Emerging Trends in Drugs, Addictions, and Health 5 (2025) 100176

Contents lists available at ScienceDirect

Emerging Trends in Drugs, Addictions, and Health

journal homepage: www.elsevier.com/locate/etdah

Child-therapist acoustic synchrony and response trajectories in autism intervention: an AI-based automated analysis using dynamic systems theory and affective computing

Giulio Bertamini [a,b,c,*], Silvia Perzolli [b], Arianna Bentenuto [b], Cesare Furlanello [d], Mohamed Chetouani [c], David Cohen [a,c], Paola Venuti [b]

[a] *Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière University Hospital, Sorbonne University, 47-83 Bd de l'Hôpital, Paris, 75013 Île-de-France, France*
[b] *Laboratory of Observation, Diagnosis, and Education, Department of Psychology and Cognitive Science, University of Trento, Via Matteo del Ben, 5B, Rovereto, 38068 TN, Italy*
[c] *Institute for Intelligent Systems and Robotics, Sorbonne University, Pyramide, T55, 4 Pl. Jussieu 65, Paris, 75005 Île-de-France, France*
[d] *HK3 Lab, Piazza Manifattura, 1, Rovereto, 38068 TN, Italy*

ARTICLE INFO

Handling Editor: Dr. Ornella Corazza.

Keywords:
Deep learning
CRQA
Interpersonal synchrony
Affective computing
Automated analysis
Autism intervention
Response trajectories

ABSTRACT

Introduction: Child-clinician interpersonal dynamics are central to psychotherapy and are increasingly acknowledged as key elements in autism intervention. However, quantitatively studying fine-grained aspects such as the child-clinician synchrony patterns poses challenges, limiting translational research. Moreover, synchrony is rarely investigated with a long-term perspective. This study employed an AI-based, fully automated computational pipeline to analyze child-clinician interpersonal acoustic synchrony through the lens of complex dynamic systems and affective computing.
Methods: We followed 25 autistic preschoolers over one year of Naturalistic Developmental Behavioral Intervention (NDBI). Three 60-minute intervention sessions, at the beginning, after three months, and after one year, were analyzed second-by-second, totaling 75 videos. After AI-based automatic speech segmentation, acoustic synchrony was assessed using Cross-Recurrence Quantification Analysis to derive interaction metrics over the entire therapy sessions employing affective prosodic features. Robust Bayesian correlation analysis was used to explore the relationship between affective acoustic synchrony and developmental learning rates at different time points.
Results: No significant associations were found at baseline, while correlations emerged after three months and became more pronounced at one year. Early in therapy, interactions with a stronger internal structure, particularly in loudness, spectral dynamics, and voice quality, were linked to higher developmental gains. After one year, the relationship between synchrony and response shifted toward metrics reflecting transition dynamics and stability. Associations with fine-grained spectral features particularly characterized this phase.
Discussion: Specific and different synchrony aspects were associated with therapy response trajectories both in the initial and latter phases of therapy. Acoustic features involved in intervention response are known to participate in the emotional content of speech, highlighting the contribution of affective aspects to therapy.
These findings provide valuable insights into the role of interpersonal synchrony in autism intervention and underscore the potential of computational methods in monitoring treatment progress.

1. Introduction

1.1. Autism intervention

Autism is a neurodevelopmental condition characterized by biological and environmental aspects. At the behavioral level, autistic individuals exhibit variations in social communication and interaction, as well as restricted repetitive patterns of behaviors and interests that may impact their life quality and adaptation abilities (APA, 2022). In children, these traits may influence the experience-dependent,

* Corresponding author at: Via Matteo del Ben, 5B, Rovereto, 38068 TN, Italy.
E-mail address: giulio.bertamini@unitn.it (G. Bertamini).

https://doi.org/10.1016/j.etdah.2025.100176
Received 16 April 2025; Received in revised form 13 June 2025; Accepted 16 June 2025
Available online 24 June 2025

experience-expectant developmental processes mediated by the interpersonal environment, ultimately impacting the acquisition of developmental milestones (Klin et al., 2020; Shultz et al., 2018; Pelphrey et al., 2011). Naturalistic Developmental Behavioral Interventions (NDBIs) comprise a variety of empirically based models of intervention that aim at promoting learning and reducing social and adaptation difficulties by fostering adaptive exchanges while working on developmental objectives through different techniques (D'agostino et al., 2023). Research has demonstrated that autism intervention can lead to significant improvements in a variety of outcome dimensions, spanning social emotional functioning and core challenges associated with autism (Song et al., 2025; Sandbank et al., 2023). However, not all children respond to intervention in the same way, and there is a growing need to identify early markers of treatment efficacy, moderator variables, and mechanisms of action to enhance intervention efficacy (Jobin, 2020; Vivanti et al., 2014). This aspect is even more crucial if considering autism in its developmental aspects (Nelson et al., 2023).

1.2. *Child-clinician interpersonal dynamics*

The developmental framework of NDBIs emphasizes interpersonal interplay as a key component in effective therapy implementation (Schreibman et al., 2015). NDBIs focus on promoting communication and social skills through structured and dynamic interactions between children and therapists. One important aspect of these interactions is interpersonal synchrony, which refers to the coordinated and temporally aligned exchange of behavioral and physiological signals between individuals, such as shared gaze, vocalizations, gestures, and affect (Feldman, 2012). Synchrony has been extensively studied in the context of typical and atypical development (Mayo and Gordon, 2020; McNaughton and Redcay, 2020; Leclère et al., 2014), and its contribution has also been studied with respect to psychotherapy (Koole and Tschacher, 2016). Synchrony is thought to facilitate social bonding, communication, mutual understanding, and to be implied in emotion co-regulation, and is both a unimodal and cross-modal phenomenon (Butler, 2011; Feldman, 2007). In the acoustic domain, it involves the temporal coordination of vocal features such as pitch, loudness, rhythm, and more nuanced spectral components between interacting partners, which are known to convey affective aspects (Weninger, et al., 2013). However, despite its theoretical importance, little is known about its role in autism therapy and, more in general, in developmental clinical contexts.

