du symbole chez l'enfant: imitation, jeu et rêve, image et représentation, ser. Actualités pédagogiques et psychologiques. Delachaux et Niestlé, 1976.

[50] A. N. Meltzoff and M. K. Moore, "Imitation et développement humain : les premiers temps de la vie," Terrain, vol. 44, pp. 71–90, 2005.

[51] M. Lafrance, Posture Mirroring and Rapport. Human Sciences Press, New York, NY, 1982, pp. 279–298.

[52] L. K. Miles, L. K. Nind, and C. N. Macrae, "The rhythm of rapport: Interpersonal synchrony and social perception," Journal of Experimental Social Psychology, vol. 45, no. 3, pp. 585 – 589, 2009.

[53] D. Lakens, "Movement synchrony and perceived entitativity," Journal of Experimental Social Psychology, vol. 46, no. 5, pp. 701 – 708, 2010.

[54] S. S. Wiltermuth and C. Heath, "Synchrony and cooperation," Psychological Science, vol. 20, no. 1, pp. 1 – 5, 2009.

[55] C. N. Macrae, O. K. Duffy, L. K. Miles, and J. Lawrence, "A case of hand waving: Action synchrony and person perception," Cognition, vol. 109, no. 1, pp. 152 – 156, 2008.

[56] L. K. Miles, L. K. Nind, Z. Henderson, and C. N. Macrae, "Moving memories: Behavioral synchrony and memory for self and others," Journal of Experimental Social Psychology, vol. 46, no. 2, pp. 457 – 460, 2010.

[57] F. Ramseyer and W. Tschacher, "Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome," Journal of Consulting and Clinical Psychology, vol. 79, no. 3, pp. 284 – 295, 2011.

[58] A. L. Bouhuys and M. M. Sam, "Lack of coordination of nonverbal behaviour between patients and interviewers as a potential risk factor to depression recurrence: vulnerability accumulation in depression," Journal of Affective Disorders, vol. 57, no. 1-3, pp. 189 – 200, 2000.

[59] G. Bird, J. Leighton, C. Press, and C. Heyes, "Intact automatic imitation of human and robot actions in autism spectrum

disorders," Proceedings of the Royal Society B Biological Sciences, vol. 274, no. 1628, pp. 3027–3031, 2007.

[60] M. Kipp, "Spatiotemporal coding in anvil," in LREC, 2008.

[61] J. Cappella, "Behavioral and judged coordination in adult informal social interactions: vocal and kinesic indicators," Pers. Soc. Psychol., vol. 72, pp. 119–131, 1997.

[62] C. Trevarthen and S. Daniel, "Disorganized rhythm and synchrony: Early signs of autism and Rett syndrome," Brain and Development, vol. 27, no. 1, pp. S25–S34, 2005.

[63] C. J., A Procedure to Measure Interactional Synchrony in the Context of Satisfied and Dissatisfied Couples' Communication. Lawrence Erlbaum, 2005, pp. 199–208.

[64] C. Saint-Georges, R. Cassel, D. Cohen, M. Chetouani, M.-C. Laznik, S. Maestro, and F. Muratori, "What studies of family home movies can teach us about autistic infants: A literature review," Research in Autism Spectrum Disorders, vol. 4, no. 3, 2010.

[65] F. Bernieri, J. Reznick, and R. Rosenthal, "Synchrony, pseudo synchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions," Journal of Personality and Social Psychology, vol. 54, no. 2, pp. 243–253, 1988.

[66] R. Feldman, "Coding interactive behavior manual (CIB)," Unpublished manuscript Ramat-Gan, Israel: Bar Ilan University, 1998.

[67] ——, "Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions." The Journal of Child Psychology and Psychiatry and Allied Disciplines, vol. 48, no. 3-4, pp. 329–354, 2007.

