Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Symbolic State-Space Framework for Quantifying the Dynamical Structure of Multimodal Conversational Interaction

Mark Wünsche

m.wuensche@fz-juelich.de

Forschungszentrum Jülich

Malin Spaniol

University Hospital Cologne

Mathis Jording

University Hospital Cologne

Martine Grice

University of Cologne

Kai Vogeley

University Hospital Cologne

Research Article

Keywords: Interaction Dynamics, Multimodal Conversation, Recurrence Quantification Analysis, Leave-One-Covariate-Out, Symbolic Dynamics, Face-to-Face Interaction

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11245090/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Symbolic State-Space Framework for Quantifying the Dynamical Structure of Multimodal Conversational Interaction

Mark Wünsche[1,2], Malin Spaniol[2], Mathis Jording[2], Martine Grice[3], Kai Vogeley[2,1]

Corresponding author: Mark Wünsche (m.wuensche@fz-juelich.de)

1 Research Centre Jülich, Institute of Neuroscience and Medicine, Cognitive Neuroscience (INM-3), Jülich, Germany

2 Department of Psychiatry, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

3 IfL-Phonetics, University of Cologne, Cologne, Germany

Abstract

Human conversation depends on the sequential coordination of multiple communicative channels. We present a general methodological framework for quantifying the dynamical organisation of multimodal conversational behaviour. Conversational behaviour is represented as trajectories through a discrete, symbolic state space that combines speech, gaze, backchannels, and laughter from both interlocutors. We apply a set of dynamical and information-theoretic analyses, namely recurrence quantification analysis (RQA), cross-recurrence analysis, information-theoretic measures (block entropies, entropy rate, excess entropy), Markov transition summaries, and a leave-one-covariate-out (LOCO) procedure for quantifying modality-specific contributions to conversational structure. As a proof-of-concept application, using a corpus of dyadic conversations recorded across three conversational contexts (introduction, a collaborative tangram task, and open discussion), we show that conversational trajectories exhibit clear recurrence structure alongside substantial local variability. Structure was systematically modulated by task demands: the most constrained collaborative task produced the strongest recurrence and the lowest entropy. Modality-specific analyses further showed that speech and gaze were the dominant structure-contributing channels, whereas backchannels and laughter contributed only weakly to global dynamical organisation. We describe the encoding, metrics, and statistical procedures so that the framework can be applied to other multimodal interaction datasets.

Keywords: Interaction Dynamics; Multimodal Conversation; Recurrence Quantification Analysis; Leave-One-Covariate-Out; Symbolic Dynamics; Face-to-Face Interaction

1 Introduction

Human face-to-face conversation is a highly coordinated multimodal process involving speech, gaze, gestures, backchannels, and facial expressions. The framework introduced below is developed and illustrated for multimodal conversation specifically, but the same representational logic applies in principle to other forms of face-to-face social interaction. Through the continuous exchange of these signals, interlocutors regulate turn-taking, signal understanding, information exchange, maintenance of mutual attention and, importantly, the success of communication, also referred to as "rapport" (e.g., Wünsche et al. 2026). Despite its apparent spontaneity, conversational interaction follows organisational principles that unfold in time during the dyadic interaction. Patterns of speaking and listening, gaze shifts including mutual gaze, and feedback signals form structured sequences that allow participants to maintain mutual understanding and coordinate the flow of their communication (Hjalmarsson et al., 2012; Skantze et al., 2014).

Understanding the structuredness of conversation is important for several reasons. First, conversational structure is the substrate upon which communicative success depends: the orderly sequencing of communicative signals enables interlocutors to maintain mutual understanding, coordinate turn transitions, and achieve shared goals (Marsh et al., 2009). Second, the degree of structuredness in interaction is itself a marker of the communicative demands placed on interlocutors: different tasks and contexts may elicit different levels of coordination, and quantifying these differences requires tools that can capture both the global organisation and the local variability of conversational behaviour. Finally, from a theoretical standpoint, the tension between conversational orderliness and flexibility speaks to long-standing debates about whether conversation is best understood as a rule-governed system, a self-organising dynamical process, or something in between (Thórisson, 2008; Kelso, 2012). Resolving this tension requires methods that can simultaneously quantify structure and variability in the same interaction.

Classical theories of conversational coordination have often conceptualised conversations simply as a structured alternation between speaker and listener roles. The foundational model of turn-taking proposed by Sacks, Schegloff, and Jefferson (1974) describes conversation as a locally managed system in which participants coordinate the allocation of turns through pragmatic cues and the projection of upcoming turn-completion points. This framework has been remarkably influential, and subsequent work has shown that its core principles —

minimal gap and overlap, one-party-at-a-time — hold with striking cross-cultural uniformity across diverse languages and communities (Stivers et al., 2009). Turn-taking models describe how conversational partners coordinate speech production and comprehension through temporal cues, pragmatic signalling, and prediction of upcoming turns (Wilson & Wilson, 2005; Levinson & Torreira, 2015). Approaches focusing on the content of conversation emphasise that interlocutors achieve mutual understanding through interactive alignment of linguistic representations at multiple levels, e.g. semantical, syntactical, lexical (Pickering & Garrod, 2004) and through the incremental establishment of common ground (Clark, 1996). Other approaches emphasise inter- and intra-personal synchronisation processes (Bloch et al., 2019; Georgescu et al., 2020; Plank et al., 2023), suggesting that communicative behaviour follows oscillatory dynamics analogous to coupled biological rhythms (Wilson & Wilson, 2005). While such frameworks successfully capture aspects of conversational rhythm, they cannot on their own account for the dynamic interplay between the fundamental sequential organisation of conversations and their substantial moment-to-moment variability.

In everyday interactions, speakers frequently deviate from canonical turn-taking structures which were predicted from highly structured research paradigms leading to artificial, structured or even partially scripted speech (Sacks et al., 1974; Kendrick et al., 2023) and instead show highly variable dynamics of e.g. speech overlaps, backchannels, and gaze shifts, with gaze, manual gesture, facial expression, and body posture all contributing to turn coordination alongside the speech signal (Goodwin, 2000; Kendrick, Holler, & Levinson, 2023; Trujillo & Holler, 2023). These local fluctuations suggest unpredictability, yet conversations rarely devolve into disorder; rather, as we, too, show below, they evolve within a constrained set of behavioural configurations that recur over time.

One possible explanation for this tension between structuredness and variability is that conversation exhibits metastable-like organisation. In metastable systems, trajectories remain globally organised while exhibiting locally unpredictable fluctuations at the same time (Brunton et al., 2017; Marwan et al., 2007). Rather than following rigid periodic patterns, such systems occupy a structured state space with local, "surface" variability. This type of dynamical organisation has been observed across biological and physical systems, including complex communication networks and neural coordination (Tognoli & Kelso, 2014; Kelso, 2012), and the concept of metastability has been proposed as a general principle of brain–behaviour coordination, capturing the coexistence of integrative and segregative tendencies (channels or subsystems at times behaving independently, and at other times coordinating as a single coherent whole) within the same system (Kelso, Dumas, & Tognoli, 2013; Engstrøm & Kelso, 2008).

Applying this idea to human communication requires representing conversational behaviour as a dynamical system evolving in a shared multimodal state space. Instead of analysing individual communicative signals in isolation, interaction can be represented as trajectories through combinations of behavioural signals produced by two or more interlocutors. Each moment of interaction corresponds to a discrete configuration of communicative activity across modalities and partners. Conversational behaviour is here encoded as k binary channels for each of n participants, yielding a symbolic state space with n^k possible interaction states (Figure 1). This state-space view is consistent with recent methodological work showing that dyadic interaction dynamics can be formally simulated, visualised, and interpreted as evolving coupled systems rather than as independent behavioural streams (Berkhout et al., 2025), and with approaches that treat multimodal alignment as a temporally structured, multilevel process (Rasenberg, Özyürek, & Dingemanse, 2020). More details on the exact formulation can be found in Section 2.2.

This representation enables the application of analytical tools from nonlinear dynamics and information theory. Recurrence analysis provides a method for detecting repeated state configurations in complex systems and has been widely used to identify hidden structure in noisy time series (Marwan et al., 2007; Zbilut et al., 1998). Information-theoretic measures such as entropy rate quantify the degree of unpredictability in system dynamics (Crutchfield & Feldman, 2003; Crutchfield & Young, 1989), while Markov-based models capture the transition structure between states. Together, these approaches allow quantifying how strongly conversational trajectories are organised and which communicative signals contribute to that organisation.