1.3. *Challenges in quantitative analysis*

While the child-clinician interpersonal dynamics are increasingly recognized as central to therapy success (Schreibman et al., 2015), quantitatively studying these fine-grained patterns of interaction poses significant challenges, particularly within the realm of autism research. In fact, this type of investigation has been traditionally limited to observational, qualitative assessment manually performed by trained observers. While providing rich insights into the therapeutic process, these techniques suffer from limited quantification, objectivity, and require intense human labor, limiting both their scalability and their translational applications. Furthermore, such methods cannot easily quantify the complexity and dynamics of real-time naturalistic interactions as a whole, as they rely more on broad, general theoretic constructs than on data-driven insights (Delaherche et al., 2012). For this, dynamic systems theory could provide a solid framework for the quantitative investigation of complex non-linear relationships between various interpersonal signals, from physiology to behavior (Webber and Marwan, 2015). One powerful technique is represented by Cross-Recurrence Quantification Analysis (CRQA, Coco and Dale, 2014), and there is a growing need to employ computational approaches in clinical research to disentangle the complexity of interpersonal aspects (Dale et al., 2011), with particular emphasis on process aspects during therapy that may ultimately impact on its outcomes (Shockley, 2005). CRQA is a non-linear method that quantifies the shared dynamics between two time series by identifying states in one series that recur in the other over time. Unlike linear correlation, CRQA captures the temporal structure and coordination patterns of coupled systems. It provides several quantitative metrics such as recurrence rate (the proportion of shared states), determinism (the predictability of shared patterns), laminarity (the presence of stable, unchanging states), and entropy (the complexity of these patterns), allowing for a detailed characterization of the degree, stability, and structure of coordination between interacting partners. Further, the ratios of determinism to recurrence and laminarity to determinism provide composite summary metrics reflecting complementary aspects of system synchronization. The Determinism-to-Recurrence ratio reflects the proportion of recurrent points that form predictable, structured sequences, representing the extent to which interactions are governed by stable, coordinated patterns rather than random or transient matches. Complementary, the Laminarity-to-Determinism reflects the balance between stability and flexibility in interpersonal dynamics. Taken together, these metrics offer complementary insights into interpersonal synchrony dynamics, capturing the balance between structured coordination and stable engagement (Marwan et al., 2007). In clinical contexts, this balance may be particularly relevant. Given the limited research specifically addressing these facets and their implications for clinical practice, there is a need for exploratory studies that characterize fine-grained synchrony dynamics over both short and long timescales, providing an empirical basis for developing data-driven clinical hypotheses to guide future research.

1.4. *The role of affective aspects and speech*

Affective synchrony and the alignment of emotional states between interacting individuals are recognized as important factors in human social interactions and crucial to child development (Bourvis et al., 2018; Leclère et al., 2014; Weisman et al., 2015). In the context of autism, where early social communication difficulties are significant, promoting affective synchrony between the child and therapist may be particularly crucial for facilitating therapeutic engagement and improving intervention outcomes (Fusaroli et al., 2017), as is the case in studies on the role of parent-child synchrony for emotional and social development (Bourvis et al., 2018; Feldman, 2012).

Interestingly, studies related acoustic synchrony to alliance (the collaborative and affective bond between therapist and client, encompassing agreement on therapy goals, tasks, and the development of mutual trust and understanding) and outcomes during psychotherapy (Schoenherr et al., 2021; Nof et al., 2020). Additional evidence indicated that synchrony is associated with therapist empathy and encodes affective aspects in clinical dyads (Imel et al., 2014).

One potential marker of therapeutic success is the degree of synchrony between the child and therapist, especially concerning affective aspects. By modeling and analyzing these interaction dynamics over time, we may be able to identify patterns that are associated with positive or more positive treatment outcomes, producing insights into therapy functioning and evolution. Research has shown that the degree of synchrony in early interactions may serve as an indicator of the development of social communication skills in children with autism (Warlaumont et al., 2010; Oller et al., 2013). Specifically, the presence of structured and predictable vocal interactions, reflecting greater emotional engagement and coordination, has been linked to higher rates of developmental progress (Cohen et al., 2013). There is also initial evidence that the child-therapist alliance may be relevant for intervention efficacy (Mössler et al., 2017).

Dynamic systems theory has been employed to study human vocal interactions (Fusaroli and Tylén, 2016). CRQA was used to study the dynamics of child-caregiver vocal interactions and its role during child development, showing that the recurrence profiles are associated with linguistic developmental outcomes in both autism and typical development (Warlaumont et al., 2014) and learning (Duong et al., 2024). However, little is known about the child-clinician dynamics during autism therapy.

1.5. Aim

The aim of this study was to explore the relationship between child-therapist interpersonal synchrony and therapy response trajectories during autism intervention over a longitudinal sample of preschool children undergoing a Naturalistic Developmental Behavioral Intervention (NDBI). Specifically, we aimed to model acoustic synchrony by means of affective computing and complex dynamic systems theory. We were particularly interested in identifying temporal patterns of synchrony that may signal favorable response trajectories from a developmental standpoint and may merit further investigation as intervention mediators. To this end, we employed an AI-based system (Bertamini et al., 2025) capable of automatically segmenting child–clinician speech on a second-by-second basis within naturalistic clinical settings. This model, consisting of a 2-layer deep learning convolutional architecture trained on clinical data for metric learning, serves as a foundation for the automatic annotation of audio signals recorded by ambient microphones by means of similarity-based classification. Basically, the system sequentially analyzes the audio stream and recognizes human voice and whether it belongs to a child or an adult based on spectrogram features. Downstream, we extracted prosodic and affective acoustic features and modeled child-therapist synchrony non-linearly across entire therapy sessions through CRQA. Finally, we conducted a robust exploratory Bayesian correlation analysis to identify features associated with treatment response at various time points.

2. Methods

2.1. Participants and procedure

A total of $N = 25$ European/Italian preschool autistic children (22 males) and from lower- to upper-middle socioeconomic backgrounds participated in the study. The children (mean chronological age=37.72 months, SD=10.06, range=[23–56]; mean developmental age=26.08 months, SD=7.23, range=[14–45]) were followed longitudinally for approximately one year (mean duration=15.2 months, SD=4.9) as part of an intervention program at the Laboratory of Observation, Diagnosis, and Education (ODFLab), University of Trento, Italy. All participants underwent an initial clinical assessment using standardized gold-standard diagnostic tools for neurodevelopmental conditions. Following this, each child began a personalized NDBI-based therapy program (2–4 h per week) conducted by licensed therapists trained in this approach. The intervention model implemented by ODFLab is grounded in the NDBI framework and follows the guidelines of the Italian National Health Institute. Specifically, therapists were certified in one or more established NDBI models, including the Early Start Denver Model (ESDM), Joint Attention Symbolic Play Emotion Regulation (JASPER), and Parent-mediated Communication-focused Treatment (PACT). Intervention plans were individualized based on each child's functional assessment, with tailored learning objectives, interaction strategies, expected evolving trajectories, and caregiver involvement strategies. For the purposes of this study, analyses were restricted to individual therapy sessions involving direct therapist-child interactions. A second complete clinical evaluation was carried out at the end of the intervention period to assess changes in developmental functioning and symptom severity. ASD diagnoses were established and confirmed in line with DSM-5 criteria and through the Autism Diagnostic Observation Schedule-2 (ADOS-2, Lord et al., 2012), administered by certified clinicians (mean ADOS-2 Calibrated Severity Score=5.16 (2.78)). Module 2 was employed in 18 cases, while seven children underwent the initial evaluation with Module Toddler. Developmental profiles were assessed through the Griffiths Mental Development Scales-Edition Revised (GMDS-ER, Luiz et al., 2006) (mean General Quotient=71.96 (14.91); mean Locomotion=78.32 (18.00); mean Personal-social=68.88 (20.42); mean Language=55.20 (25.05); mean Eye-hand coordination=73.72 (18.75); mean Performance=88.08 (22.73).