[68] R. Feldman, M. Keren, O. Gross-Rozval, and S. Tyano, "Mother-child touch patterns in infant feeding disorders: Relation to maternal, child, and environmental factors," Journal of the American Academy of Child & Adolescent Psychiatry, vol. 43, no. 9, pp. 1089 – 1097, 2004.

[69] D. Dollberg, R. Feldman, and M. Keren, "Maternal representations, infant psychiatric status, and mother–child relationship in clinic-referred and non-referred infants," European Child & Adolescent Psychiatry, vol. 19, pp. 25–36, 2010, 10.1007/s00787-009-0036-5.

[70] D. Silberstein, R. Feldman, J. M. Gardner, B. Z. Karmel, J. Kuint, and R. Geva, "The mother–infant feeding relationship across the first year and the development of feeding difficulties in low-risk premature infants," Infancy, vol. 14, no. 5, pp. 501–525, 2009.

[71] R. Feldman and A. I. Eidelman, "Direct and indirect effects of breast milk on the neurobehavioral and cognitive development of premature infants," Developmental Psychobiology, vol. 43, no. 2, pp. 109–119, 2003.

[72] R. Feldman, A. Weller, L. Sirota, and A. I. Eidelman, "Skin-to-skin contact (kangaroo care) promotes self-regulation in premature infants: sleep-wake cyclicity, arousal modulation, and sustained exploration." Developmental Psychology, vol. 38, no. 2, pp. 194–207, 2002.

[73] R. Feldman, A. I. Eidelman, and N. Rotenberg, "Parenting stress, infant emotion regulation, maternal sensitivity, and the cognitive development of triplets: A model for parent and child influences in a unique ecology." Child Development, vol. 75, no. 6, pp. 1774–91, 2004.

[74] G. Chittaranjan, O. Aran, and D. Gatica-Perez, "Inferring truth from multiple annotators for social interaction analysis," in Neural Information Processing Systems (NIPS) Workshop on Modeling Human Communication Dynamics (HCD), 2011, p. 4.

[75] U. Altmann, "Studying movement synchrony using time series and regression models," I. A. Esposito, R. Hoffmann, S. Hübler, and B. W. (Eds.), Eds., 2011, p. 23.

[76] R. Levitan, A. Gravano, and J. Hirschberg, "Entrainment in speech preceding backchannels," in Annual Meeting of the Association for Computational Linguistics (ACL/HLT), 2011, pp. 113–117.

[77] Štefan Beňuš, A. Gravano, and J. Hirschberg, "Pragmatic aspects of temporal accommodation in turn-taking," Journal of Pragmatics, vol. 43, no. 12, pp. 3001 – 3027, 2011.

[78] J. C. Acosta and N. G. Ward, "Achieving rapport with turn-by-turn, user-responsive emotional coloring," Speech Commun., vol. 53, pp. 1137–1148, November 2011.

[79] E. Delaherche and M. Chetouani, "Multimodal coordination: exploring relevant features and measures," in Second International Workshop on Social Signal Processing, ACM Multimedia 2010, 2010.

[80] A. V. Barbosa, E. Vatikiotis-Bateson, M. Oberg, and R.-M. Déchaine, "An instantaneous correlation algorithm for assessing intra and inter subject coordination during communicative behavior," in Modeling Human Communication Dynamics,NIPS Workshop 2010, 2010.

[81] O. Oullier, G. C. de Guzman, K. J. Jantzen, J. A. Scott Kelso, and J. Lagarde, "Social coordination dynamics: Measuring human bonding," Social Neuroscience, vol. 3, no. 2, pp. 178–192, 2008.

[82] D. C. Richardson and R. Dale, "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension," Cognitive Science, vol. 29, no. 6, pp. 1045–1060, 2005.

[83] R. C. Schmidt, B. O'Brien, and R. Sysko, "Self-organization of between-persons cooperative tasks and possible applications to sport," Int J Sport Psychol., vol. 30, p. 558–579, 1999.