Existing recurrence-based tools for interpersonal coordination are informative but limited in two respects. First, cross-recurrence quantification analysis (CRQA) as typically applied — for example via the widely used crqa R package (Coco & Dale, 2014), validated against the benchmark MATLAB crptoolbox (Marwan et al., 2007) — compares two separate time series at a time, most often a single behavioural channel recorded from each interlocutor, such as eye-fixation sequences or continuous body-movement intensity (Fusaroli et al., 2014), rather than several communicative channels from both interlocutors jointly. Second, both categorical and continuous CRQA require a delay and/or radius parameter to be estimated for each dataset (the CTcrqa and optimizeParam functions of crqa; Fraser & Swinney, 1986; Marwan et al., 2007): an unsuitable choice can systematically over- or under-estimate recurrence, making results difficult to compare across studies without careful, dataset-specific manual tuning. The framework introduced below addresses both limitations by representing multiple communicative channels from both interlocutors jointly as a single combinatorial state, dispensing with delay-

embedding altogether by defining recurrence directly on exact state identity, and pairing this representation with a channel-ablation procedure that quantifies which channels drive the observed structure (Section 2.9).

The present paper addresses three methodological questions:

- How can multimodal conversational behaviour be formalised as trajectories through a discrete, symbolic state space?
- Which tools from nonlinear dynamics and information theory can quantify the resulting structure, and how can the contribution of individual communicative channels to that structure be isolated?
- Do conversational contexts that differ in their communicative demands produce systematically different degrees of structuredness, and can this framework detect such differences?

To address these questions, multimodal behavioural streams from dyadic conversations were encoded as symbolic state sequences and analysed using recurrence quantification analysis, cross-recurrence analysis, information-theoretic measures, Markov transition summaries, and a leave-one-covariate-out (LOCO) procedure.

As an example and proof of concept of this approach we analysed a corpus of conversations between two interlocutors with channels for speech, gaze, backchannels, and laughter. These four channels were chosen because they jointly index vocal, visual, and affective dimensions of conversational coordination and are widely used markers of engagement and coordination in the interaction literature. In principle, the same representational logic extends to any set of behavioural channels that can be time-aligned and discretised.

The results show that conversational interaction combines strong sequential structure with persistent local variability, and that this structure is systematically modulated by conversational context and by the communicative channels available to the system — supporting the view that human conversation operates as a metastable dynamical system. The corpus used to illustrate the framework includes dyads that varied in clinical composition (autistic-autistic and non-autistic-non-autistic dyads). Diagnostic group is not a focus of the present analyses, therefore, analyses are based on the data of both groups and leave out group comparisons. A full between-group comparison will follow in separate work (Wünsche et al. in prep.). For transparency, a purely descriptive task comparison is provided, but therefore only based on the non-autistic subgroup.

2 Methods

2.1 Data and conversational tasks

The framework requires behavioural data that are time-aligned across interlocutors and encoded as discrete channels (e.g., binary presence/absence indicators) sampled at a fixed temporal resolution; any dataset meeting this format can be substituted for the example used below. Here, behavioural data are derived from a corpus of dyadic face-to-face conversations (Spaniol et al., 2024; Spaniol, 2025). 20 German-speaking dyads (8 autistic-autistic and 12 non-autistic-non-autistic dyads), constituting approximately 9.7 hours of total conversation time, with available annotation are included in the analyses. Participants' mean age was 37.0 years (SD = 12.7; range 22-60), and 13 dyads were matched for gender. Each interaction lasted for 30 minutes and involved two participants whose communicative behaviour was automatically annotated (gaze), semi-automatically annotated (speech, via automatic voice-activity detection with manual verification and correction), and manually annotated (backchannel, laughter) from audio and video recordings, with speech captured via microphones and gaze behaviour via eye-tracking glasses. Backchannels were coded by trained annotators from a closed set of vocal feedback tokens (e.g., ja, okay, mmhm, mm, genau, multi-unit responses, or other; see Spaniol et al., 2024, for the full coding scheme), and laughter was coded as any audible laughter regardless of its communicative function. Dyads were recorded across three conversational contexts designed to elicit different forms of communicative coordination:

- introduction (i): informal small talk allowing participants to become acquainted;
- tangram task (t): a collaborative problem-solving task requiring participants to jointly reconstruct tangram figures through verbal description and visual coordination;
- discussion (d): an open conversation reflecting on the preceding task.

These contexts differ systematically in their cognitive and communicative demands. The introduction phase emphasises spontaneous social interaction, the tangram task requires structured collaborative coordination, and the discussion phase allows for more flexible conversational dynamics. More detail on the instructions and stimulus materials can be found in Spaniol (2025, dissertation) or at OSF: osf.io/z5crx. This variation in task structure is exploited below as a validation case for the framework: if the method is sensitive to genuine differences in interactional structure, task should systematically modulate the recovered dynamical summaries.

2.2 Multimodal behavioural encoding

Conversational behaviour was represented in binary behavioural channels, indicating for each interlocutor the presence or absence of the respective signals (here: gaze directed at partner's face, speech production, backchannels, and laughter). At each time point, the interaction state is represented as a vector (here $x_t = (g_1, s_1, b_1, l_1, g_2, s_2, b_2, l_2)$, where g, s, b, and l denote gaze, speech, backchannels, and laughter, respectively, for participants 1 and 2). This representation captures the joint multimodal structure of conversational behaviour (channel identities, their number, and their order within the vector are chosen by the user to match the behaviours of interest and the available data) and allows interactions to be analysed as trajectories within a discrete symbolic state space. The toolbox treats all channels equally without consideration of theoretical implications. In the example used here, however, the four channels differ in their theoretical status, which matters for interpreting results even though it plays no role in the computations themselves. Speech and gaze are regarded here as the most fundamental channels: speech constitutes the core substrate of conversational turn-taking, and gaze regulates joint attention throughout the interaction, so both fluctuate frequently between presence and absence over the course of an interaction. Backchannels and laughter, by contrast, are more narrowly defined categories assumed to index specific coordinative and affective functions (e.g., signalling continued attention or affiliative stance) and occur comparatively sparsely, with correspondingly fewer transitions between presence and absence.

2.3 Temporal discretisation

To reduce high-frequency noise while preserving temporal ordering, behavioural streams can be aggregated into fixed-width temporal bins. Within each bin, a channel is coded as positive if the corresponding behaviour occurs at least once within the bin (a logical "any" operator). This procedure reduces measurement noise and computational complexity while preserving the sequential structure of the interaction. As a general guideline, bin width should be set relative to the channel with the coarsest native temporal resolution or the highest measurement noise; analysing at a finer resolution than that channel affords adds no information. In the example dataset analysed here, raw behavioural annotations were recorded at approximately 5 ms temporal resolution and were aggregated into bins of 25 ms.

2.4 Symbolic state-space representation

Each binary state vector was encoded as a symbolic state identifier by interpreting the binary variables as a bit vector (implemented by summing each channel weighted by successive powers of two). Because each channel is binary, the symbolic state space contains n^k (in our example $2^8 = 256$) possible multimodal interaction states.

Conversational behaviour can therefore be conceptualised as a trajectory x^1, x^2, x^3, …, x_t through a discrete symbolic state space, where each x_t represents one of n^k possible multimodal configurations of communicative signals at time t. Figure 1 illustrates this symbolic representation and the resulting interaction trajectories through state space.

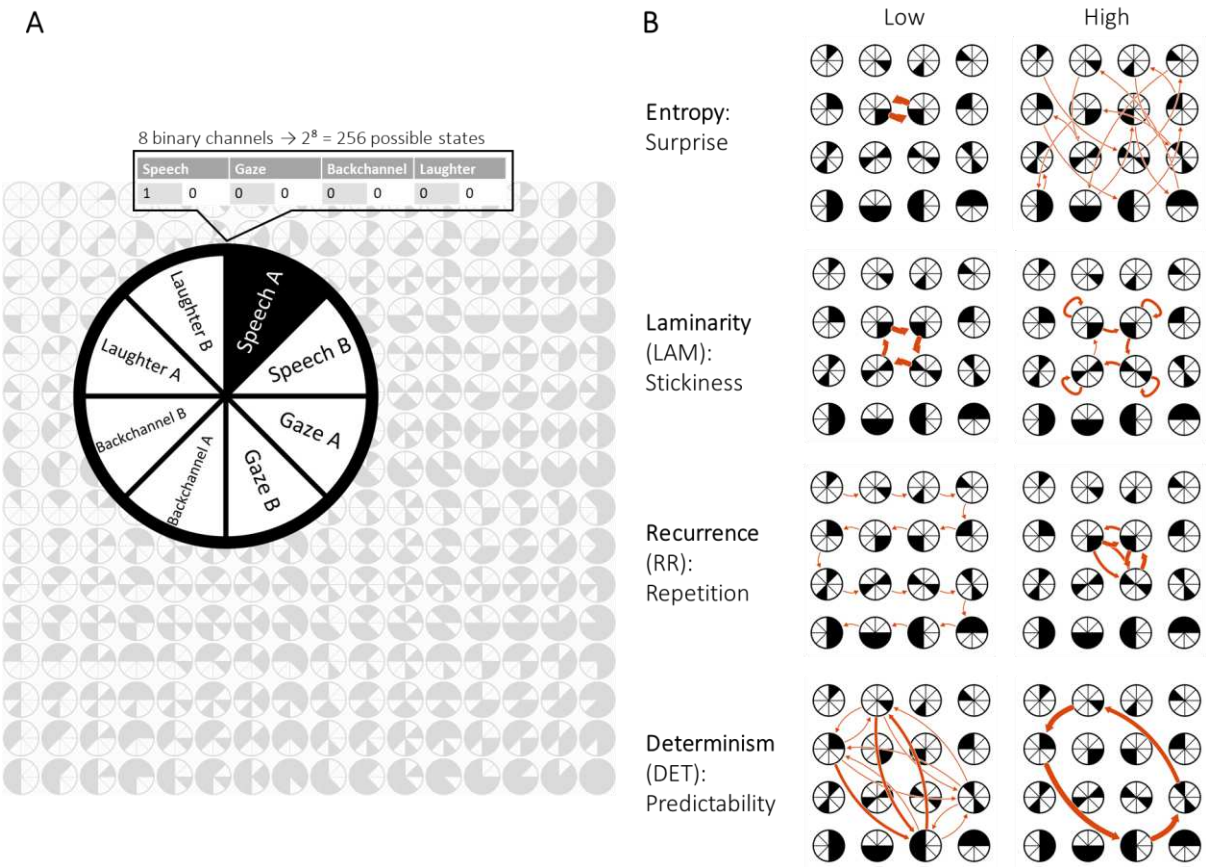


Figure 1. Symbolic state-space representation of multimodal conversation.