The clinical team included 17 European/Italian therapists (14 females), all trained under the same NDBI model, working under consistent supervision. Inclusion criteria for participation in this study included: (i) a DSM-5 diagnosis of ASD made before the age of five; (ii) availability of two complete clinical assessments approximately one year apart; (iii) uninterrupted participation in the NDBI intervention during the study period; and (iv) the presence of vocal output (linguistic or non-linguistic). The study was conducted in accordance with the latest revision of the Declaration of Helsinki (World Medical Association, 2024) and was approved by the Ethics Committee of the University of Trento (protocol number: 2020–042). Written informed consent was obtained from all participants' caregivers. Data collection spanned from 2015 to 2021.

For each child, three therapy sessions were analyzed, for a total of 75 videos. Therapy sessions were video-recorded using ceiling-mounted cameras, and audio was captured via single room microphones. One session was selected immediately following the baseline evaluation (T0 months), a second session was selected after about 3 months of intervention (T3 months), and a third session was collected slightly before the follow-up assessment (T12 months). The entire duration of each session (approximately 60 min) was analyzed using the automated processing pipeline described in the following sections.

2.2. Measures

Developmental progress was quantified using Developmental Learning Rates (LRs) (Klintwall et al., 2013), which provide a standardized estimate of change in developmental functioning over time. Children's developmental profiles were assessed via the Griffiths Mental Development Scales-Edition Revised (GMDS-ER, Luiz et al., 2006). In addition to the developmental Z-quotient scores (mean=100; SD=15), the GMDS-ER yields developmental age equivalents measured in months to compare children's chronological and developmental ages with respect to developmental milestones. They can be used to compute LRs, defined as the change in developmental age equivalents (in months) divided by the number of months elapsed between the two clinical assessments. For each child, LR yielded a rate-like metric that captures the slope of developmental change over time. By definition, an LR of 1 reflects typical developmental pacing (i.e., one month of developmental gain per month of real time). LRs below 1 indicate slower developmental progress (i.e., increasing developmental lag), while values above 1 suggest accelerated or "catch-up" development during the intervention period, narrowing their developmental gap. Therefore, LRs allow for the evaluation of response trajectories while accounting for developmental changes that may occur naturally over time, independent of the intervention itself.

2.3. Automated speech segmentation

A deep learning system performed the automated segmentation of therapy sessions' audio recordings. The system was validated in Bertamini et al. (2025) and consists of a two-layer classification pipeline trained on naturalistic clinical data. Two shared-weight neural networks perform the similarity based second-by-second classification on Mel-Frequency Cepstral Coefficients (MFCCs) spectrograms. The first network identifies segments containing human voice presence, whereas the second distinguishes speakers. The architecture consists of a set of dilated 2d-convolutions incorporating channel-wise attention mechanisms. The network is trained to learn an optimized latent embedding space to project sound patterns into feature vectors such as those belonging to the same class are clustered together, minimizing intra-class distances, and far from other classes, maximizing inter-class separation. This process leverages learning by generating triplets of anchor-positive-negative examples, allowing effective learning also in smaller datasets compared to traditional deep learning solutions. Further, these models are also suitable for domain adaptation and few-shot learning, which is relevant for application on novel data without retraining extensively the system. Developed using noisy, unstructured interactions between autistic preschoolers and clinicians, our system is specifically adapted to handle non-linguistic vocalizations, common in this population. Validation was conducted through a robust cross-validation framework, with the voice activity detection module achieving strong performance. Speaker diarization also performed reliably. Agreement with human annotations in masked conditions was high, with a three-rater Fleiss's kappa of 0.87, indicating strong consistency. The DL model also proved to be highly scalable and flexible, requiring just a few annotated examples to align to different data and contexts successfully.

To monitor reliability in the current study, we extracted a random sample of 300 one-second vocal segments, balanced by speaker and session, which were then manually annotated by a rater masked to model predictions. Performance confirmed to be high (balanced accuracy=0.89; F1-score=0.89; sensitivity=0.89; specificity=0.88; precision=0.88; MCC=0.78; AUCROC=0.89; AUCPR=0.84). No noise segments were present in the evaluated subset. Inter-rater agreement accounting for chance agreement and measured via Cohen's kappa was good ($k = 0.77$).

2.4. Feature extraction

In the present study, we employed openSMILE (Open-source Speech and Music Interpretation by Large-space Extraction) to extract acoustic features from child–caregiver vocal interactions. This open-source framework is widely recognized for its ability to automatically derive a broad range of acoustic parameters, spanning both low-level descriptors (LLDs) and higher-order functionals (Eyben et al., 2016). OpenSMILE has become a standard tool across various domains, including affective computing, speech science, and clinical research (Fusaroli et al., 2017; Oller et al., 2013). Specifically, we used the eGeMAPSv0 feature set (Eyben et al., 2016), which was developed for applications in paralinguistic and affect-related speech processing, with a focus on clinical and psychological contexts. A summary of the 25 extracted LLDs is provided in Table 1. Acoustic features were extracted from second-level, automatically annotated audio segments, producing distinct time series for the child and the caregiver. These were then aggregated into 500 ms intervals. Periods of mutual silence lasting 25 s or more were removed. When a speaker was silent during a given segment, features were computed from a silent audio placeholder to maintain consistent time indexing. In cases of overlapping vocalizations, identical feature values were assigned to both speaker vectors. Prior to applying CRQA, each time series was independently standardized to ensure comparability across individuals.

2.5. Cross-recurrence quantification analysis

CRQA is a non-linear method for analyzing time-series data, grounded in the framework of complex dynamical systems theory. It extends classical Recurrence Quantification Analysis (RQA) to the simultaneous analysis of two distinct time series, allowing for the quantification of temporal coupling between interacting systems (Marwan and Kurths, 2002). Unlike traditional approaches such as cross-correlation, CRQA can capture recurrent, non-linear, and non-stationary dynamics, making it particularly well-suited for studying real-world, multimodal interactions (Varni et al., 2017). It enables the exploration of both synchronous and time-lagged patterns of coordination, yielding a set of metrics that characterize shared temporal structures and adaptive processes. As such, CRQA provides valuable insights into mutual regulation and interpersonal alignment, especially in the context of naturalistic human communication (Webber and Zbilut, 2005). A summary of CRQA metrics used in this study is provided in Table 2. We specifically focused on two composite metrics: the Laminarity-to-Determinism ratio (LAM/DET) and the Determinism-to-Recurrence rate ratio (DET/REC). These metrics offer a more integrative perspective on the dynamics of interaction, with LAM/DET reflecting the balance between stable, persistent states and overall predictability and DET/REC capturing the extent to which recurrent patterns are structured and deterministic rather than random or fragmented.

Table 1
openSMILE low-level descriptors from the eGeMAPSv02 feature set (Eyben et al., 2016).