[84] R. Schmidt, C. Carello, and M. Turvey, "Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people," Journal of Experimental Psychology: Human Perception and Performance, vol. 16, no. 2, pp. 227 – 247, 1990.

[85] K. Shockley, M.-V. Santana, and C. A. Fowler, "Mutual interpersonal postural constraints are involved in cooperative conversation," Journal of Experimental Psychology: Human Perception and Performance, no. 29, 2003.

[86] D. Richardson, R. Dale, and K. Shockley, Synchrony and swing in conversation: Coordination, temporal dynamics, and communication. Oxford University Press, 2008.

[87] K. D. Shockley, A. A. Baker, M. J. Richardson, and C. A. Fowler,

"Articulatory constraints on interpersonal postural coordination," Journal of Experimental Psychology: Human Perception and Performance, no. 33, pp. 201–208, 2007.

[88] K. T. Ashenfelter, S. M. Boker, J. R. Waddell, and N. Vitanov, "Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation," J Exp Psychol Hum Percept Perform, vol. 35, no. 4, pp. 1072–91, 2009.

[89] S. M. Boker, M. Xu, J. L. Rotondo, and K. King, "Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series,," Psychological Methods, vol. 7, no. 3, pp. 338 – 355, 2002.

[90] R. Rienks, R. Poppe, and D. Heylen, "Differences in head orientation behavior for speakers and listeners: An experiment in a virtual environment," ACM Trans. Appl. Percept., vol. 7, pp. 2:1–2:13, January 2010.

[91] N. Campbell, "Multimodal processing of discourse information; the effect of synchrony," 2008 Second International Symposium on Universal Communication, pp. 12–15, 2008.

[92] ——, "Automatic detection of participant status and topic changes in natural spoken dialogues," 2008.

[93] ——, "An audio-visual approach to measuring discourse synchrony in multimodal conversation data," in Interspeech, 2009, pp. 2159–2162.

[94] G. Varni, G. Volpe, and A. Camurri, "A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media," Multimedia, IEEE Transactions on, vol. 12, no. 6, pp. 576 –590, oct. 2010.

[95] G. Varni, A. Camurri, P. Coletta, and G. Volpe, "Emotional entrainment in music performance," in Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on, sept. 2008, pp. 1 –5.

[96] A. Camurri, G. Varni, and G. Volpe, "Measuring entrainment in small groups of musicians," in Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on, sept. 2009, pp. 1 –4.

[97] G. Varni, A. Camurri, P. Coletta, and G. Volpe, "Toward a real-time automated measure of empathy and dominance," in CSE (4), 2009, pp. 843–848.

[98] S. M. Boker and J. L. Rotondo, Symmetry building and symmetry breaking in synchronized movement. Amsterdam, Netherlands: John Benjamins Publishing Company, 2002, vol. 42, pp. 163–171.

[99] E. Keller and W. Tschacher, "Prosodic and gestural expression of interactional agreement," COST 2102 Workshop (Vietri), pp. 85–98, 2007.

[100] D. Reidsma, A. Nijholt, W. Tschacher, and F. Ramseyer, "Measuring multimodal synchrony for human-computer interaction," in Proceedings of the International Conference on CYBERWORLDS 2010, A. Sourin, Ed. Los Alamitos: IEEE Computer Society Press, October 2010, pp. 67–71, synchrony, nonverbal communication, measurement, virtual humans, HCI.

[101] F. Ramseyer and W. Tschacher, "Nonverbal synchrony or random coincidence? How to tell the difference," in Development of Multimodal Interfaces: Active Listening and Synchrony, A. Esposito et al., Eds. Springer Berlin / Heidelberg, 2010, vol. 5967, pp. 182–196.

[102] C. Nagaoka and M. Komori, "Body movement synchrony in psychotherapeutic counseling: A study using the video-based quantification method," Ieice Transactions On Information And Systems, vol. E91-D, no. 6, pp. 1634–1640, 2008.

