DIMS Dashboard for Exploring Dynamic Interactions and Multimodal Signals

Grace Qiyuan Miao
q.miao@ucla.edu

James Trujillo
j.p.trujillo@uva.nl

Landry S. Bulls
landry.s.bulls.gr@dartmouth.edu

Mark A. Thornton
Mark.A.Thornton@dartmouth.edu

Rick Dale
rdale@ucla.edu

Wim Pouw
w.pouw@tilburguniversity.edu

Abstract

Social interaction is a complex, multimodal phenomenon with varying timescales and meaning-making structures. Research in this area has progressed along two largely separate paths: qualitative researchers focus on fine-grained analysis, while quantitative researchers computationally identify broader patterns. To bridge this gap and promote cross-disciplinarity, we developed the Dynamic Interaction and Multimodal Signals (DIMS) Dashboard, an open tool for visualizing multimodal data, enabling a qualitative-quantitative synergy in social interaction research. We overview its development, and conduct a proof-of-concept qualitative-quantitative ("quali-quanti") analysis using neural and behavioral time-series data combined with video recording. Our exploratory case study reveals that 80% of segments with sharply increased neural activations in the right temporoparietal juncture align with highly engaged interaction, while 20% correspond to topic transitions. Through triangulation with qualitative insights, we observed that social brain synchrony relates in meaningful moments to head motion synchrony. Finally, we discuss how visualization tools like DIMS enhance multimodal, cross-disciplinary research in social interaction, and tool development.

Keywords: multimodal signals; social interaction; data visualization dashboard; dynamic data display

Introduction

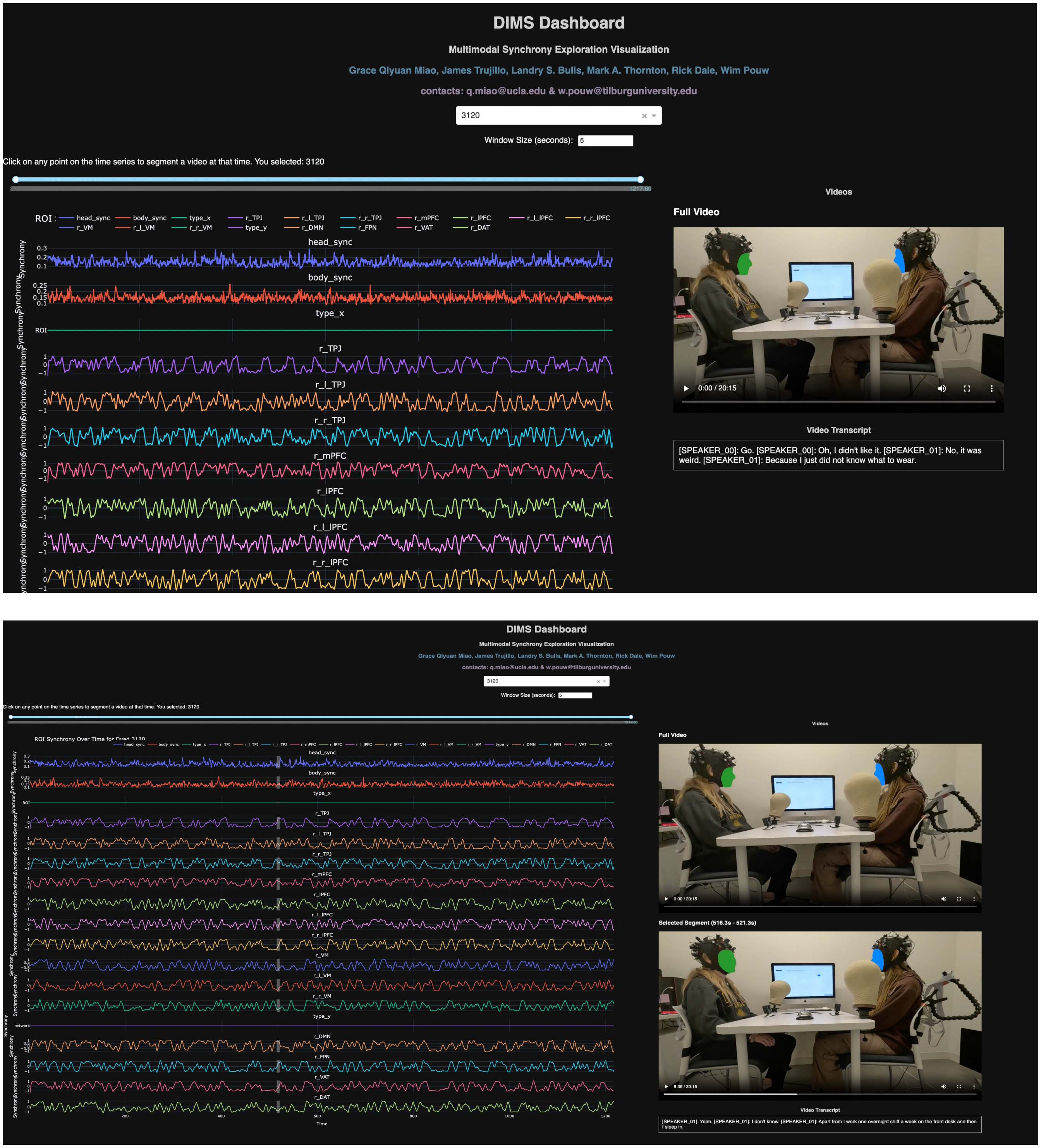
Social interaction is a fundamental human experience that occurs everyday, but is far from trivial. Social interactions are complex by nature as they involve moment-by-moment monitoring of subtle cues that are colloquially described as change of vibes from interaction partners. These subtle cues could be a sentence, a nod, or a frown, all of which are external bodily components of interaction partners' internal mental, neurocognitive processes (Wheatley et al., 2024; Cheong et al., 2023; King-Casas et al., 2008). These components operate on different time scales, but collectively make sense (Thibault, 2020). Successful social interactions require skills to understand these cues in context and instantaneously repair misunderstanding through tweaking word choice, prosody, gesture; moving closer or farther apart from one another; or making and breaking eye contact (Wheatley et al., 2024; Clark, 1996; Cooney & Wheatley, in press; Wohltjen & Wheatley, 2021; Di Paolo et al., 2018).