Feature	Description
Loudness	Perceived intensity of sound, related to amplitude
alphaRatio	Ratio of energy in the low frequencies (alpha range)
hammarbergIndex	Spectral balance, indicating presence of harmonics in speech
mfcc1 to mfcc4	First four Mel-Frequency Cepstral Coefficients (MFCCs)
F0semitoneFrom27.5H	Fundamental frequency (F0) deviation from 27.5 Hz in semitones, indicating pitch change
shimmerLocaldB	Local amplitude variation (shimmer), indicating breathiness or roughness
HNRdBACF	Harmonics-to-noise ratio (HNR), reflecting vocal fold vibration smoothness
logRelF0-H1-H2	Logarithmic ratio between F0 and harmonic components, related to voice quality
logRelF0-H1-A3	Logarithmic ratio of F0 to harmonics, capturing voice roughness or breathiness
spectralFlux	Changes over time in the sound spectrum changes, measures how quickly or sharply the sound changes
F1, F2, F3 frequencies	Frequency of the formants (F1, F2, F3), related to vowel sound production
F1, F2, F3 bandwidths	Bandwidth of the formants (F1, F2, F3), indicative of vocal tract resonance
F1, F2, F3 amplitudes	Amplitude of formants (F1, F2, F3) relative to F0: vocal resonance and tonal quality

2.6. Data analysis plan

To answer our research question, we employed a rigorous Data Analysis Plan (DAP) based on CRQA and Bayesian correlation analysis. Given the exploratory nature of this study, the limited sample size, and the high dimensionality of the data, we opted for a Bayesian approach to better account for uncertainty, reduce the risk of overfitting, and generate more informative estimates of effect sizes. This framework also enabled the incorporation of prior knowledge and facilitated interpretation even under constrained data conditions (Kruschke, 2014). Correlation analysis was performed to link intervention LRs and the two composite CRQA metrics (DET/REC, LAM/DET), at T0 months, T3 months, and T12 months. Further, we considered T0-T3 and T0-T12 residualized change scores. Residualized change scores were derived by extracting residuals from the linear model *post* ~ *pre* to take into account variations over time conditioned on baseline value. The Bayesian correlation analysis was conducted on standardized data using medium-strength informative priors to stabilize parameter estimation. For each estimate, we reported 95 % Credible Intervals (CIs). CIs indicate the range of parameter values that contain the true parameter with a given probability, conditional on the observed data and the specified prior distribution, and are derived directly from the posterior distribution of the parameter. Correlations were computed on rank-based inverse normal transformed data in case of violations of normality (Bishara and Hittner, 2012). To enhance robustness and mitigate potential biases due to parameter sensitivity in CRQA, we conducted a sensitivity analysis by repeating the CRQA across a set of eight different parameter combinations. These involved variations in embedding dimension ([2, 4]), time delay ([2, 4]), and fixed radius for Euclidean

Table 2
Summary of CRQA metrics (Marwan et al., 2007; Webber and Zbilut, 2005; Shockley, 2005; Coco and Dale, 2014).

Metric	Description	Interpretation	Increase	Decrease
Recurrence rate (REC)	Proportion of recurrent points in the recurrence plot	How often two time-series revisit the same state	Increased synchronization or common influences	Weaker interaction or increased randomness in the system
Determinism (DET)	Proportion of recurrence points forming diagonal lines	Higher DET suggests more predictable and structured interactions	Stronger temporal dependence and shared dynamic patterns	More stochastic behavior or less structured coordination
Laminarity (LAM)	Proportion of recurrence points forming vertical lines	Reflects intermittency: higher LAM reflects indicates tendency of systems to remain in similar states	Increased stability, prolonged, sustained interaction or slowly gradual changes. It may also indicate rigid behavior	More dynamic or less constrained interaction, perhaps also more unpredictable or unstructured
Laminarity to Determinism ratio (LAM/DET)	Ratio of LAM to DET. Balances intermittency and predictability	Higher values indicate dominance of stability over predictable sequences	Prolonged stable states with fewer or smoother transitions with respect to overall predictability	More dynamic, less persistent interactions, less stable states with respect to overall predictability
Determinism to Recurrence Rate ratio (DET/REC)	Ratio of DET to REC. Measures structure relative to recurrence	Higher values suggest more structured, less random recurrence patterns	Interactions are increasingly systematic and rule-governed	Interactions are becoming more random or less structured

distance ([1, 1.5]). This multi-parameter approach allowed us to capture varying complexity in temporal dynamics and recurrence thresholds. Correlations with a Bayes Factor (BF) greater than 3 were retained, and only those consistently observed in at least four of the eight tested configurations (thus demonstrating stability) were selected for further interpretation. BF was interpreted using the Lee & Wagenmakers (2014) scheme, with BF>3 suggesting moderate evidence in favor of the alternative hypothesis and BF>10 indicating strong evidence. This strategy ensured that reported effects were not only statistically meaningful but also robust across variations in analysis assumptions and modeling parameters. CRQA was performed using PyRQA (Rawald et al., 2017). Statistical analysis was performed in R 4.4.3 using the BayesFactor package (Morey et al., 2015).

3. Results

Relevant results (BF>3 and CIs not containing zero) of the correlation analysis are reported in Table 3. The complete set of correlations performed can be found in the Supplementary Material.

The correlation analysis revealed a number of correlations between the selected synchrony metrics and LR that showed a specific organization over time and spanned different acoustic features. No significant correlations emerged at baseline, but correlations emerged after three months and evolved at one year.

Specifically, in the first three months of intervention, evidence for a weak-to-moderate positive correlation was found between the LR and the residualized change of the DET/REC ratio among acoustic features involving harmonic-to-noise ratio (HNRdBACF: $r = 0.38$), harmonic structure (logRelF0-H1-H2: $r = 0.40$), sound intensity (Loudness: $r = 0.44$), and spectral dynamics (spectralFlux: $r = 0.43$). These correlations suggest that an increase in the deterministic structure of recurrences in the first months of intervention is already associated with a greater LR at one year. Specifically, for loudness and spectral flux, the evidence was strong (BF>10). The 95 % credible intervals for the Bayesian correlations, which represent the range within which the true parameter lies with 95 % probability given the observed data and prior information, did not include zero, indicating evidence for a meaningful association.

At 3 months, some correlations emerged, consistently related to the DET/REC ratio for voice quality (logRelF0-H1-H2: $r = 0.38$), speech volume (Loudness: $r = 0.48$), and spectral dynamics (spectralFlux: $r = 0.41$). Correlations were weak-to-moderate, and positive. Also in this case, the BF for the correlation between learning rate and both spectral flux and loudness DET/REC suggested strong evidence. CIs excluded zero, further indicating supporting a meaningful association.

At 12 months, the DET/REC ratio for the first MFCC was positively and moderately correlated with the LR ($r = 0.40$; CI=[0.06, 0.67]).

Regarding residualized changes in synchrony metrics between baseline and 12 months, we found weak-to-moderate correlations in different acoustic domains. These included spectral features such as alphaRatio ($r = 0.36$) and spectralFlux ($r = 0.39$), fundamental frequency measured by F0semitoneFrom27.5 Hz ($r = 0.39$), formant characteristics including F1amplitudeLogRelF0 ($r = 0.41$), F1frequency ($r = 0.39$), F2amplitudeLogRelF0 ($r = 0.41$), and F3amplitudeLogRelF0 ($r = 0.40$), speech quality measures like harmonics-to-noise ratio (HNRdBACF: $r = 0.38$) and shimmerLocaldB ($r = 0.38$), intensity indexed by Loudness ($r = 0.37$), and cepstral features including mfcc2 ($r = 0.43$), mfcc3 ($r = 0.39$), and mfcc4 ($r = 0.43$). Unlike previous analyses, these correlations specifically involved the LAM/DET metric from CRQA, which reflects the balance between stable and structured coordination. Bayes Factors for the correlations with the second and fourth MFCCs indicated strong evidence of association, with values exceeding 10. All CIs excluded zero, indicating evidence for a meaningful association.