A: Illustration of the formalising of the eight binary channels (speech, gaze, backchannel and laughter for each of the two interlocutors A and B) and the resulting 256 (2^8) possible states (depicted in the background). Each state is drawn as a circle divided into eight wedges representing binary channels (black for 1 and white for 0). Depicted is one example state where only Speech of interlocutor A is active. B: Visualisation of the quantification of system structure. For illustration, 16 arbitrarily chosen states out of the 256 are arranged as a 4 × 4 matrix. For each of the four measures (Entropy, Laminarity, Recurrence, Determinism) an example trajectory (orange arrows) with a low (left column) and a high (right column) value is shown.. A thicker arrow marks a transition that was repeated, and a circular arrow marks a state in which the trajectory stayed for at least two consecutive time steps. . Entropy is the measure of unpredictability: the low example alternates between two states (H = 1.00 bits), the high example visits all 16 states in irregular order (H = 4.00 bits). Laminarity describes how long the dyad remains in a specific state: in the low example the trajectory changes state at every step (LAM = 0.00), in the high example it stays in every state before moving on (LAM = 1.00). Recurrence rate assess how often the conversation revisits a specific state: no revisitations of states in the low example (RR = 0.00), revisitations of the same three states in the high example (RR = 0.29). Determinism measures order: it captures whether repetitions in the system follow the same path or not. In the low example, the same four states are visited but in different sequences (DET = 0.20), in the high example the four states are always repeated in the same sequence (DET = 1.00). All values stated here for the different examples were computed solely for illustration purposes and based on 16 time steps in the drawn sequence.

2.5 Recurrence analysis of conversational dynamics

To characterise the dynamical organisation of conversational trajectories, recurrence analysis was applied. Recurrence plots provide a visual representation of when a dynamical system revisits previously observed states over time (Marwan et al., 2007). For a trajectory of symbolic states x_t, recurrence between time points i and j is defined as $R_{i,j} = 1$ if $d(x_i, x_j) \leq \varepsilon$, and 0 otherwise, where $d(\cdot)$ denotes the Hamming distance between binary state vectors and ε is the recurrence threshold. In the present analysis, recurrence was defined using exact state identity, $\varepsilon = 0$, meaning that recurrence occurs only when two interaction states are identical (min line length for diagonal/vertical structures = 2).

2.6 Recurrence quantification analysis (RQA)

To quantify the structure of recurrence patterns, standard RQA metrics were computed (Marwan et al., 2007; Zbilut et al., 1998):

- Recurrence rate (RR): the proportion of state pairs that are recurrent, $RR = (1/N^2) \Sigma_i \Sigma_j R_{i,j}$.
- Determinism (DET): the fraction of recurrence points forming diagonal line structures, reflecting repeated sequential patterns.
- Laminarity (LAM): the fraction of recurrence points forming vertical line structures, indicating persistence of interaction states.
- Trapping time (TT): the average duration of laminar episodes, reflecting how long the system remains within a given interaction state.

Together, these measures quantify complementary aspects of sequential organisation in conversational trajectories, computed once per dyad and task segment (as are the cross-recurrence, information-theoretic, Markov, and LOCO measures in Sections 2.7-2.9); these dyad × task values are the units entering the statistical models in Section 2.10, computed here on the full eight-channel joint dyadic state vector (labelled throughout).

2.7 Cross-recurrence analysis of interlocutor-specific trajectories

To characterise interpersonal coupling between interaction partners, cross-recurrence analysis (CRQA) was applied to interlocutor-specific behavioural state trajectories. Cross-recurrence methods quantify how two behavioural trajectories revisit similar states over time, and are useful for analysing interpersonal coordination in social interaction (Fusaroli et al., 2014; Coco & Dale, 2014; Wallot & Leonardi, 2018). Whereas the RQA metrics above were computed on the joint dyadic state vector, cross-recurrence analysis compared the state trajectory of

partner 1 with the state trajectory of partner 2. To this end, a state trajectory was constructed separately for each interlocutor's behavioural stream (in the example: speech, gaze to partner, backchannels, and laughter). As in Section 2.5, cross-recurrence between time points i and j was defined via the Hamming distance, but now computed between the two interlocutors' own state vectors at those time points – $y^{(1)}_i$ for interlocutor 1 and $y^{(2)}_j$ for interlocutor 2 – again with $\varepsilon = 0$.

2.8 Information-theoretic and Markov summaries

Beyond recurrence-based metrics, two complementary families of dynamical summaries were computed:

- Information-theoretic measures: exact block entropies were computed for block lengths up to $k = 6$, from which an entropy-rate estimate and an excess entropy (a measure of long-range predictive structure) were derived, alongside a Lempel–Ziv complexity estimate for the binarised sequence (Lempel & Ziv, 1976).
- Markov transition summaries: first-order transition matrices were estimated from the symbolic state sequence, from which an entropy rate was derived following formal treatments of entropy rate for discrete Markov-type processes (Smart et al., 2024), alongside a spectral-gap proxy, a mixing-time proxy, and the dominant (top-k) transitions. Equivalently, the symbolic state space and its observed transitions define a directed, weighted transition network, in which the n^k states form nodes and the estimated transition probabilities form edge weights; the measures above are standard descriptors of such a network's structure.

These measures characterise complementary aspects of the dynamics: information-theoretic measures quantify how much long-range structure and unpredictability the sequence contains overall, while Markov summaries characterise the local transition structure between adjacent states.

2.9 Modality contributions: leave-one-covariate-out (LOCO) analysis

To quantify how strongly each communicative channel contributes to the overall dynamical organisation of the interaction, a leave-one-covariate-out (LOCO) procedure was implemented, a permutation-based channel-ablation approach similar in spirit to variable-importance procedures (e.g., permutation importance; Breiman, 2001). For each dyad, task, and communicative channel (in our example speech, gaze, backchannels, laughter), the channel was removed from the multimodal encoding and all RQA, cross-recurrence, information-theoretic, and Markov

metrics were recomputed on the reduced ("drop") representation. The LOCO delta for a given metric was defined as the difference between the value obtained from the full model and the value obtained from the reduced model (full − drop): a large delta in either direction indicates that the channel shapes the dynamical organisation of the trajectory, whereas a delta near zero indicates that it contributes little. Because removing a channel reduces the number of distinguishable states, the reduced representation typically shows higher recurrence and lower entropy; a negative delta for recurrence-based metrics (and a positive delta for entropy-based metrics) therefore means that including the channel lowers recurrence and raises entropy, i.e. that the channel adds distinguishable states and sequential variety. Column permutation tests (permuting the dropped channel in time, independently per dyad × task) were used to assess whether the observed delta exceeded what would be expected from an unstructured channel with the same marginal statistics. This separates the contribution of a channel's temporal organisation from that of its overall activity, which differs strongly between frequently changing channels (here speech and gaze) and sparse ones (here backchannels and laughter). Following the same procedure, the LOCO delta can also be tested for a specific channel from individual interlocutors in order to estimate to what extend one interlocutor's behaviour in this channel structures the conversation.