The importance and complexity of social interaction have attracted scholarly interests across qualitative and quantitative methodological traditions. In the last 50 years, qualitative scholars in the conversation analysis (CA) tradition developed systematic techniques to explore the everyday worlds of ordinary people through in-depth microanalysis of the language they use in interaction, focusing on social actions such as requests, offers, questions, etc. (Sidnell & Stivers, 2012). Quantitative scholars captured global patterns of linguistic or nonverbal signals within large interactional datasets, using a variety of computational tools including cross-recurrence quantification analysis (Alviar et al., 2023; Wallot, 2017), natural language processing (Giulianelli & Fernandez, 2021; Yazdian, Chen, & Lim, 2022), convergence entropy (Rosen & Dale, 2024), and more. Recently, social neuroscientists started taking a second-person neuroscience perspective (Schilbach et al., 2013) to investigate the interacting minds during social interaction (Wheatley et al., 2024) using portable neuroimaging techniques like EEG and functional near infrared spectroscopy (fNIRS) (Pinti et al., 2018; Pinti et al., 2020).

Decades of research in these traditions have deepened our understanding of human social interaction. For example, qualitative conversation analysis research has shown how different word choices form action, territory of knowledge, and epistemic stance (Heritage, 2012); statistical cross-cultural analysis uncovered universal turn-taking gaps across languages (Stivers et al., 2009); and social neuroscience identified neurocognitive patterns linked to friendship (Parkinson et al., 2018). Yet, qualitative and quantitative methods of inquiry tend to be separate from one another, likely due to differences in theoretical perspectives (e.g., bottom-up vs. top-down), implementation (e.g., manual vs. computational), and research aims (e.g., examining targeted samples of gesture vs. quantifying body motion over thousands of samples).

Ideally, challenges should not divide disciplines, but rather foster collaboration and tool development that leverages each discipline's strengths for understanding social interaction (Dale et al., 2022). To this end, we developed the DIMS Dashboard for dynamic interaction and multimodal signals, which integrates audiovisual, kinematic, neural, and transcript data into one space. By displaying the qualitative data source (i.e. video) side-by-side with quantitative signals extracted from the same source (i.e., neural fluctuations, body movements, synchrony scores), we aim to provide a methodological solution to quali-quanti collaboration.

This paper presents the DIMS Dashboard (Fig. 1) as a tool to bridge the qualitative-quantitative divide and support mixed-method research. Multimodal, time-varying data and scholars probing into temporal dynamics are especially likely to benefit from more versatile ways of disseminating research data. As elaborated in the Discussion section, integrating interactive dashboards alongside traditional static research could help validate complex measures and enhance reproducibility.



DIMS Dashboard

Multimodal Synchrony Exploration Visualization

Grace Qiyuan Miao, James Trujillo, Landry S. Bulls, Mark A. Thornton, Rick Dale, Wim Pouw

contacts: q.miao@ucla.edu & w.pouw@tilburguniversity.edu

3120

Window Size (seconds): 5

Click on any point on the time series to segment a video at that time. You selected: 3120

Videos

Full Video

Video Transcript

[SPEAKER_00]: Go. [SPEAKER_00]: Oh, I didn't like it. [SPEAKER_01]: No, it was weird. [SPEAKER_01]: Because I just did not know what to wear.