After 12 months of intervention, correlations involving the same set of features emerged for both residualized changes and coherent strengths. Correlations with the largest supporting evidence, with a BF exceeding 10, involved F1 amplitude ($r = 0.43$), and the second ($r = 0.43$) and fourth ($r = 0.44$) MFCCs. Also in this case, all CIs excluded zero, indicating evidence for a meaningful association. These correlations suggest that a long-term greater presence (and a long-term increase) of prolonged stable states with fewer, smoother transitions is associated with higher LR (developmental improvements).

Taken together, these results highlight that dynamic changes in acoustic synchrony as reflected by specific CRQA metrics across multiple speech features are meaningfully and differentially associated with learning rate improvements over time, providing insights about specific aspects of affective temporal coordination in developmental progress that characterize intervention.

4. Discussion

This exploratory study investigated the child-therapist interpersonal synchrony in terms of affective prosodic aspects of their acoustic interaction from the perspective of complex dynamic systems and affective computing. Specifically, we were interested in disclosing associations between the synchrony dynamics and response trajectories over the evolution of the therapeutic paths during a developmental intervention for autistic preschoolers. Our analysis was grounded in an AI-based system that allowed the automated processing of entire therapy sessions to derive synchrony metrics, effectively overcoming the limitations of observational research with respect to quantification and the labor-intensive nature of manual analysis. Our findings revealed nuanced relationships between specific synchrony patterns and developmental trajectories, shedding light on how early acoustic synchrony in therapeutic settings might serve as an indicator of treatment efficacy. We found correlations both between learning rates at different points in

Table 3
Bayesian correlational analysis for the association between DET/REC and LAM/DET of acoustic features and response trajectories (LR).

Time	Feature	BF	r	CI lower	CI upper
DET/REC					
Change(0 m,3 m)	HNRdBACF	5.05	0.38	0.04	0.66
Change(0 m,3 m)	logRelF0-H1-A3	7.14	0.40	0.07	0.67
Change(0 m,3 m)	Loudness	21.54	0.44	0.12	0.70
Change(0 m,3 m)	spectralFlux	11.63	0.43	0.10	0.69
T3m	logRelF0-H1-H2	4.80	0.38	0.04	0.65
T3m	Loudness	42.14	0.48	0.16	0.72
T3m	spectralFlux	14.38	0.41	0.08	0.68
T12m	mfcc1	6.33	0.40	0.06	0.67
LAM/DET					
Change(0 m,12 m)	alphaRatio	3.80	0.36	0.02	0.64
Change(0 m,12 m)	F0semitoneFrom27.5Hz	5.66	0.39	0.05	0.66
Change(0 m,12 m)	F1amplitudeLogRelF0	9.44	0.41	0.08	0.68
Change(0 m,12 m)	F1frequency	6.50	0.39	0.06	0.66
Change(0 m,12 m)	F2amplitudeLogRelF0	8.60	0.41	0.08	0.68
Change(0 m,12 m)	F3amplitudeLogRelF0	7.25	0.40	0.06	0.67
Change(0 m,12 m)	HNRdBACF	5.32	0.38	0.05	0.66
Change(0 m,12 m)	logRelF0-H1-H2	3.63	0.36	0.01	0.64
Change(0 m,12 m)	Loudness	4.37	0.37	0.03	0.65
Change(0 m,12 m)	mfcc2	10.17	0.43	0.10	0.69
Change(0 m,12 m)	mfcc3	5.58	0.39	0.05	0.66
Change(0 m,12 m)	mfcc4	11.93	0.43	0.10	0.69
Change(0 m,12 m)	shimmerLocaldB	5.23	0.38	0.04	0.66
Change(0 m,12 m)	spectralFlux	5.69	0.39	0.05	0.66
T12m	alphaRatio	4.44	0.37	0.03	0.65
T12m	F0semitoneFrom27.5Hz	5.41	0.39	0.05	0.66
T12m	F1amplitudeLogRelF0	11.40	0.43	0.10	0.69
T12m	F1frequency	6.98	0.40	0.06	0.67
T12m	F2amplitudeLogRelF0	8.68	0.41	0.08	0.68
T12m	F3amplitudeLogRelF0	8.49	0.41	0.08	0.68
T12m	HNRdBACF	5.07	0.38	0.04	0.65
T12m	logRelF0-H1-H2	3.95	0.36	0.02	0.64
T12m	Loudness	5.16	0.38	0.04	0.66
T12m	mfcc2	11.09	0.43	0.11	0.69
T12m	mfcc3	7.58	0.41	0.08	0.68
T12m	mfcc4	16.04	0.44	0.12	0.70
T12m	shimmerLocaldB	5.00	0.38	0.04	0.66
T12m	spectralFlux	6.86	0.40	0.07	0.67

BF: Bayes Factor. r: Pearson's r correlation coefficient. CI lower: lower limit of the 95 % credible interval. CI upper: upper limit of the 95 % credible interval.

therapy and between changes over time (adjusted for baseline) and developmental outcomes.

Across different phases of the intervention, a number of correlations emerged that involved different aspects of child-therapist synchrony and acoustic dimensions. Initially, no correlation emerged at baseline. Given that baseline sessions have been selected at the very beginning of therapy, this may reflect the absence of mutual engagement and a situation where no child-therapist interaction routines were yet established or scaffolded, in line with evidence on interpersonal synchrony (Kinreich et al., 2017).

Notably, both correlations about changes in the first months of intervention and at three months regarded synchrony aspects involving the presence of deterministic patterns in the recurrence structure on a number of acoustic features. Therefore, more successful therapies were already associated with a higher degree of predictability in the interaction patterns between the child and therapist in the first period of the intervention, as reflected by the structure of their vocal exchanges. These patterns suggest that more structured and consistent communication is linked to greater developmental progress over time, potentially indicating a foundation for more effective therapeutic engagement and better outcomes. Correlations involved acoustic features related to voice quality, loudness, harmonic structure, and spectral dynamics. In summary, our results pointed out that at the beginning of the intervention, it may be important to structure an exchange characterized by affective attunement equipped with a consistent structure to enhance therapy response. In the first phase of the intervention, the strongest evidence of the association with response involved loudness and spectral dynamics.

When we looked at correlations and considered changes with respect to baseline values after one year, a different picture emerged. Correlations specifically involved another aspect of child-therapist synchrony, that is the balance between intermittency/stability of the exchange with respect to its predictability. A higher value in this metric indicates that the vocal interaction is characterized by more prolonged stable states relative to strictly deterministic sequences, suggesting smoother transitions and possibly a less deterministic interaction structure. Acoustic features involved in this phase included voice quality and dynamics, pitch and formants, and spectral characteristics, further implicating the affective content of speech (Ververidis and Kotropoulos, 2006). Among these features, aspects related to MFCCs and first formant amplitude showed the most prominent evidence.