2.10 Statistical analysis of the example data

Note, the toolbox presented here only outputs metrics for conversational organisation but does not compare or statistically analyse them as statistical analyses are highly dependent on the respective scientific questions and hypotheses. The following description of the statistical procedures are therefore meant as a showcase. Task effects on each dynamical summary were assessed with ordinary least squares (OLS) models of the form metric ~ C(task) × C(diagnosis), i.e. task, diagnostic group, and their interaction (diagnostic group enters as a covariate but is not analysed here, see Section 1), fitted per metric across dyads (bounded proportions RR, DET, and LAM were logit-transformed prior to modelling; trapping time and entropy-rate metrics were modelled on their raw scale). Omnibus term tests for the task factor were obtained via Wald tests on the fitted model, and heteroscedasticity-robust (HC3) coefficient estimates were additionally computed to characterise the direction and magnitude of effects. Benjamini–Hochberg false-discovery-rate (FDR) correction was applied across metrics within each omnibus term. In addition, pairwise contrasts between each pair of tasks were obtained from the same HC3-robust model (two as treatment-coded coefficients against a reference task, the third via a linear contrast of those coefficients; all evaluated at the reference diagnostic group), with Benjamini–Hochberg FDR correction applied

within each metric across its three pairwise contrasts. Descriptive statistics (counts, means, medians) were tabulated by task, and boxplots and descriptive mean ± 95% CI plots were produced for each dynamical summary.

3 Results

3.1 Global structure of multimodal conversational dynamics

Dyadic interactions were represented as trajectories through an eight-dimensional symbolic state space encoding gaze, speech, backchannels, and laughter for both participants. Across the corpus, recurrence quantification analysis (RQA) features averaged across the three conversational tasks (reported by task in Section 3.2) showed clear evidence of repeated revisitation of previously occupied interaction states: mean recurrence rate was 0.154, mean determinism was in the low-to-moderate range (0.391) – well below the DET > 0.80 range conventionally taken to indicate near-periodic dynamics (see Figure 2), and mean laminarity was comparatively high (0.573), indicating that conversational trajectories not only revisited states but also maintained them for extended intervals (mean trapping time = 2.852 bins). The same held for the cross-recurrence (CRQA) metrics computed on interlocutor-specific trajectories (Section 2.7): mean cross-recurrence rate was 0.318, mean cross-determinism was 0.621, mean cross-laminarity was 0.763, and mean cross-trapping time was 4.229 bins. Recurrence plots showed pronounced diagonal structures, indicating repeated sequential motifs, alongside vertical structures reflecting temporarily persistent states, confirming that conversational behaviour was not randomly distributed across the 256-state space but repeatedly revisited constrained regions of it.

The distribution of determinism across all dyad-task segments, shown in Figure 2, illustrates this pattern directly: while a majority of segments showed low-to-moderate determinism, a subset of segments approached much higher levels of sequential order, consistent with a system that combines a dominant flexible regime with occasional strongly ordered episodes rather than a single homogeneous level of structure.

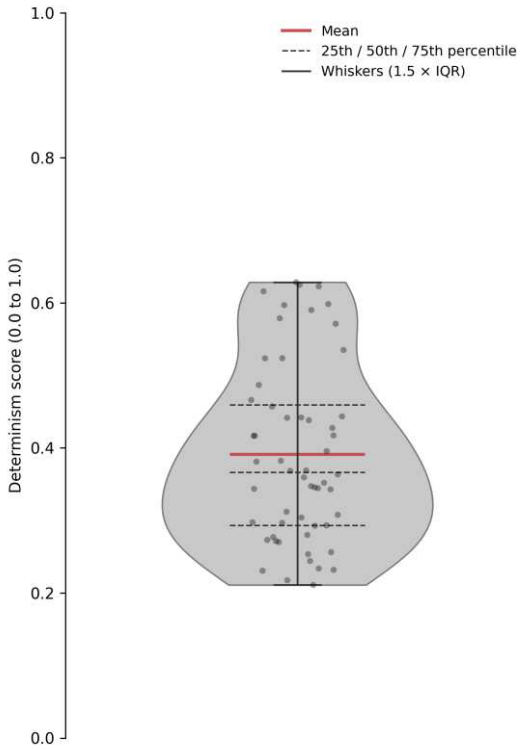


Figure 2. Distribution of determinism (DET) across all dyad tasks, shown as a violin plot. One datapoint corresponds to one task recording for one dyad, overlaid as points; the solid line marks the mean, the dashed lines mark the 25th, 50th, and 75th percentiles, and the whiskers extend to the most extreme values within 1.5 × the interquartile range.

3.2 Task effects on recurrence, determinism, laminarity, and entropy

Recurrence-based structure was present in every dyad-task segment analysed (recurrence rate and determinism were greater than zero throughout), but its magnitude varied systematically with conversational context. For the metrics derived from the joint dyadic state trajectory, the tangram task showed the strongest structure: mean recurrence rate increased from 0.123 in the introduction and 0.119 in the discussion to 0.224 in the tangram task. Mean determinism rose from 0.339 (introduction) and 0.336 (discussion) to 0.505 (tangram). Mean laminarity increased from 0.518 (introduction) and 0.528 (discussion) to 0.677 (tangram), and mean trapping time rose from 2.726 (introduction) and 2.764 (discussion) to 3.078 (tangram).

The same pattern appeared for the cross-recurrence metrics: mean recurrence rate was 0.261 (introduction), 0.271 (discussion), and 0.429 (tangram); determinism increased from 0.561 (introduction) and 0.573 (discussion) to 0.736 (tangram); laminarity rose from 0.718 (introduction) and 0.730 (discussion) to 0.847 (tangram); and trapping time increased from 3.728 (introduction) and 3.910 (discussion) to 5.105 (tangram).

At the same time, entropy rate was lowest in the tangram task, as shown in Figure 3: mean entropy rate was 2.641 (introduction), 2.684 (discussion), and 2.172 (tangram).

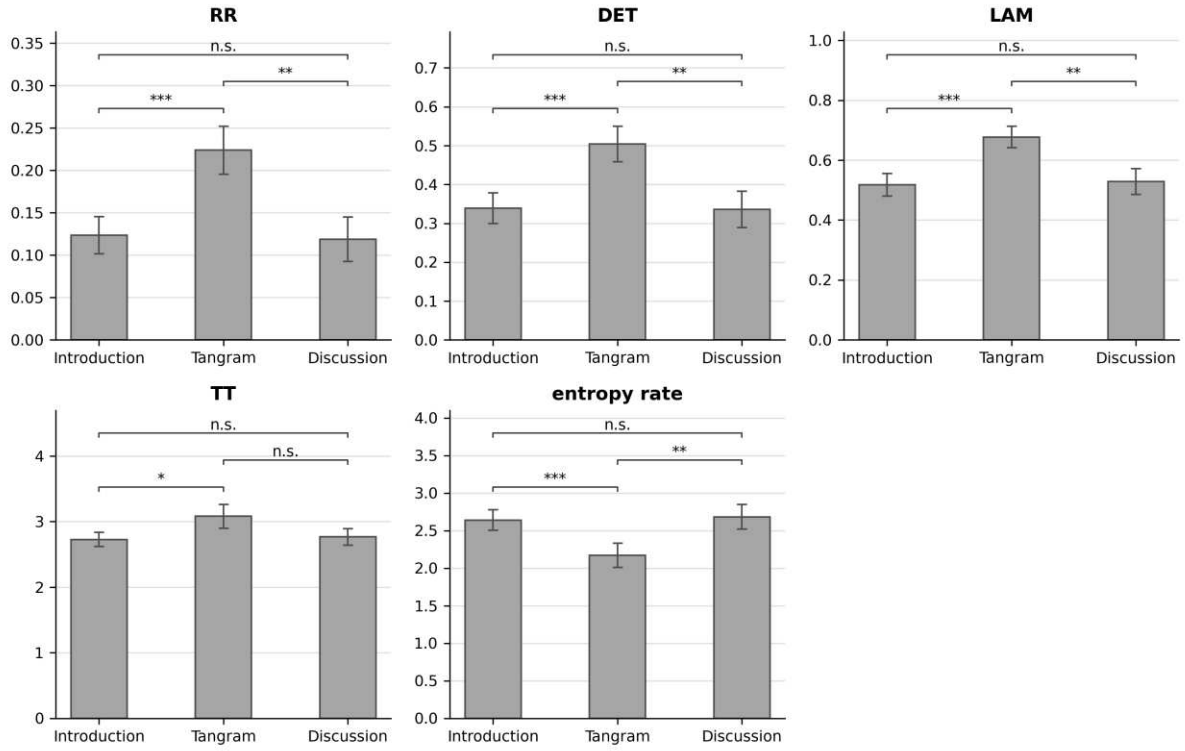


*Figure 3. Recurrence, laminarity, trapping-time, and entropy-rate metrics by conversational task. Bars show the mean (with 95% CI) of each metric across dyad-task segments; RR = recurrence rate; DET = determinism; LAM = laminarity; TT = trapping time. Brackets mark pairwise contrasts between tasks (HC3-robust, Benjamini-Hochberg FDR-corrected within each metric; *** q < .001, ** q < .01, * q < .05, n.s. = not significant).*