DIMS Dashboard

Multimodal Synchrony Exploration Visualization

Grace Qiyuan Miao, James Trujillo, Landry S. Bulls, Mark A. Thornton, Rick Dale, Wim Pouw

contacts: q.miao@ucla.edu & w.pouw@tilburguniversity.edu

3120

Window Size (seconds): 5

Click on any point on the time series to segment a video at that time. You selected: 3120

ROI Synchrony Over Time for Dyad 3120

Videos

Full Video

Selected Segment (516.3s - 521.3s)

Video Transcript

[SPEAKER_01]: Yeah. [SPEAKER_01]: I don't know. [SPEAKER_01]: Apart from I work one overnight shift a week on the front desk and then I sleep in.

Figure 1. The DIMS Dashboard.
Top: Screenshot of the initial setting (https://tinyurl.com/dimsdashboard), displaying time series of head synchrony, body synchrony, and neural synchrony across various social brain regions or networks on the left, and video data source with transcript on the right.
Bottom: Screenshot with the "rapid video preview" function, demonstrating a custom-set window size of 5 seconds. When a user selects a specific time point on the time series (highlighted in gray), a 5-second video segment corresponding to the selection appears with transcripts updated accordingly.

Methods

Application Building

Building on EnvisionBox, an open-source pedagogical platform for social signal processing, we drew inspiration from its dynamic dashboard module[1] linking video to "gesture networks" of gesture embedding space capturing kinematic similarities via dynamic time warping (Pouw & Dixon, 2020). This approach for visualizing gesture recurrences has recently been extended to support qualitative investigation of multimodal vocal performance (Pearson, Nuttall, & Pouw, 2024).

In this current work, we further developed the *multimodal* potential of such a data visualization tool by displaying conversation video and transcript alongside neurocognitive activity and bodily synchrony extracted during the same conversation. To support exploration of long interactions, we added a "rapid video preview" function, which loads an additional short video preview with a custom-set window length (e.g. 5 seconds) when the researcher selects a specific moment in the neural or bodily synchrony time series (Fig.1). This feature makes it easy for researchers to qualitatively examine key moments in the time series of various bio-behavioral signals (e.g., peaks in neural and bodily synchrony) by linking them directly to specific video clips, enabling a more integrated interpretation of the data.

The DIMS Dashboard is available for public access (https://tinyurl.com/dimsdashboard).

Backend Infrastructure The dashboard is built with Plotly Dash (Python 3.8), a Python framework for interactive web applications. We added JavaScript modules to enable rapid video segment selection not supported natively by Dash.

Launching The app runs on a server using Apache2 for HTTP support. It can also be launched locally by executing the Python script in a terminal.

Multimodal Data

The ConvoConnect Dataset consists of 70 stranger dyads engaging in 20-minute get-to-know-you conversations over either shallow or deep topics. During these conversations, participants were equipped with a portable neuroimaging technology–functional near-infrared spectroscopy (fNIRS)– to capture neurocognitive activities, as well as audiovisual recordings for movement analysis (Miao et al., 2024). The DIMS Dashboard features two sample dyads from this multimodal communication dataset.

Brain Recording The fNIRS montage covers cortical regions implicated in social interactions, including medial prefrontal cortex (mPFC), temporo-parietal junction (TPJ), lateral prefrontal cortex (lPFC) and superior parietal lobule (SPL). Based on the widely used Yeo et al. (2011) parcellation, this montage could also be characterized in a network-based approach, covering regions in the default networks (DMN), the frontoparietal network (FPN), the ventral attention network (VAN), and the dorsal attention network (DAN).

Video Recording Three GoPro cameras were placed in the room to record conversations and nonverbal behaviors. Specifically, one camera was placed in front of each participant to record their facial expressions and the third camera captured both participants together. The latter is displayed in the DIMS Dashboard as video data, and the former was used for bodily synchrony analysis.

Transcripts Understanding conversation requires knowing who spoke, what was said, and when. We used WhisperX (Bain et al., 2023), an enhanced version of Whisper (Radford et al., 2023), to perform both speaker diarization—identifying who spoke each segment in a single-source audio—and transcription—converting audio into text.

WhisperX uses forced alignment (Moreno et al., 1998) to generate precise word- and segment-level timestamps, mapping each utterance to its time and speaker. The outputs of diarization and transcription are automatically merged, producing a structured transcript with speaker labels and word-level timing. Each speaker's transcript is hierarchically organized into continuous utterance segments, enabling precise alignment with neural, behavioral, and audiovisual modalities for integrated analysis.