Notably, acoustic features emerging from our analysis were associated with emotional expression of speech in previous research works (Kamiloğlu et al., 2020; Eyben et al., 2015; Guzman et al., 2013; Scherer, 2013; Weninger et al., 2013; Patel et al., 2011, 2010), and can be considered markers of affective content, also in autism (Hubbard et al., 2017). Also MFCCs, despite being less directly interpretable, reflect fine-grained spectral characteristics and are extensively used in emotion recognition from a computational perspective, with excellent levels of accuracy in determining emotional content dimensions (Biswas et al., 2023). Moreover, research about the interconnection of speech and emotion has highlighted a complex picture and a high degree of feature interdependence, which may explain the aspects emerged from our analysis (Chen et al. 2012; Sundberg et al., 2011).

Taken together, and considering the strength of the associations, loudness and spectral flux emerged as particularly relevant features in the first phase of the intervention, reflecting the internal structure of the child-therapist exchange. In the later phase, fine-grained spectral characteristics related to stability, variety, and transitions in the interaction became more prominent. These findings may highlight potential targets for future research on prosodic and affective synchrony in autism intervention. Features identified at different time points were also largely consistent with those associated with longitudinal changes in synchrony metrics.

Our results seem to be in line with other studies employing computational approaches for the longitudinal understanding of prosodic signatures of autism and their relationships with developmental outcomes or autism severity (Eni et al., 2025; Godel et al., 2023). Notably, this study also underlines their potential role in autism intervention as mediated by child-therapist interpersonal synchrony.

It is also important to emphasize that while some synchrony metrics, such as higher predictability or stability, may reflect desirable features of structured and coordinated interactions, they could also emerge in the context of rigid or overly deterministic exchanges. In fact, both predictability and stability may reflect adaptive and effective exchanges characterized by shared co-regulation and variability embedded within an organized structure. Alternatively, they may also indicate rigid, stereotyped interactions or overly repetitive and predictable patterns. The main result underscored by the present findings focuses on the need to consider the balance between complementary aspects of synchrony, i.e., including both stability and flexibility, in line with evidence on synchrony complexity (Mayo and Gordon, 2020). Importantly, such patterns may also vary over time, reflecting different phases of therapy, and may be differentially associated with the developmental response, as our results seem to suggest. For example, in early stages of intervention, more predictable interactions may support engagement and joint attention, while as therapy progresses, smoother transitions, shared control, and sustained synchronization may emerge, fostering more reciprocal, flexible, co-regulated exchanges. Capturing these evolving synchrony patterns may help identify clinically meaningful profiles and inform optimal levels of synchrony.

4.1. Implications

The findings of this study provide significant insights into the relationship between acoustic synchrony and therapy response in children with autism, contributing to both clinical practice and research. From a clinical perspective, these results imply that the computational assessment of interpersonal synchrony might provide valuable feedback to clinicians. Our analysis involved acoustic features which are known to participate in the emotional content of speech, pointing out the relevance of affective aspects between the child and the therapist for intervention efficacy, underscored by developmental models of intervention (Schreibman et al., 2015) and developmental studies (Bourvis et al., 2018; Cohen et al., 2013; Feldman, 2012). Although exploratory, this evidence also suggests that different phases of the intervention may require specific adjustments and further investigation to optimize therapeutic paths, in line with clinical experience. In fact, it is common to scaffold clinical interaction with autistic children by first focusing on establishing predictable affective routines, and then enriching them by introducing variations. These dynamics could be reflected in the child-therapist prosodic synchronization, as suggested by evidence in clinical contexts (Lahiri et al., 2022; Bone et al., 2014). These findings suggest that monitoring the balance between stability and flexibility in therapist-child synchrony over time may help clinicians tailor interventions, supporting structured engagement in the early phases, while gradually fostering more flexible, reciprocal, and sustained interactions that promote social-communication development as therapy unfolds.

4.2. Limitations and future research

This study is not without limitations, the main one being the sample size which challenges the power of the analysis. To compensate, we employed a Bayesian continuous approach and interpreted evidence only if at least moderate. However, given the heterogeneity of autism, despite the robust methodology and rigorous pipeline, caution should be used while interpreting these exploratory results, especially with respect to their generalization, that needs to be addressed in further studies. Moreover, our sample included a predominance of male children and female therapists, also challenging generalizability with respect to gender. Further study should extend the exploratory findings discussed in this study. Additionally, this study did not include therapist variables or child baseline characteristics, focusing solely on developmental trajectories while controlling for spontaneous evolution. The correlational nature of this study did not allow the investigation of causal relationships. Even if correlations in the first phase of the intervention suggest the existence of early mechanisms that influence therapy efficacy, further analysis should focus on investigating whether interpersonal synchrony could be a mechanism of action or a marker of successful therapies with a predictive approach. However, such an insight would have a significant clinical relevance in both cases. Another limitation is represented by the unimodal nature of our analysis, which did not include other behavioral information and focused solely on prosodic aspects of language. In the same line, this study aimed at featuring specific aspects of the therapeutic process related to short-term and long-term synchrony, which remains particularly under-investigated. However, further research should address longitudinal synchrony trajectories by analyzing more videos to identify robust synchrony profiles.

4.3. Conclusion

This approach offers novel, data-driven lenses to examine affective synchrony in therapeutic contexts and could significantly enhance clinical research by leveraging extensive automated computation. To our knowledge, no previous research applied automated, advanced computational techniques to full-session acoustic data in naturalistic developmental therapy, marking a methodological advancement in the field and potentially disclosing fine-grained features of therapy dynamics linked to response. Future studies should consolidate this methodology on a larger scale in terms of sample size and longitudinal time points to enable digital phenotyping of affective synchrony and disclose its potential role as an active driver of therapeutic change. Moreover, it will be important to better characterize effective synchrony profiles by integrating child and therapist variables as well as clinical expertise, in order to refine intervention personalization and optimize therapeutic outcomes.

Declarations

Clinical trial number: not applicable.

Ethics approval and consent to participate

This study was approved by the Research Ethics Board of the University of Trento (Protocol number: 2020–042) and complies with the principles laid down in the last version of the Declaration of Helsinki. All participants gave their informed consent to participate in this research.

Consent for publication

All the authors read and approved the manuscript and gave consent for publication. All the material presented in the manuscript is original and does not require other consent for publication.

Availability of data and materials

The aggregated anonymized data can be shared upon reasonable request to the corresponding author. Source data can not be shared due to privacy and ethics limitations for personal biometric and clinical data acquired in clinical contexts, as indicated by the ethics protocol. The source code and trained models are available and will be released in a public repository to be employed by other researchers.

Funding statement

This research was supported by ERA Per Med Joint Transnational Call for Proposals (2021) for Multidisciplinary Research Projects on Personalized Medicine (grant. ID: 779282) - Development of Clinical Support Tools for Personalized Medicine Implementation for the project TECH-TOYS: Acquire digiTal biomarkErs in infanCy witH sensorized TOYS for early detection and monitoring of neuro- developmental disorders (ERAPERMED2021-309).

CRediT authorship contribution statement

Giulio Bertamini: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Silvia Perzolli:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Arianna Bentenuto:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Conceptualization. **Cesare Furlanello:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Mohamed Chetouani:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Conceptualization. **David Cohen:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Conceptualization. **Paola Venuti:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.etdah.2025.100176.