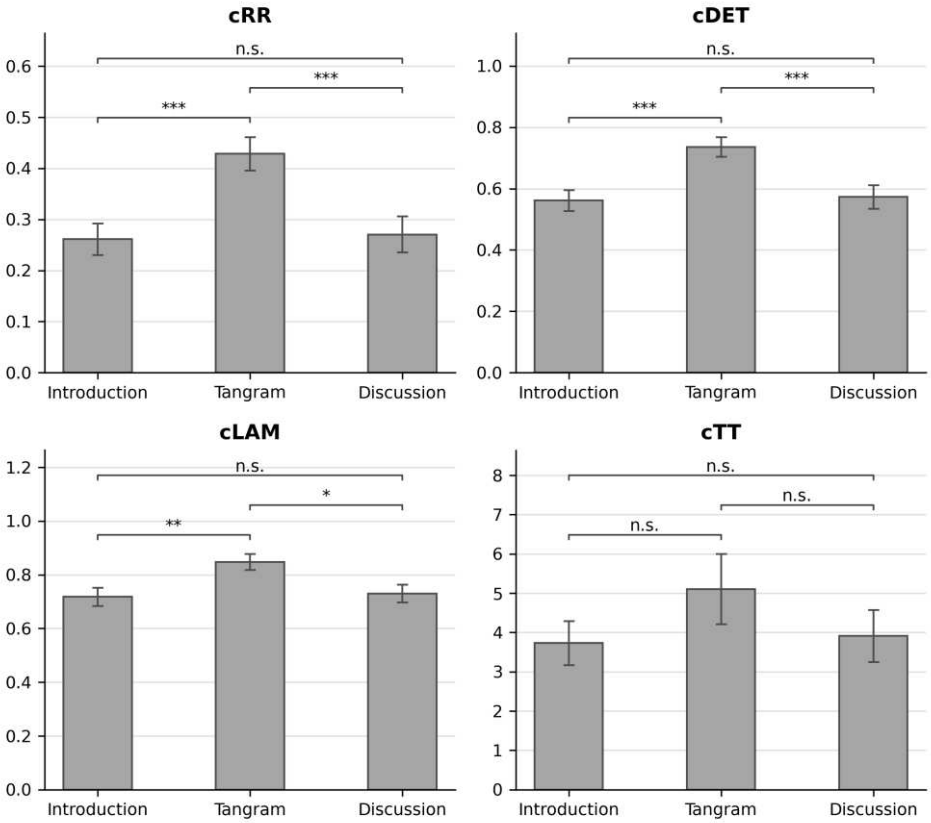


Figure 4. Cross-recurrence metrics by conversational task. Bars show the mean (with 95% CI) of each metric across dyad-task segments; the prefix "c" denotes the cross-recurrence analogue of each metric (cRR, cDET, cLAM, cTT). Brackets mark pairwise contrasts between tasks, as in Figure 3.

This pattern indicates that the more tightly constrained collaborative task (tangram) was associated with both stronger recurrence structure and lower local unpredictability.

Robust fixed-effects term tests were applied to assess these task effects. Metrics differed across tasks as follows (Wald χ^2 statistics; df = 2 throughout): RR (statistic = 13.32, df = 2, $p < .001$, $q < .001$), DET (statistic = 10.46, $p < .001$, $q < .001$), LAM (statistic = 12.33, $p < .001$, $q < .001$), TT (statistic = 4.48, $p = .0163$, $q = .0183$). The same held for cRR (statistic = 19.69, $p < .001$, $q < .001$), cDET (statistic = 21.43, $p < .001$, $q < .001$), cLAM (statistic = 11.86, $p < .001$, $q < .001$), cTT (statistic = 4.20, $p = .0206$, $q = .0206$), and entropy rate (statistic = 10.86, $p < .001$, $q < .001$). Together, these results show that the framework is sensitive to real differences in interactional structure across conversational contexts that were expected a priori, based on the tasks' differing communicative demands (Section 2.1), independent of the observed data. This supports its use as a general-purpose measurement tool (Table 1).

Table 1. Fixed-effects omnibus Wald tests for the task term across all nine dynamical summary metrics.

Metric	Statistic	df	p	q (FDR)
RR	13.32	2	<.001	<.001
DET	10.46	2	<.001	<.001
LAM	12.33	2	<.001	<.001
TT	4.48	2	.0163	.0183
cRR	19.69	2	<.001	<.001
cDET	21.43	2	<.001	<.001
cLAM	11.86	2	<.001	<.001
cTT	4.20	2	.0206	.0206
entropy rate	10.86	2	<.001	<.001

Metrics are as defined in Figures 3-4 (RR, DET, LAM, TT and their cross-recurrence analogues, plus Markov entropy rate); p-values are Benjamini-Hochberg FDR-corrected within each omnibus term (q).

3.3 Modality-specific contributions to conversational dynamics

LOCO analyses showed that speech and gaze were the dominant contributors to the structure of conversational dynamics, whereas backchannels and laughter had markedly smaller effects. When channels were dropped across all observations, removing speech produced the largest changes in recurrence-based metrics, with higher recurrence without the channel than with it (mean RR 0.154 with all channels vs 0.445 without speech): the mean LOCO delta (full − drop) was −0.303 for DET, −0.291 for RR, −0.232 for LAM, and −3.619 for TT. Removing gaze also substantially changed the recurrence structure in the same direction (mean RR 0.309 without gaze), with mean deltas of −0.256 for DET, −0.154 for RR, −0.205 for LAM, and −0.737 for TT. By contrast, removing backchannels changed DET by only −0.001 and RR by −0.007, while removing laughter changed DET by −0.003 and RR by −0.003.

The same ranking appeared in the information-theoretic metrics. For the six-step block entropy (cH6), the mean LOCO delta was 1.948 for speech, 0.928 for gaze, 0.016 for backchannels, and 0.008 for laughter. For excess entropy, the corresponding deltas were 19.916 (speech), 19.097 (gaze), 0.819 (backchannels), and 0.324 (laughter). Removing speech or gaze thus substantially reduced block entropy and excess entropy (positive deltas:

both were higher with the channel than without it), whereas removing backchannels or laughter had comparatively modest consequences (Table 2). Column permutation tests (Section 2.9) indicated that these contributions reflect the temporal organisation of the channels and not only their overall activity: compared with permuting the dropped channel in time, the observed data showed higher recurrence rate, determinism, laminarity, and trapping time and lower block entropy and entropy rate when speech was retained (median per-segment permutation $p = .005$ for each metric; mean $|z| = 5.9–15.4$), a weaker pattern for gaze (median $p \leq .02$ for DET, LAM, TT, block entropy, and entropy rate; RR n.s.), a mixed pattern for backchannels, and no reliable effect for laughter (all median $p > .80$).

Table 2. Leave-one-covariate-out (LOCO) mean deltas (full minus drop) by communicative channel, across all dyads and tasks.

Channel	RR	DET	LAM	TT	H6	excess entropy
Speech	-0.291	-0.303	-0.232	-3.619	1.948	19.916
Gaze	-0.154	-0.256	-0.205	-0.737	0.928	19.097
Backchannels	-0.007	-0.001	0.005	0.002	0.016	0.819
Laughter	-0.003	-0.003	-0.002	0.002	0.008	0.324

Task-wise permutation contrasts showed that the task dependence of the LOCO effects differed between the two dominant channels. The effect of dropping gaze was weaker in the tangram task than in discussion, with LOCO deltas that were smaller in magnitude in tangram by 0.102 for RR, 0.225 for DET, 0.197 for LAM, and 0.635 for TT (all $p \leq .001$; Table 3); the same pattern held relative to introduction. Speech showed the opposite pattern: its effects were stronger in the tangram task than in discussion and introduction for the recurrence, laminarity, trapping-time, and entropy metrics reported here (RR, DET, LAM, TT, six-step block entropy, excess entropy, and entropy rate; all $p \leq .01$). The contribution of speech was thus most evident when the interaction imposes strong structural demands, whereas that of gaze was most evident in the less constrained conversational contexts.

Table 3. Task-wise permutation contrast for dropping gaze.

Metric	Difference in |LOCO delta|, discussion minus tangram	p (perm)

RR	0.102	≤.001
DET	0.225	≤.001
LAM	0.197	≤.001
TT	0.635	≤.001

Values show how much larger the LOCO effect of dropping gaze (absolute delta, full − drop) was in the discussion context than in the tangram task; p-values are from the task-wise permutation contrast.

3.4 Summary of findings

Overall, the results show that conversational interaction combines clear global structure with persistent local variability. Recurrence plots and RQA metrics demonstrated that dyads repeatedly revisited constrained regions of multimodal state space rather than moving through it randomly. Entropy rates nonetheless remained substantial, especially in the more open-ended conversational contexts (Introduction and Discussion), indicating that this structure is not rigidly deterministic. The tangram task consistently produced the strongest recurrence structure and the lowest entropy, showing that the framework is sensitive to independently motivated differences in task demand. Across analyses, speech and gaze emerged as the principal channels shaping this organisation, whereas backchannels and laughter exerted small effects.

