Supplementary material for:

Local and global rhythmic dynamics in small-group conversations

Journal of Nonverbal Behavior

Arodi Farrera[1,2], Caleb Rascon[1] and Gabriel Ramos-Fernández[1]*

[1] Institute for Research on Applied Mathematics and Systems, National Autonomous University of Mexico
[2] Institute of Anthropological Research, National Autonomous University of Mexico

*Corresponding author: ramosfer@alumni.upenn.edu

Methodological details

Participants

ID	EN2001a	EN2001d	EN2001e	EN2001b	EN2002a	EN2002b
Nomenclature	1A	1B	1C	2A	2B	2C
MEE073					H0	H3
FEO070					H1	H0
FEO072					H2	H2
MEE071					H3	H1
FEO066	H1	H0	H0	H0		
MEO069	H4	H1	H1	H1		
FEO065	H2	H2	H2	H2		
MEE068	H0	H4	H4	H3		
MEE067	H3	H3	H3			

Table 1_supp. Participant sample composition for all meetings. The Participants IDs are denoted as H1 through H5.

Data

Onsets of voiced utterances:

The data extraction method results in a set of onsets that represent the start of voiced utterances. These acoustic events are sensitive to non-verbal acoustic events such as laughter and coughing. This is expected in more spontaneous and real-world contexts (non-laboratory conditions as in this work), which are necessary to understand the scope of rhythmic abilities (Henry et al., 2021). Notwithstanding, these onsets can be seen as domain-general elementary units that capture the presence and absence of acoustic behavior from which the temporal structure can be analyzed.

Since the presence of acoustic events at any given moment depends on speech rate, rhythm, and tempo are closely related (Auer et al., 1999). To minimize the effects of this dependency on our results, the onsets used to calculate the structural organization and those used to calculate the tempo were extracted in two slightly different ways. In the first case, the set of onsets was extracted over the entire audio file that mixes the speech signals of all the participants. In the second case, this audio file was also used, but the onsets were only calculated within dialogue interventions. Note that these datasets differ in that the first one highlights the fluency of acoustic events in the entire meeting (including prolonged silences), while the second to the fluency of these events between successive speaking turns.

BeatRoot configuration used:

To estimate the BPM from a given set of onsets, BeatRoot employs a multi-agent architecture to provide different hypotheses of beat sequences (one for each agent). Each agent is evaluated in terms of how well its hypothesis fits a clustered version of the given set of onsets as well as how evenly-spaced its beats are. Each agent is then iteratively modified to better predict the set of onsets. The agent with the highest score is used to calculate the final BPM estimate. The following configuration was used: a cluster size of 3, the momentum/reactiveness of all agents was set to 0.35, and a border-like time window of 0.005 s. was applied to each side of an agent's hypothesis length to accept an onset as part of its beat sequence, a tolerance time window of 0.05 s. was applied to each side of an agent's hypothesis length to let outside onsets affect its BPM estimate, and 25 starting onsets were employed to start the multi-agent search. Additionally, the following score matrix was used to evaluate each agent: a weight of 10 is given to the amount of hits the beat sequence obtained, a weight of -3 is given to the amount of misses, a weight of 5 is given to how well spaced it is related to the clustered version of the given set of onsets, and a weight of -0.01 to the average distance between each hit and its respective onset.

This process was applied to each dialogue intervention in each meeting recording and resulted in a set of BPM estimates, one for each intervention. It is important to specify that when analyzing some of the meetings, because of the nature of the voice of some individuals (low volume, short and unintelligible interventions, etc.) some interventions were ignored. However, only a few interventions were affected. Additionally, it is important to note that this method provides one BPM estimation per intervention, regardless of its length. Although it would be of interest to estimate the variation of the BPM throughout each intervention, it has been found (Yuan et al., 2006) that speaking rates tend to stay invariant throughout interventions of lengths that are similar to those investigated here (8 to 30 words). Thus, for the time being, we assume one BPM estimation per intervention; we will carry out further analysis where we consider a variable BPM per intervention in future work.

Dialogue interventions:

A dialogue intervention is an intervention whose start time is less than 2 seconds apart from the end time of the last intervention. This definition captures both interventions in which only one participant speaks (Figure 1a_supp, C) and interventions in which participants are in dialogue (Figure 1a_supp, A and B), including when their interventions overlap.

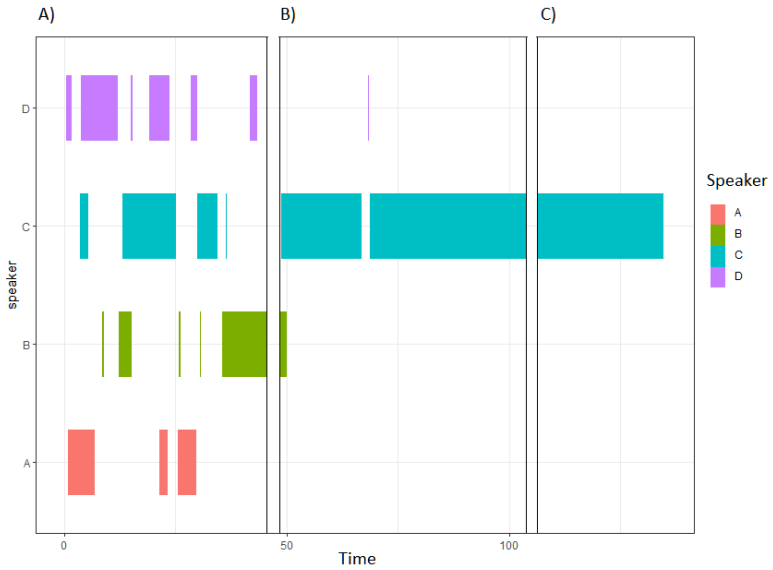

Figure 1a_supp. Dialogue interventions.