References

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787.
Bertamini, G., Furlanello, C., Chetouani, M., Cohen, D., Venuti, P., 2025. Automated segmentation of child-clinician speech in naturalistic clinical contexts. Res. Dev. Disabil. 157, 104906.
Bishara, A.J., Hittner, J.B., 2012. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. Psychol. Methods 17 (3), 399.
Biswas, M., Sahu, M., Agrebi, M., Singh, P.K., Badr, Y., 2023. Speech emotion recognition using deep CNNs trained on log-frequency spectrograms. Innovations in Machine and Deep Learning: Case Studies and Applications. Springer Nature Switzerland, Cham, pp. 83–108.
Bone, D., Lee, C.C., Black, M.P., Williams, M.E., Lee, S., Levitt, P., Narayanan, S., 2014. The psychologist as an interlocutor in autism spectrum disorder assessment: insights from a study of spontaneous prosody. J. Speech Lang. Hear. Res. 57 (4), 1162–1177.
Bourvis, N., Singer, M., Saint Georges, C., Bodeau, N., Chetouani, M., Cohen, D., Feldman, R., 2018. Pre-linguistic infants employ complex communicative loops to engage mothers in social exchanges and repair interaction ruptures. R. Soc. Open. Sci. 5 (1), 170274.
Butler, E.A., 2011. Temporal interpersonal emotion systems: the "TIES" that form relationships. Pers. Soc. Psychol. Rev. 15 (4), 367–393.
Chen, X., Yang, J., Gan, S., Yang, Y., 2012. The contribution of sound intensity in vocal emotion perception: behavioral and electrophysiological evidence. PLoS One 7 (1), e30278.
Coco, M.I., Dale, R., 2014. Cross-recurrence quantification analysis of categorical and continuous time series: an R package. Front. Psychol. 5, 510.
Cohen, D., Cassel, R.S., Saint-Georges, C., Mahdhaoui, A., Laznik, M.C., Apicella, F., Chetouani, M., 2013. Do parentese prosody and fathers' involvement in interacting facilitate social interaction in infants who later develop autism? PLoS One 8 (5), e61402.
D'Agostino, S.R., Dueñas, A.D., Bravo, A., Tyson, K., Straiton, D., Salvatore, G.L., Pellecchia, M., 2023. Toward deeper understanding and wide-scale implementation of naturalistic developmental behavioral interventions. Autism 27 (1), 253–258.
Dale, R., Warlaumont, A.S., Richardson, D.C., 2011. Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. Int. J. Bifurc. Chaos 21 (04), 1153–1161.
Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., Cohen, D., 2012. Interpersonal synchrony: a survey of evaluation methods across disciplines. IEEE Trans. Affect. Comput. 3 (3), 349–365.
Duong, S., Davis, T.J., Bachman, H.J., Votruba-Drzal, E., Libertus, M.E., 2024. Dynamic structures of parent-child number talk: an application of categorical cross-recurrence quantification analysis and companion to Duong et al.(2024). Quant. Methods Psychol. 20, 137–155.
Eyben, F., Scherer, K.R., Schuller, B.W., Sundberg, J., Andre, E., Busso, C., et al., 2016. The Geneva minimalistic acoustic parameter Set (GeMAPS) for voice research and affective computing. IEEE Trans. Affect. Comput. 7 (2), 190–202. Apr 1.
Eyben, F., Salomão, G.L., Sundberg, J., Scherer, K.R., Schuller, B.W., 2015. Emotion in the singing voice—A deeper look at acoustic features in the light of automatic classification. EURASIP J. Audio Speech Music Process. 2015, 1–9.
Eni, M., Zigel, Y., Ilan, M., Michaelovski, A., Golan, H.M., Meiri, G., Dinstein, I., 2025. Reliably quantifying the severity of social symptoms in children with autism using ASDSpeech. Transl. Psychiatry 15 (1), 14.
Feldman, R., 2007. Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. J. Child Psychol. Psychiatry 48 (3–4), 329–354.
Feldman, R., 2012. Parent-infant synchrony: a biobehavioral model of mutual influences in the formation of affiliative bonds. Monogr. Soc. Res. Child Dev. 77 (2), 42–51.
Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D.M., Gaigg, S.B., 2017. Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. Autism Res.: Off. J. Int. Soc. Autism Res. 10 (3), 384–407. https://doi.org/10.1002/aur.1678.
Fusaroli, R., Tylén, K., 2016. Investigating conversational dynamics: interactive alignment, interpersonal synergy, and collective task performance. Cogn. Sci. 40 (1), 145–171.
Godel, M., Robain, F., Journal, F., Kojovic, N., Latrèche, K., Dehaene-Lambertz, G., Schaer, M., 2023. Prosodic signatures of ASD severity and developmental delay in preschoolers. NPJ Digit. Med. 6 (1), 99.
Guzman, M., Correa, S., Muñoz, D., Mayerhoff, R., 2013. Influence on spectral energy distribution of emotional expression. J. Voice: Off. J. Voice Found. 27 (1), 129. e1–129.e10.
Hubbard, D.J., Faso, D.J., Assmann, P.F., Sasson, N.J., 2017. Production and perception of emotional prosody by adults with autism spectrum disorder. Autism Res. 10 (12), 1991–2001.
Kamiloğlu, R.G., Fischer, A.H., Sauter, D.A., 2020. Good vibrations: a review of vocal expressions of positive emotions. Psychon. Bull. Rev. 27, 237–265.
Klin, A., Micheletti, M., Klaiman, C., Shultz, S., Constantino, J.N., Jones, W., 2020. Affording autism an early brain development re-definition. Dev. Psychopathol. 32 (4), 1175–1189.
Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., Feldman, R., 2017. Brain-to-brain synchrony during naturalistic social interactions. Sci. Rep. 7 (1), 17060.
Klintwall, L., Eldevik, S., Eikeseth, S., 2013. Narrowing the gap: effects of intervention on developmental trajectories in autism. Autism 19 (1), 53–63. Nov 8.
Koole, S.L., Tschacher, W., 2016. Synchrony in psychotherapy: a review and an integrative framework for the Therapeutic Alliance. Front. Psychol. 7 (862). Jun 14.
Kruschke, J. (2014). Doing bayesian data analysis: a tutorial with R, JAGS, and Stan.
Imel, Z.E., Barco, J.S., Brown, H.J., Baucom, B.R., Baer, J.S., Kircher, J.C., et al., 2014. The association of therapist empathy and synchrony in vocally encoded arousal. J. Couns. Psychol. 61 (1), 146–153.
Jobin, A., 2020. Varied treatment response in young children with autism: a relative comparison of structured and naturalistic behavioral approaches. Autism 24 (2), 338–351.
Lahiri, R., Nasir, M., Kumar, M., Kim, S.H., Bishop, S., Lord, C., Narayanan, S., 2022. Interpersonal synchrony across vocal and lexical modalities in interactions involving children with autism spectrum disorder. JASA Express Lett. 2 (9).
Leclère, C., Viaux, S., Avril, M., Achard, C., Chetouani, M., Missonnier, S., Cohen, D., 2014. Why synchrony matters during mother-child interactions: a systematic review. PLoS One 9 (12), e113571.
Lee, M.D., Wagenmakers, E.J., 2014. Bayesian Cognitive modeling: A practical Course. Cambridge university press.
Lord, C., Rutter, M., DiLavore, P.C., Risi, S., Gotham, K., Bishop, S., 2012. Autism Diagnostic Observation Schedule, 2nd edition. Western Psychological Services, Torrance, CA.
Luiz, D., Barnard, A., Knoesen, N., Kotras, N., Horrocks, S., McAlinden, P., et al., 2006. Griffiths Mental Development Scales—Extended Revised: Two to Eight Years: Administration Manual. Hogrefe, Oxford, UK.
Marwan, N., Kurths, J., 2002. Nonlinear analysis of bivariate data with cross recurrence plots. Phys. Lett. A 302 (5–6), 299–307.
Marwan, N., Romano, M.C., Thiel, M., Kurths, J., 2007. Recurrence plots for the analysis of complex systems. Phys. Rep. 438 (5–6), 237–329.
Mayo, O., Gordon, I., 2020. In and out of synchrony-behavioral and physiological dynamics of dyadic interpersonal coordination. Psychophysiology, e13574. Mar 28.
McNaughton, K.A., Redcay, E., 2020. Interpersonal Synchrony in autism. Curr. Psychiatry Rep. 22 (3). Feb 5.
Morey, R.D., Rouder, J.N., Jamil, T., & Morey, M.R.D. (2015, September). *Package 'bayesfactor'*.
Mössler, K., Gold, C., Aßmus, J., Schumacher, K., Calvet, C., Reimer, S., et al., 2017. The therapeutic relationship as predictor of change in music therapy with young children with Autism spectrum disorder. J. Autism. Dev. Disord. 49 (7), 2795–2809. Sep 21.
Nelson C.A., Sullivan E., Engelstad A.M. Annual Research Review: early intervention viewed through the lens of developmental neuroscience. 2023 Jul 12.
Nof, A., Amir, O., Goldstein, P., Zilcha-Mano, S., 2020. What do these sounds tell us about the therapeutic alliance: acoustic markers as predictors of alliance. Clin. Psychol. Psychother. 28 (4), 807–817. Dec 27.
Oller, D.K., Buder, E.H., Ramsdell, H.L., Warlaumont, A.S., Chorna, L., Bakeman, R., 2013. Functional flexibility of infant vocalization and the emergence of language. Proc. Natl. Acad. Sci. U.S.A 110 (16), 6318–6323. https://doi.org/10.1073/pnas.1300337110.
Patel, S., Scherer, K.R., Björkner, E., Sundberg, J., 2011. Mapping emotions into acoustic space: the role of voice production. Biol. Psychol. 87 (1), 93–98.
Patel, S., Scherer, K.R., Sundberg, J., Björkner, E., 2010. Acoustic markers of emotions based on voice physiology. In: Proceedings of the speech prosody, 100865, pp. 1–4.
Pelphrey, K.A., Shultz, S., Hudac, C.M., Vander Wyk, B.C, 2011. Research Review: constraining heterogeneity: the social brain and its development in autism spectrum disorder. J. Child Psychol. Psychiatry 52 (6), 631–644. Jan 19.
Rawald, T., Sips, M., Marwan, N., 2017. PyRQA—Conducting recurrence quantification analysis on very long time series efficiently. Comput. Geosci. 104, 101–108. Jul 1.
Sandbank, M., Bottema-Beutel, K., LaPoint, S.C., Feldman, J.I., Barrett, D.J., Caldwell, N., Woynaroski, T., 2023. Autism intervention meta-analysis of early childhood studies (Project AIM): updated systematic review and secondary analysis. BMJ 383.