4 Discussion

The present framework shows that multimodal conversational interaction can be described as a structured dynamical system with local variability, by representing dyadic behaviour as trajectories through a symbolic state space defined by combinations of gaze, speech, backchannels, and laughter. The results are consistent with a metastability-informed interpretation of conversational dynamics, in which trajectories are globally constrained yet locally flexible: they return to preferred regions of state space without becoming locked into a strictly repetitive cycle (Brunton et al., 2017; Marwan et al., 2007; Tognoli & Kelso, 2014). Across the corpus, recurrence structure was clearly present, but determinism remained in the low-to-moderate range for most segments rather than approaching the very high levels expected for highly ordered periodic systems (Figure 2). Entropy rates remained substantial in a complementary fashion, especially in the less constrained conversational contexts. This indicates that conversational structure maintains order while preserving the flexibility necessary for adaptive interpersonal coordination.

This coexistence of structure and variability, as observed in the task comparison, where the more constrained tangram condition produced both higher recurrence and lower entropy, is the hallmark of metastable-like organisation (Kelso, 2012; Kelso, Dumas, & Tognoli, 2013). In coordination dynamics, metastability refers to a regime in which the tendencies of system components to function independently (segregation) and to coordinate as a coherent whole (integration) coexist without resolving into a single stable state or a fully disordered one (Engstrøm & Kelso, 2008; Tognoli et al., 2020). The present findings suggest that conversation may exhibit such a regime: interlocutors' behavioural channels are neither perfectly synchronised nor fully independent, but occupy a middle ground that allows both global coordination and local adaptivity.

4.1 Conversation as a metastable dynamical system

More broadly, this framework extends existing recurrence-based approaches to interpersonal coordination — such as cross-recurrence quantification analysis (CRQA; Coco & Dale, 2014) — from the single-channel case to the joint, multi-channel case, and pairs this extension with complementary information-theoretic summaries and a channel-ablation (LOCO) procedure. This channel-importance approach is conceptually related to permutation-based methods used elsewhere in recurrence-based complex-systems analysis (e.g., Bhardwaj & Das, 2019, applying RQA to an ecological trophic-chain model), but has not, to our knowledge, previously been combined with multimodal symbolic RQA/CRQA for human interaction data. The framework is best understood not as a replacement for CRQA-style methods but as an extension of them, allowing joint multimodal structure and each channel's relative contribution to it to be quantified within a single, unified framework rather than through separate single-channel analyses. Applied to the present corpus, this makes it possible to ask not only whether conversation is structured, but also which communicative channels are responsible for that structure — a question standard, single-channel recurrence methods cannot address.

Taken together, the results suggest that conversation can be understood as a metastable, multimodal dynamical system. This interpretation rests on three empirical features of the data. First, interaction trajectories repeatedly revisited specific regions of state space, as shown by the recurrence plots and non-zero recurrence rates. Second, the system displayed sequential organisation beyond isolated state revisitation, as reflected in determinism, laminarity, and trapping time. Third, this organisation coexisted with substantial entropy and only moderate determinism, indicating that the system remained locally variable rather than rigidly periodic. We do not propose a fixed quantitative threshold for “metastable” or “sequentially organised”; these terms describe the qualitative pattern observed here – non-trivial recurrence combined with incomplete determinism – relative to the two

limiting cases of fully random and fully periodic behaviour, rather than a claim that a specific numerical cutoff has been crossed. This framework helps reconcile a long-standing tension in how conversation can be described (Thórisson, 2008): conversation seems neither tightly clock-like nor essentially noisy but globally constrained while locally exploratory.

The present state-space approach is useful because it makes this interplay of structure and variability measurable across a series of channels. By encoding multimodal interaction as symbolic trajectories, it becomes possible to observe whether the system revisits prior configurations, how persistent those configurations are, and how much unpredictability remains in the transitions between them — going beyond traditional analyses focused on single modalities or isolated summary counts, and beyond the pairwise, single-channel comparisons typical of standard CRQA (Coco & Dale, 2014; see Section 1). It also complements accounts that emphasise oscillatory coordination by showing that interaction can be structured without being reducible to simple periodic synchronisation (Wilson & Wilson, 2005), extending this coordination-dynamics logic from simple rhythmic motor coordination (Haken et al., 1985) to complex multimodal conversation. This is consistent with broader work on interpersonal coordination showing that joint action supports, and is supported by, continuous behavioural alignment between partners (Marsh et al., 2009). It also fits evidence that the tendency for interlocutors' behaviours to converge across multiple modalities during communication is itself a synchronised, temporally structured process rather than an incidental by-product of interaction (Louwerse et al., 2012). Finally, it aligns with dynamical-systems approaches to social coordination that model interpersonal interaction as a self-organising process governed by lawful coupling dynamics (Coey, Varlet, & Richardson, 2012; Tognoli et al., 2020; Cornejo et al., 2017).

More broadly, the findings align with complexity-based approaches to communication, which argue that social interaction emerges from nonlinear coupling across multiple behavioural channels and interacting individuals (Sherry, 2015; Thórisson, 2008), and with multidimensional recurrence approaches showing that coordination in joint action can be quantified at the level of multivariate behavioural state spaces rather than single behavioural streams (Wallot et al., 2016). The growing use of recurrence-based dynamical methods well beyond their original motor-control or psycholinguistic applications — for instance in quantifying dyadic social-interaction dynamics between strangers of different backgrounds (Hamer et al., 2022) and in modelling recurrence structure in ecological trophic-chain systems (Bhardwaj & Das, 2019) — further supports treating the present framework as a general-purpose, transferable measurement tool rather than a method specific to this corpus. Recent work on analysing behavioural time series has also demonstrated the value of combining recurrence-based methods with other nonlinear and information-theoretic techniques for characterising multimodal interaction dynamics (Xu et

al., 2020), and dynamical frameworks for modelling interpersonal synchrony and complementarity as time-varying processes have begun to formalise the distinction between coordinated same-behaviour patterns and coordinated complementary patterns in conversation (Miao, Dale, & Galati, 2023).

The metastability interpretation should nonetheless be framed with appropriate restraint. The present analyses do not demonstrate chaos in the strict mathematical sense, nor do they establish attractor structure in the formal sense used in dynamical systems theory. Nonetheless, the data do show a pattern that is empirically consistent with a system with metastable organisation: structured recurrence, persistent local states, and ongoing variability within a constrained space of possible configurations.

4.2 Task effects on conversational structure

More generally, the framework can detect differences in interactional structure that are driven by external task demands — an important validation of its sensitivity, and a capability standard single-channel recurrence measures do not by themselves provide. A second central finding concerns the relevance of conversational context. The collaborative tangram task consistently produced the strongest recurrence structure and the lowest entropy. Both state-based and transition-based recurrence metrics increased in the tangram condition, while entropy rate decreased relative to the less constrained contexts of introduction and discussion. This pattern suggests that the demands of collaborative problem-solving force the interaction into a more limited subset of viable multimodal configurations. In the tangram task, participants must coordinate reference (e.g., agreeing on how to verbally describe an ambiguous tangram shape), monitor comprehension, and maintain shared task structure; these demands plausibly reward a more constrained behavioural structure, thereby reducing the effective degrees of freedom available to the system. By contrast, open the discussion allowed greater variability and showed weaker recurrence structure, indicating more flexible exploration of the interactional landscape, with the introduction phase potentially occupying a slightly more intermediate position.

4.3 Modality-specific contributions to conversational structure

Across LOCO analyses and across tasks, speech and gaze produced by far the largest changes in recurrence-based and information-theoretic summaries when removed from the model, whereas backchannels and laughter had comparatively small effects. More generally, the leave-one-covariate-out procedure introduced here (Section 2.9) makes it possible, for any multimodal dataset, to ask which communicative channels are most important for a

system's overall dynamical organisation shifting the scope beyond the question of whether the system is structured overall.

Speech can be understood as a fundamental structural channel because it organises the core alternation of speaker and listener roles. In conversational interaction, speech does not merely transmit propositional content; it also imposes a sequential scaffold on the interaction by structuring turn occupation, turn transition, and the temporal distribution of listening behaviour. This interpretation is consistent with classical work on turn-taking and conversational timing (Sacks et al., 1974; Wilson & Wilson, 2005; Levinson & Torreira, 2015) and with models of interactive alignment in which interlocutors achieve mutual understanding by aligning their representations at multiple linguistic levels (Pickering & Garrod, 2004). Removal of speech substantially raised recurrence and determinism and lowered block and excess entropy, indicating that speech adds distinguishable, sequentially organised states and thus contributes not only to local turns but to the global organisation of the interaction trajectory.

Gaze emerged as a second major structure-contributing channel. This is theoretically important because gaze is often discussed as a supportive or regulatory signal, but the present findings suggest a more substantial role in the dynamics of the conversation. Removing gaze strongly affected laminarity, trapping time, and informational structure: without gaze, interaction states persisted longer and entropy was lower, indicating that gaze contributes to the fine-grained differentiation of interaction states over time. This interpretation is in accordance with studies showing that gaze is crucial for joint attention, backchannel invitation, feedback timing, and the coordination of mutual orientation in interaction (Kendon, 1967; Hjalmarsson et al., 2012; Skantze et al., 2014; Hessels, 2020; Degutyte & Astell, 2021), and with evidence that eye contact marks the rise and fall of shared attention between interlocutors, dynamically regulating the coupling between partners rather than simply maintaining it (Wohltjen & Wheatley, 2021). The multimodal nature of turn coordination, in which gaze and manual gestures work alongside the speech signal to facilitate turn transitions (Kendrick, Holler, & Levinson, 2023), further underscores the importance of gaze as a structural channel in the dynamics of interaction.