[103] M. Komori, K. Maeda, and C. Nagaoka, "A video-based quantification method of body movement synchrony: An application for dialogue in counseling," Japanese Journal of Interpersonal and Social Psychology, no. 7, pp. 182–196, 2007.

[104] X. Sun, K. P. Truong, M. Pantic, and A. Nijholt, "Towards visual and vocal mimicry recognition in human-human interactions," in IEEE International Conference on Systems, Man, and Cybernetics, SMC 2011: Special Session on Social Signal Processing, E. Tunstel, S. Nahavandi, and A. Stoica, Eds. USA: IEEE Computer Society, November 2011, pp. 367–373.

[105] M. J. Richardson, K. L. Marsh, and R. Schmidt, "Effects of visual and verbal interaction on unintentional interpersonal coordination," Journal of Experimental Psychology: Human Perception and Performance, vol. 31, no. 1, pp. 62 – 79, 2005.

[106] P. Amazeen, R. C. Schmidt, and M. Turvey, "Frequency detuning of the phase entrainment dynamics of visually coupled rhythmic movements," Biological Cybernetics, vol. 72, pp. 511–518, 1995, 10.1007/BF00199893.

[107] R. C. Schmidt, M. Bienvenu, P. A. Fitzpatrick, and P. G. Amazeen, "A comparison of intra- and interpersonal interlimb coordination: Coordination breakdowns and coupling strength," Journal of Experimental Psychology: Human Perception and Performance, vol. 24, no. 3, pp. 884 – 900, 1998.

[108] R. C. Schmidt, N. Christianson, C. Carello, and R. Baron, "Effects of social and physical variables on between-person visual coordination," Ecological Psychology, no. 6, pp. 159–183, 1994.

[109] R. C. Schmidt and B. O'Brien, "Evaluating the Dynamics of Unintended Interpersonal Coordination," Ecological Psychology, vol. 9, no. 3, pp. 189–206, 1997.

[110] M. J. Richardson, K. L. Marsh, R. W. Isenhower, J. R. Goodman, and R. Schmidt, "Rocking together: Dynamics of intentional and unintentional interpersonal coordination," Human Movement Science, vol. 26, no. 6, pp. 867 – 891, 2007.

[111] X. Sun, K. Truong, A. Nijholt, and M. Pantic, "Automatic visual mimicry expression analysis in interpersonal interaction," in Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR-W'11), Workshop on CVPR for Human Behaviour Analysis, Colorado Springs, USA, June 2011, pp. 40–46.

[112] J. Webber and C. L. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," Journal of Applied Physiology, no. 76, p. 965973, 1994.

[113] O. Oullier and J. A. Kelso, Social coordination from the perspective of coordination dynamics. Springer-Verlag, 2009.

[114] R. C. Schmidt and M. Turvey, "Phase-entrainment dynamics of visually coupled rhythmic movements," Biological Cybernetics,

vol. 70, pp. 369–376, 1994, 10.1007/BF00200334.

[115] J. Issartel, L. Marin, and M. Cadopi, "Unintended interpersonal co-ordination: can we march to the beat of our own drum?" Neuroscience Letters, vol. 411, no. 3, pp. 174 – 179, 2007.

[116] A. Fogel, Two principles of communication : co-regulation and framing, ser. International library of psychology. Routledge, 1993.

[117] D. M. Messinger, P. Ruvolo, N. V. Ekas, and A. Fogel, "Applying machine learning to infant interaction: The development is in the details," Neural Networks, vol. 23, no. 8-9, pp. 1004 – 1016, 2010, social Cognition: From Babies to Robots.

[118] M. S. Magnusson, "Discovering hidden time patterns in behavior: T-patterns and their detection." Behavior Research Methods, Instruments, & Computers, vol. 32, no. 1, pp. 93–110, Feb. 2000.

[119] M. Plumet and C. Tardif, "Understanding the functioning of social interaction with autistic children," 2005.