Scherer, K.R., 2013. Vocal markers of emotion: comparing induction and acting elicitation. Comput. Speech. Lang. 27 (1), 40–58.
Schoenherr, D., Strauss, B., Stangier, U., Altmann, U., 2021. The influence of vocal synchrony on outcome and attachment anxiety/avoidance in treatments of social anxiety disorder. Psychotherapy 58 (4), 510–522. Dec.
Schreibman, L., Dawson, G., Stahmer, A.C., Landa, R., Rogers, S.J., McGee, G.G., Halladay, A., 2015. Naturalistic developmental behavioral interventions: empirically validated treatments for autism spectrum disorder. J. Autism. Dev. Disord. 45, 2411–2428.
Shockley, K. (2005). Cross recurrence quantification of interpersonal postural activity. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, 142–177.
Shultz, S., Klin, A., Jones, W., 2018. Neonatal transitions in social behavior and their implications for autism. Trends Cogn. Sci. (Regul. Ed.) 22 (5), 452–469.
Song, J., Reilly, M., Reichow, B., 2025. Overview of meta-analyses on naturalistic developmental behavioral interventions for children with autism spectrum disorder. J. Autism. Dev. Disord. 55 (1), 1–13.
Sundberg, J., Patel, S., Bjorkner, E., Scherer, K.R., 2011. Interdependencies among voice source parameters in emotional speech. IEEe Trans. Affect. Comput. 2 (3), 162–174.
Varni, G., Hupont, I., Clavel, C., Chetouani, M., 2017. Computational study of primitive emotional contagion in dyadic interactions. IEEe Trans. Affect. Comput. 11 (2), 258–271.
Ververidis, D., Kotropoulos, C., 2006. Emotional speech recognition: resources, features, and methods. Speech. Commun. 48 (9), 1162–1181.
Vivanti, G., Prior, M., Williams, K., Dissanayake, C., 2014. Predictors of outcomes in autism early intervention: why don't we know more? Front. Pediatr. 2, 58.
Warlaumont, A.S., Oller, D.K., Dale, R., Richards, J.A., Gilkerson, J., Xu, D., 2010. Vocal interaction dynamics of children with and without autism. In: Proceedings of the Annual Meeting of the Cognitive Science Society, 32.
Warlaumont, A.S., Richards, J.A., Gilkerson, J., Oller, D.K., 2014. A social feedback loop for speech development and its reduction in autism. Psychol. Sci. 25 (7), 1314–1324. https://doi.org/10.1177/0956797614531023.
Webber, C.L., Marwan, N. (Eds.), 2015. Recurrence Quantification Analysis. Understanding Complex Systems. Springer International Publishing, Cham.
Webber Jr, C.L., Zbilut, J.P., 2005. Recurrence quantification analysis of nonlinear dynamical systems. Tutor. Contemp. Nonlinear Methods Behav. Sci. 94 (2005), 26–94.
Weisman, O., Chetouani, M., Saint-Georges, C., Bourvis, N., Delaherche, E., Zagoory-Sharon, O., Feldman, R., 2015. Dynamics of non-verbal vocalizations and hormones during father-infant interaction. IEEE Trans. Affect. Comput. 7 (4), 337–345.
Weninger, F., Eyben, F., Schuller, B.W., Mortillaro, M., Scherer, K.R., 2013. On the acoustics of emotion in audio: what speech, music, and sound have in common. Front. Psychol. 4, 292.
World Medical Association, 2024. World medical association declaration of Helsinki. JAMA 333 (1). Oct 19.