By contrast, backchannels and laughter exerted much smaller effects on the overall dynamical organisation. This does not imply that these modalities are unimportant for communication; rather, their functional contribution may be more local and modulatory than globally structuring. Backchannels may regulate alignment, responsiveness, and affective stance on short timescales without determining the broad trajectory of state-space movement. Laughter may similarly punctuate interaction or shift its social tone without acting as a principal anchor of the

overall system. From the perspective of metastability, speech and gaze appear to function as the dominant information-bearing channels that shape the behavioural structuredness of the conversation, whereas backchannels and laughter contribute local perturbations. This division of communicative contribution is consistent with other accounts that distinguish between structural and modulatory signals in multimodal communication (Trujillo & Holler, 2023; Rasenberg, Özyürek, & Dingemanse, 2020). It is worth noting that the dominance of speech and gaze observed here is likely tied to the nature of the present data and communicative setting, in which participants engaged in an unconstrained, audible, face-to-face conversation with no strong competing demands on the vocal channel. In communicative settings where speech is less available or less informative — for example, in very noisy environments, in signed conversation, or in situations involving confidential information that discourages open verbal exchange — gaze or other channels might instead emerge as the primary structural anchor of the interaction. The present framework is agnostic to which channel plays this structuring role and could in principle be used to test such context-dependent shifts in future work.

5 Outlook

The present framework demonstrates that multimodal conversational interaction can be efficiently analysed as trajectories through a symbolic state space, and that this representation reveals systematic dynamical structure in face-to-face communication. Several directions for future methodological work follow naturally.

A first avenue concerns time-resolved dynamical analyses. The current approach characterises conversational trajectories using metrics computed over entire interaction segments. Sliding-window recurrence or entropy analyses could be used to examine how dynamical properties change from moment to moment, allowing the identification of transitions between more stable and more variable interaction phases within a single conversational task, and testing whether specific communicative events — such as gaze shifts, backchannels, or turn transitions — act as local triggers for changes in the system's dynamical regime. Future work could also move beyond recurrence and transition summaries by estimating directed information flow between modalities or partners, for example using transfer entropy to test whether changes in one channel improve prediction of subsequent changes in another (Schreiber, 2000).

Second, the symbolic state representation used here could be extended to incorporate richer behavioural information — for example, kinematic data such as manual gestures, head nods, or blinks, which have been suggested to serve as additional anchors of local variability in interaction (Trujillo & Holler, 2023) — by reading

out continuous signals as graded variables (e.g., the amplitude of the voice, or the spatial distance between a partner's gaze and their interlocutor's face) rather than binary presence/absence indicators. A related extension concerns quantifying chaos more formally. The present analyses deliberately avoid continuous phase-space reconstruction and instead rely on exact recurrence within a discrete, combinatorial state space (Section 1); this choice is well suited to the multichannel, presence/absence nature of the behavioural data but stops short of the classical toolkit used to characterise low-dimensional continuous chaotic systems, such as Lyapunov exponents, which quantify the rate of divergence of nearby trajectories (Dingwell, 2006) and have been applied to detect chaotic regimes in domains as varied as vessel motion (McCue & Troesch, 2011) and, more generally, in the broader mathematical framework of chaotic dynamics (Cvitanović et al., 2016; Chandre et al., 2012). Extending the present framework to continuous or graded behavioural signals could make such classical measures applicable in future work, complementing the present symbolic, discrete-state approach rather than replacing it.

A further extension concerns larger and more diverse interaction datasets. The present analyses focused on dyadic conversations within three specific contexts; additional interaction settings — multi-party conversations, instructional interactions, different groups of interlocutors, or clinical consultations — could reveal how conversational dynamics adapt to different communicative environments, and applications to clinical populations of interest are a natural extension of this general-purpose framework (see Wünsche et al., in prep., for one such application).

Finally, the present approach opens possibilities for linking behavioural dynamics to neural and cognitive processes. Because conversational interaction unfolds through tightly coordinated multimodal behaviour, dynamical summaries of interaction trajectories could potentially be related to neural measures of interpersonal coordination or cognitive models of communication.

6 Conclusion

This paper introduced a general methodological framework for quantifying the dynamical organisation of multimodal interaction by modelling it as trajectories through a symbolic state space defined by combinations of communicative signals, together with a symbolic transition-network representation of the resulting state dynamics and a leave-one-covariate-out procedure for identifying which channels drive that organisation — tools that, to our knowledge, existing single-channel recurrence methods do not by themselves provide. Applied here to a corpus of dyadic conversations as an illustrative case. Recurrence analysis revealed that conversational behaviour

repeatedly revisits specific regions of this state space, indicating strong sequential organisation, while entropy-based measures demonstrated that conversational trajectories retain substantial local variability. These findings support the view that conversational interaction is neither random nor rigidly periodic, but instead evolves within a constrained behavioural landscape in which stable interaction states coexist with flexible moment-to-moment variation. This organisation depends strongly on interaction context: the structured collaborative task produced higher recurrence and lower entropy than the less constrained conversational contexts. The analyses further revealed clear modality-specific contributions, with speech and gaze emerging as the primary channels shaping the structure of interaction trajectories, and backchannels and laughter contributing comparatively smaller effects on global dynamical organisation and rather contributing more locally. By formalising conversational behaviour as trajectories through a multimodal symbolic state space, this framework provides a general, transferable quantitative method for analysing the sequential organisation of human interaction, applicable to any dataset in which multiple communicative channels can be time-aligned and discretised.

List of abbreviations

RQA, recurrence quantification analysis; CRQA, cross-recurrence quantification analysis; RR, recurrence rate; DET, determinism; LAM, laminarity; TT, trapping time; cRR, cDET, cLAM, cTT, cross-recurrence analogues of RR, DET, LAM, and TT; LOCO, leave-one-covariate-out; OLS, ordinary least squares; HC3, heteroscedasticity-consistent (robust) covariance estimator, version 3; FDR, false discovery rate.

Declarations

Availability of data and materials

The raw behavioural annotation data analysed in the present study are available from the corresponding author on reasonable request; further detail on the recording protocol is given in Spaniol (2025, dissertation) and at OSF: osf.io/z5crx.

The analyses (RQA/CRQA, information-theoretic and Markov summaries, LOCO procedure, and statistical models) used in this paper are openly available. Project home page: https://doi.org/10.5281/zenodo.23090538. Archived version: https://doi.org/10.5281/zenodo.23090538. Programming language: Python (>=3.9). Other requirements: NumPy, SciPy, pandas, statsmodels, matplotlib. Licence: MIT License.

Competing interests

The authors declare that they have no competing interests.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. MG and KV received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252. MG and MJ received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 563096772. KV received funding from the German Ministry of Research and Education [Project SIMSUB; grant ID 01GP2215].

Authors' contributions

MW conceived the methodological framework, implemented the analysis pipeline, analysed the data, and drafted the manuscript. MS collected and annotated the conversational corpus and contributed to data interpretation and critically revised the manuscript. MJ contributed to framework development and critically revised the manuscript. MG and KV supervised the project and critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

Berkhout, S. W., Schuurman, N. K., & Hamaker, E. L. (2025). A tool to simulate and visualize dyadic interaction dynamics. *Psychological Methods*, *30*(3), 599–621. https://doi.org/10.1037/met0000575

Bhardwaj, R., & Das, S. (2019). Recurrence quantification analysis of a three level trophic chain model. *Heliyon*, *5*(8), e02182. https://doi.org/10.1016/j.heliyon.2019.e02182

Bloch, C., Vogeley, K., Georgescu, A. L., & Falter-Wagner, C. M. (2019). INTRApersonal synchrony as constituent of INTERpersonal synchrony and its relevance for autism spectrum disorder. Frontiers in Robotics and AI, 6, 73. https://doi.org/10.3389/frobt.2019.00073

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Brunton, S. L., Brunton, B. W., Proctor, J. L., Kaiser, E., & Kutz, J. N. (2017). Chaos as an intermittently forced linear system. *Nature Communications*, *8*(1), 19. https://doi.org/10.1038/s41467-017-00030-8

Chandre, C., Diakonos, F. K., & Schmelcher, P. (2012). Discrete iteration relaxation method. In P. Cvitanović, R. Artuso, R. Mainieri, G. Tanner, & G. Vattay (Eds.), *Chaos: Classical and Quantum* (14th ed.). Niels Bohr Institute. http://ChaosBook.org/version14/chapters/ChaosBook.pdf

Clark, H. H. (1996). *Using Language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00510

Coey, C. A., Varlet, M., & Richardson, M. J. (2012). Coordination dynamics in a socially situated nervous system. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00164

Cornejo, C., Cuadros, Z., Morales, R., & Paredes, J. (2017). Interpersonal Coordination: Methods, Achievements, and Challenges. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01685

Crutchfield, J. P., & Young, K. (1989). Inferring statistical complexity. *Physical Review Letters*, *63*(2), 105–108. https://doi.org/10.1103/PhysRevLett.63.105

Crutchfield, J. P., & Feldman, D. P. (2003). Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos*, *13*(1), 25–54. https://doi.org/10.1063/1.1530990

Cvitanović, P., Artuso, R., Mainieri, R., Tanner, G., & Vattay, G. (2016). *Chaos: Classical and Quantum*. Niels Bohr Inst. http://ChaosBook.org/

Degutyte, Z., & Astell, A. (2021). The Role of Eye Gaze in Regulating Turn Taking in Conversations: A Systematized Review of Methods and Findings. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.616471

Dingwell, J. B. (2006). Lyapunov Exponents. In *Wiley Encyclopedia of Biomedical Engineering*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780471740360.ebs0702

Engstrøm, D. A., & Scott Kelso, J. (2008). COORDINATION DYNAMICS OF THE COMPLEMENTARY NATURE. *Gestalt Theory*, *30*, 121–134.

Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, *33*(2), 1134–1140. https://doi.org/10.1103/PhysRevA.33.1134

Fusaroli, R., Konvalinka, I., & Wallot, S. (2014). Analyzing Social Interactions: The Promises and Challenges of Using Cross Recurrence Quantification Analysis. In N. Marwan, M. Riley, A. Giuliani, & Jr. Webber

Charles L. (Eds.), *Translational Recurrences* (pp. 137–155). Springer International Publishing. https://doi.org/10.1007/978-3-319-09531-8c9

Georgescu, A. L., Koeroglu, S., Hamilton, A. F., Vogeley, K., Falter-Wagner, C. M., & Tschacher, W. (2020). Reduced nonverbal interpersonal synchrony in autism spectrum disorder independent of partner diagnosis: A motion energy study. Molecular Autism, 11(1), 11. https://doi.org/10.1186/s13229-019-0305-1

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, *32*(10), 1489–1522. https://doi.org/10.1016/S0378-2166(99)00096-X

Haken, H., Kelso, J. A., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*(5), 347–356. https://doi.org/10.1007/BF00336922

Hamer, K., López Pérez, D., Drogosz, M., & Goworek, H. (2022). The dynamics of dyadic interactions between people of different ethnicities depend on their identification with all humanity. *Scientific Reports*, *12*(1), 21671. https://doi.org/10.1038/s41598-022-25905-9

Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review*, *27*(5), 856–881. https://doi.org/10.3758/s13423-020-01715-w

Hjalmarsson, A., Oertel, C., & Speech, K. (2012). *Gaze direction as a Back-Channel inviting cue in dialogue*. *Joint Proceedings of the Intelligent Virtual Agents 2012 Workshops:Santa Cruz, CA, September 15, 2012*.

Kelso, J. A. S. (2012). Multistability and metastability: Understanding dynamic coordination in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 906–918. https://doi.org/10.1098/rstb.2011.0351

Kelso, J. A. S., Dumas, G., & Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Networks*, *37*, 120–131. https://doi.org/10.1016/j.neunet.2012.09.003

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, *26*, 22–63. https://doi.org/10.1016/0001-6918(67)90005-4

Kendrick, K. H., Holler, J., & Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *378*(1875), 20210473. https://doi.org/10.1098/rstb.2021.0473

Lempel, A., & Ziv, J. (1976). On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory*, *22*(1), 75–81. https://doi.org/10.1109/TIT.1976.1055501

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00731

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior Matching in Multimodal Communication Is Synchronized. *Cognitive Science*, *36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*(2), 320–339. https://doi.org/10.1111/j.1756-8765.2009.01022.x

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5), 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

McCue, L. S., & Troesch, A. W. (2011). Use of Lyapunov Exponents to Predict Chaotic Vessel Motions. In M. Almeida Santos Neves, V. L. Belenky, J. O. de Kat, K. Spyrou, & N. Umeda (Eds.), *Contemporary Ideas on Ship Stability and Capsizing in Waves* (pp. 415–432). Springer Netherlands. https://doi.org/10.1007/978-94-007-1482-3c23

Miao, G. Q., Dale, R., & Galati, A. (2023). (Mis)align: A simple dynamic framework for modeling interpersonal coordination. *Scientific Reports*, *13*(1), 18325. https://doi.org/10.1038/s41598-023-41516-4

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Plank, I. S., Traiger, L. S., Nelson, A. M., Koehler, J. C., Lang, S. F., Tepest, R., Vogeley, K., Georgescu, A. L., & Falter-Wagner, C. M. (2023). The role of interpersonal synchrony in forming impressions of autistic and non-autistic adults. Scientific Reports, 13(1), 15306. https://doi.org/10.1038/s41598-023-42006-3

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*(11), e12911. https://doi.org/10.1111/cogs.12911

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

Schreiber, T. (2000). Measuring Information Transfer. *Physical Review Letters*, *85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461

Sherry, J. (2015). The Complexity Paradigm for Studying Human Communication: A Summary and Integration of Two Fields. *Review of Communication Research*, *3*, 22–65. https://doi.org/10.12840/issn.2255-4165.2015.03.01.007

Skantze, G., Hjalmarsson, A., & Oertel, C. (2014). Turn-taking, feedback and joint attention in situated human–robot interaction. *Speech Communication*, *65*, 50–66. https://doi.org/10.1016/j.specom.2014.05.005

Smart, B., Roughan, M., & Mitchell, L. (2024). The entropy rate of Linear Additive Markov Processes. *PLOS ONE*, *19*(4), e0295074. https://doi.org/10.1371/journal.pone.0295074

Spaniol, M., Wehrle, S., Janz, A., Vogeley, K., & Grice, M. (2024). The influence of conversational context on lexical and prosodic aspects of backchannels and gaze behaviour. *Speech Prosody 2024*, 607–611. https://doi.org/10.21437/SpeechProsody.2024-123

Spaniol, M. (2025). Face to face: exploring human social interaction through gaze, speech and neurodiversity [PhD thesis]. University of Cologne. https://osf.io/z5crx

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Thórisson, K. R. (2008). Modeling Multimodal Communication as a Complex System. In I. Wachsmuth & G. Knoblich (Eds.), *Modeling Communication with Robots and Virtual Humans* (Vol. 4930, pp. 143–168). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-79037-2c8

Tognoli, E., & Kelso, J. A. S. (2014). The Metastable Brain. *Neuron*, *81*(1), 35–48. https://doi.org/10.1016/j.neuron.2013.12.022

Tognoli, E., Zhang, M., Fuchs, A., Beetle, C., & Kelso, J. A. S. (2020). Coordination Dynamics: A Foundation for Understanding Social Behavior. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00317

Trujillo, J. P., & Holler, J. (2023). Interactionally Embedded Gestalt Principles of Multimodal Human Communication. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *18*(5), 1136–1159. https://doi.org/10.1177/17456916221141422

Wallot, S., & Leonardi, G. (2018). Analyzing Multivariate Dynamics Using Cross-Recurrence Quantification Analysis (CRQA), Diagonal-Cross-Recurrence Profiles (DCRP), and Multidimensional Recurrence Quantification Analysis (MdRQA) – A Tutorial in R. *Frontiers in Psychology*, *9*, 2232. https://doi.org/10.3389/fpsyg.2018.02232

Wallot, S., Roepstorff, A., & Mønster, D. (2016). Multidimensional Recurrence Quantification Analysis (MdRQA) for the Analysis of Multidimensional Time-Series: A Software Implementation in MATLAB

and Its Application to Group-Level Data in Joint Action. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01835

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957–968. https://doi.org/10.3758/BF03206432

Wohltjen, S., & Wheatley, T. (2021). Eye contact marks the rise and fall of shared attention in conversation. *Proceedings of the National Academy of Sciences*, *118*(37), e2106645118. https://doi.org/10.1073/pnas.2106645118

Wünsche, M., Kaspar, K., & Vogeley, K. (2026). Reframing rapport across communicational domains. *Frontiers in Communication*, *11*, 1746483. https://doi.org/10.3389/fcomm.2026.1746483

Wünsche, M., Spaniol, M., Grice, M., & Vogeley, K. (2026, June 4). Modality-Specific Contributions to Multimodal Interaction Dynamics. *Book of Abstracts of The 6th International Symposium on Complex Systems, June 03-05, 2026, La Rochelle, France*. ISCS2026. https://doi.org/10.5281/zenodo.20543415

Xu, T. L., de Barbaro, K., Abney, D. H., & Cox, R. F. A. (2020). Finding Structure in Time: Visualizing and Analyzing Behavioral Time Series. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01457

Zbilut, J. P., Giuliani, A., & Webber, C. L. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, *246*(1–2), 122–128. https://doi.org/10.1016/S0375-9601(98)00457-5