[120] K. Sakaguchi and G. K. Jonsson, "Initial interpersonal attraction between mixed-sex dyad and movement synchrony," The Hidden Structure of Interaction From Neurons to Culture Patterns, pp. 108–120, 2005.

[121] K. Grammer, K. B. Kruck, and M. S. Magnusson, "The courtship dance: Patterns of nonverbal synchronization in opposite-sex encounters," Journal of Nonverbal Behavior, vol. 22, no. 1, pp. 3–29, 1998.

[122] M. T. Anguera, "Analysis of symmetry / asymmetry in social interaction," The Hidden Structure of Interaction From Neurons to Culture Patterns, pp. 52–70, 2005.

[123] C. Saint-Georges, A. Mahdhaoui, M. Chetouani, R. S. Cassel, M.-C. Laznik, F. Apicella, P. Muratori, S. Maestro, F. Muratori, and D. Cohen, "Do parents recognize autistic deviant behavior long before diagnosis? Taking into account interaction using computational methods," PLoS ONE, vol. 6, no. 7, p. e22393, 07 2011.

[124] A. Mahdhaoui and M. Chetouani, "Understanding parent-infant behaviors using non-negative matrix factorization," in Proceedings of the Third COST 2102 international training school conference on Toward autonomous, adaptive, and context-aware multimodal interfaces: theoretical and practical issues. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 436–447.

[125] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," Environmetrics, vol. 5, no. 2, pp. 111–126, 1994.

[126] A. Strehl and J. Ghosh, "Cluster ensembles a knowledge reuse framework for combining multiple partitions," in Journal on Machine Learning Research, 2002, pp. 583–617.

[127] D. Ozkan, K. Sagae, and L.-p. Morency, "Latent mixture of discriminative experts for multimodal prediction modeling," Computational Linguistics, no. August, pp. 860–868, 2010.

[128] D. Neiberg and J. Gustafson, "Predicting speaker changes and listener responses with and without eye-contact," in INTERSPEECH, 2011, pp. 1565–1568.

[129] L. Huang, L.-P. Morency, and J. Gratch, "A multimodal end-of-turn prediction model: learning from parasocial consensus sampling," in The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3, ser. AAMAS '11. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 1289–1290.

[130] C.-C. Lee and S. Narayanan, "Predicting interruptions in dyadic spoken interactions." in ICASSP'10, 2010, pp. 5250–5253.

[131] N. G. Ward, O. Fuentes, and A. Vega, "Dialog prediction for a general model of turn-taking," in INTERSPEECH, 2010, pp. 2662–2665.

[132] K. R. Thorisson, Natural Turn-Taking Needs No Manual: Computational Theory And Model, From Perception to Action. Dordrecht, The Netherlands: Kluwer Acedemic Publishers, 2002, pp. 173–207.

[133] N. Cathcart, J. Carletta, and E. Klein, "A shallow model of backchannel continuers in spoken dialogue," in Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics - Volume 1, ser. EACL '03. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003, pp. 51–58.

[134] A. Gravano and J. Hirschberg, "Backchannel-inviting cues in task-oriented dialogue." in INTERSPEECH. ISCA, 2009, pp. 1019–1022.

[135] L.-P. Morency, I. Kok, and J. Gratch, "Predicting listener backchannels: A probabilistic multimodal approach," in Proceedings of the 8th international conference on Intelligent Virtual Agents, ser. IVA '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 176–190.

[136] J. Nadel, K. Prepin, and M. Okanda, "Experiencing contingency and agency: First step toward self-understanding in making a mind?" Interaction Studies, vol. 6, no. 3, pp. 447–462, 2005.

[137] E. Delaherche and M. Chetouani, "Characterization of coordination in an imitation task : human evaluation and automatically computable cues," in 13th International Conference on Multimodal Interaction, 2011.

[138] X. Sun, J. Lichtenauer, M. Valstar, A. Nijholt, and M. Pantic, "A multimodal database for mimicry analysis," in Affective Computing and Intelligent Interaction, Part I, ser. Lecture Notes in Computer Science, S. D'Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds., vol. 6974. Berlin, Germany: Springer Verlag, October 2011, pp. 367–376.