Interpretation of three of the metrics used

In this supplementary material, we simulated three example time series to guide and aid in the interpretation of the structural complexity metrics used, i.e. the RQA measurements.

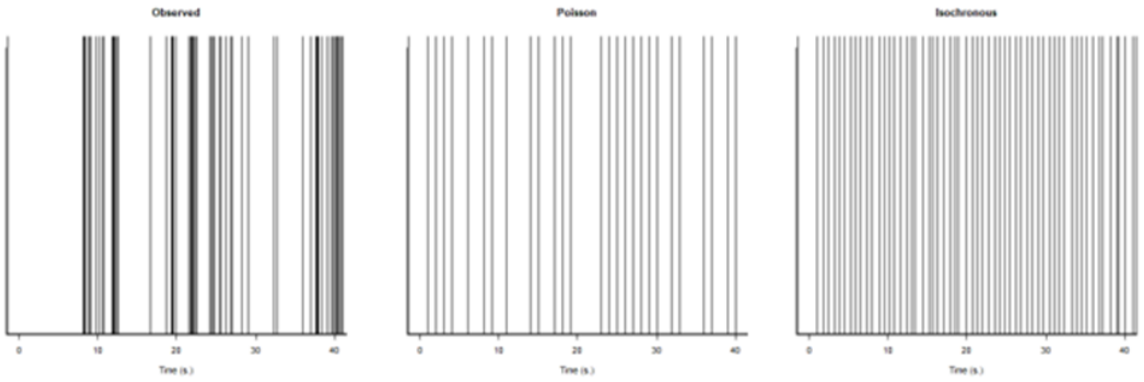

Figure 2a_supp. Plot of an extract of an observed time series and three example time series with 3000 events and a set of inter-event-intervals drawn from a Poisson (λ = 0.7) and periodic (or isochronous) (μ= 0.7, σ = 0.1) distribution.

Recurrence quantification analysis

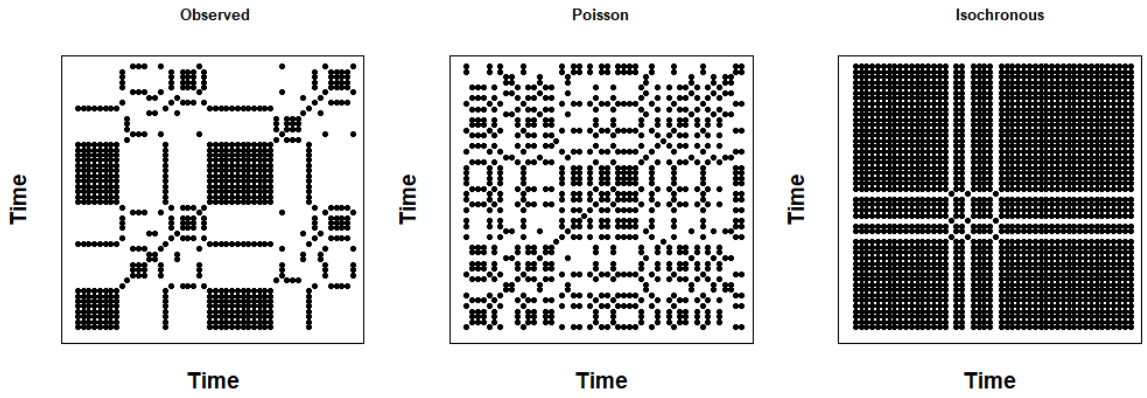

Figure 3a_supp. RQA measurements. Recurrence plot of an extract of three time series: %LAM, %DET and ENTR quantify, respectively, the percentage of vertical lines, the percentage of diagonal lines, and the variability in length of the diagonal lines. To perform a RQA on categorical time series, we used the following parameter settings of the crqa() function from the crqa package in R: delay = 1; embed = 1; rescale = 0; radius = 0.0001; normalize = 0; mindiagline = 2; minvertline = 2; tw = 1; whiteline = FALSE; recpt = FALSE; side = "both"; method = 'rqa'; metric = 'euclidean'; datatype = "categorical".

This analysis shows that the time series have a different non-linear organization of recurrent events. The higher the %DET and %LAM values, the more regular the recurrent events. The higher the ENTR values the more variability is observed in diagonal lines. Higher values of %DET, %LAM, and ENTR can be interpreted as a highly structured organization, while lower values as highly unstructured.

Accordingly, compared to the other simulated time series, the Poisson time series shows a less structured organization (%DET = 52.23, %LAM = 49.13, ENTR = 0.663), while the Isochronous or periodic time series shows a highly structured organization (%DET = 98.56, %LAM = 99.59, ENTR = 2.39). The Observed time series on the other hand shows an intermediate structured organization (%DET = 68.57, %LAM = 78.72, ENTR = 2.04).

References:

Henry, M., Cook, P., de Reus, K., Nityananda, V., Rouse, A., Kotz, S. (2021). An ecological approach to measuring synchronization abilities across the animal kingdom.

Auer, P., Couper-Kuhlen, E., Muller, F. (1999). Language in time. The rhythm and tempo of spoken interaction. Oxford studies in sociolinguistics.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In Ninth International Conference on Spoken Language Processing