[139] A. N. Meltzoff, P. K. Kuhl, J. Movellan, and T. J. Sejnowski, "Foundations for a new science of learning," Science, vol. 325, no. 5938, pp. 284–288, 2009.

[140] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, F. Nori, L. Fadiga, B. Wrede, K. Rohlfing, E. Tuci, K. Dautenhahn, J. Saunders, and A. Zeschel, "Integration of action and language knowledge: A roadmap for developmental robotics," IEEE Transactions on Autonomous Mental Development, pp. 167–195, 2010.

[141] K. Prepin and P. Gaussier, "How an agent can detect and use synchrony parameter of its own interaction with a human?" in Development of Multimodal Interfaces: Active Listening and Synchrony, A. Esposito et al., Eds., vol. 5967. Springer Berlin / Heidelberg, 2010, pp. 50–65.

[142] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy, "Creating rapport with virtual agents," in IVA '07: Proceedings of the 7th international conference on Intelligent Virtual Agents. Springer-Verlag, 2007, pp. 125–138.

[143] M. Michalowski, R. Simmons, and H. Kozima, "Rhythmic attention in child-robot dance play," in Proceedings of RO-MAN 2009, 2009.

[144] H. Kozima, M. Michalowski, and C. Nakagawa, "Keepon," International Journal of Social Robotics, vol. 1, pp. 3–18, 2009, 10.1007/s12369-008-0009-8.

[145] J. J. Diehl, L. M. Schmitt, M. Villano, and C. R. Crowell, "The clinical use of robots for individuals with autism spectrum disorders: A critical review," Research in Autism Spectrum Disorders, vol. In Press, pp. –, 2011.

[146] A. Tartaro and J. Cassell, "Playing with virtual peers: bootstrapping contingent discourse in children with autism," in Proceedings of the 8th international conference on International conference for the learning sciences - Volume 2, ser. ICLS'08. International Society of the Learning Sciences, 2008, pp. 382–389.

[147] M. Milgram and H. Atlan, "Probabilistic automata as a model for epigenesis of cellular networks," Journal of Theoretical Biology, vol. 103, pp. 523–547, 1983.

[148] R. P. N. Rao, A. P. Shon, and A. N. Meltzoff, "A bayesian model of imitation in infants and robots," In Imitation and Social Learning in Robots Humans and Animals, pp. 217–247, 2007.

[149] D. Cohen, "Probabilistic epigenesis: An alternative causal model for conduct disorders in children and adolescents," Neuroscience & Biobehavioral Reviews, vol. 34, no. 1, pp. 119 – 129, 2010.

[150] C.-c. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, P. G. Georgiou, and S. S. Narayanan, "An analysis of pca-based vocal entrainment measures in married couples' affective spoken interactions," In Proceedings of InterSpeech, no. August, pp. 3101–3104, 2011.

[151] J. F. Cohn, "Advances in behavioral science using automated facial image analysis and synthesis," IEEE Signal Processing Magazine, vol. 27, no. November, pp. 128–133, 2010.

Emilie Delaherche received the M.S. degree in Engineering from ESME Sudria (School of Engineering), Paris, 2005. She received the M.S. degree in Machine Learning and Signal Processing from the University Pierre and Marie Curie (UPMC), Paris, 2010. She is currently a PhD student in the Artificial Perception research group at the Institute of Intelligent Systems and Robotics. The topic of her thesis is the analysis of human-centered interaction dynamics. Her research interests include machine learning applications in robotics and vision and multimodal human-computer interaction.

Mohamed Chetouani is the head of the Artificial Perception research group. He received the M.S. degree in Robotics and Intelligent Systems from the University Pierre and Marie Curie (UPMC), Paris, 2001. He received the PhD degree in Speech Signal Processing from the same university in 2004. In 2005, he was an invited Visiting Research Fellow at the Department of Computer Science and Mathematics of the University of Stirling (UK). Dr. Chetouani was also an invited researcher at the Signal Processing Group of Escola Universitaria Politecnica de Mataro, Barcelona (Spain). He is currently an Associate Professor in Signal Processing and Pattern Recognition at the University Pierre et Marie Curie. His research activities, carried out at the Institute of Intelligent Systems and Robotics, cover the areas of non-linear signal processing, feature extraction, pattern classification and fusion for human centered interaction analysis: verbal and non-verbal communication, physiological signals. He is an Associate Editor of the Cognitive Computation Journal (Springer) and a Guest Editor for a Special Issue in Speech Communication on "Non-Linear and Non-Conventional Speech Processing". He served as a chairman of several international workshops on non-linear speech processing, human-robot interaction and human centered multimodal signal processing.

Ammar Mahdhaoui received the BACHELOR degree in computer science from the University of Monastir (FSM), Tunisia in 2006, and the MASTER degree in computer science from the University of Grenoble, France, in 2007. He received the PHD degree for his studies on Machine Learning and Speech Processing in 2010, from the University Pierre and Marie Curie (UPMC), one of French's repeatedly highest ranked and among its first Excellence Universities. He was a Ph.D student with the Institute of Intelligent Systems and Robotics (ISIR UMR7222), UPMC, from 2007 to 2010, and a Doctor Research Assistant with UPMC, from 2010 to 2011. He is actually a Research Engineer, post-doc, with France Télécom Orange-Labs, in the Tangible interactions enhancing user experience group (TAP). His research activities cover the areas of Pattern Recognition, Speech Processing and Human-Machine Interaction. Dr. Mahdhaoui is a member of different scientific societies (e.g., ISCA, AFIA, AFCP, ISIS).

David Cohen received a M.S. in neurosciences from the UPMC and the Ecole Normale Supérieure in 1987, and a M.D. from Necker School of Medicine in 1992. He specialized in child and adolescent psychiatry and certified in 1993. His first field of research was severe mood disorders in adolescent, topic of his PhD in neurosciences (2002). He is Professor at the UPMC and head of the department of Child and Adolescent Psychiatry at La Salpêtrière hospital in Paris. His group runs research programs in the field of autism and other pervasive developmental disorder, severe mood disorder in adolescent and childhood onset schizophrenia and catatonia. He is a member of the International Association of Child and Adolescent Psychiatry and Allied Disciplines, the European College of Neuro-Psychopharmacology, the European Society of Child and Adolescent Psychiatry, and the International Society of Adolescent Psychiatry.

Catherine Saint-Georges received a MS degree in Psychopathology and Neurosciences from UPMC (Paris 6 University) and a MD from Paris XI University of Medicine in 2000. She specialized in child and adolescent psychiatry in 2001. She received the PhD degree in neuroscience from ED3C School (UPMC) in 2011. Her PhD topic dealt with dynamic of early interactions of autistic infants with their parents, with a special focus on synchrony, reciprocity and motherese. She's currently post-doc at the Institute of Intelligent Systems and Robotic (CNRS UMR 7222), where she's involved in research on autistic children and early mother-infant synchrony in infants at risk for autism. She's also Child Psychiatrist at la Salpêtrière Hospital and in André Boulloche Center (Paris).

Sylvie Viaux received a M.S. in Neurosciences from the UPMC in 2011, and a M.D. from Tours School of Medecine in 2004. She specialized in Child and Adolescent Psychiatry and certified in 2005. She is child psychiatrist in the Department of Child and Adolescent Psychiatry at Pitié Salpêtrière University Hospital in Paris. Her field of research is early mother infant interactions in context of fetal scan soft markers and neglect and collaboration about interaction modelisation with the ISIR Insitute in UPMC. She is member of the World Association of Infant Mental Health and the French Society of Child and Adolescent Psychiatry.