Preprocessing

Masking We masked participants' faces for privacy protection using a custom Python script based on SAM2 (Ravi et al., 2024) to identify and mask regions of interest.

Bodily synchrony We calculated head rotation using MediaPipe (Pouw, 2024; Lugaresi et al., 2019) and took the speed of rotation (angular speed) as our movement time series. Time series were smoothed using a Savitsky-Golay filter with a span of 13 and a polynomial order of 7. We then used Windowed Cross-Correlation (window size: 125 frames, step of 20 frames) to calculate the correlation between the two speaker's head movements. Correlation across time lags is normalized using Fisher's Transform, and the mean of the absolute per-time-lag correlations is taken as our index of interpersonal synchrony.

Neural synchrony We took a similar windowed correlation approach (25 samples at ~5.08Hz, with a step of 2) to assess interpersonal brain synchrony. We computed the windowed correlations for brain regions defined by both anatomical structures (e.g. mPFC, left TPJ, right lPFC) and the brain networks (e.g. DMN, FPN, VAN). The correlation time series were smoothed with a Savitzky-Golay with a span of 21 and a polynomial order of 3.

Since fNIRS detects BOLD signals, which typically have a delayed response of about 5 seconds relative to the cognitive processes that recruit oxygen resources needed for neuron firing (Tesler, Linne, & Destexhe, 2023), we shifted the fNIRS synchrony measures 5 seconds back in time. In this way, we capture a more real-time measure of interpersonal brain dynamics.

[1] https://envisionbox.org/embedded_dynamicvisualizer.html

Results: Quali-Quanti Exemplification

We performed a qualitative case study on case 3120 to demonstrate how the DIMS Dashboard may be used as a qualitative-quantitative research integration tool. From social neuroscience literature, we know that the temporal parietal juncture (TPJ) is a brain region that is implicated in social functions (Krall et al., 2015; Lieberman, 2022). Motivated by TPJ's role in social interaction, we focused on neural synchrony at the right TPJ (Fig. 2), with an attempt to qualitatively examine the social interaction moments that take place during certain rTPJ activation patterns.

Upon visual inspection of the time series in Fig. 2, we noticed that this neural synchrony time series fluctuates, whereby participants' neural activities dynamically align and diverge with each other across time. Specifically, we noticed many sharp rises (such as the one highlighted in grey towards the right) of different lengths. To provide sufficient information for qualitative analysis in context, we selected 10-second-long rises, and found 10 segments appropriate for qualitative analysis. An example selection 10-second window is colored in grey in Fig. 2.

Among these 10 segments, we found that the sharply increased pattern of neural synchrony in right TPJ corresponds to highly engaged segments (i.e., qualitatively-observed engagement in the interaction) in video recording in 8 segments, which exhibited verbal or nonverbal signals that suggest such engagement. Meanwhile, we were surprised to observe 2 segments with a different type of activity, prompt switching, that invoked the same neural activation pattern.

Firstly, most 10-second segments with sharply increased neural synchrony in right TPJ corresponded with high engagement in the interaction, as observed from the audiovisual recordings. The high engagement was exhibited in both verbal and nonverbal forms: Verbal engagements included expressing agreement by giving additional examples to support the other person's point (e.g. 1075.4s - 1085.4s), repeating each others' exact words (e.g. 812.6s - 822.6s: "*...definitely different*", "*Definitely different*"; "*I guess there's more weekdays.*" "*Okay so weekdays.*"), performing overlapping speech during storytelling with assessments (e.g. 330.2s - 340.2s "*That sounds cool.*"; 504.2s - 514.2s "*That's so early.*" and "*That's still early.*") or rhetorical question (e.g. 1107.4s - 1117.4s: "*Why would I be ... drinking water?*"), and laughing together (e.g. 763.0s - 773.0s). Nonverbal engagements included frequent head nods (e.g. 113.0s - 123.0s: more than 10 nods; 952.2s - 962.2s: roughly 8 nods) and moving their bodies in a similar frequency (e.g. 763.0s - 773.0s).

Secondly, we observed that 2 segments with sharply increased neural synchrony in right TPJ correspond with prompt transitions from one discussion topic to the next (45.4s - 55.4s; 647.4s - 657.4s). Both segments start with one participant finishing up the last words for the previous prompt, then clicking on and reading the next prompt, and one participant beginning to answer this prompt.

In sum, in this short demonstration of using the DIMS Dashboard, we found that 80% of segments with sharply increased neural activations correspond with highly engaged interaction, and 20% of such segments corresponded with prompt transitions. Our first observation provides a face validity check for social neuroscience literature (Redcay & Schilbach, 2019), while the second observation potentially inspires new research direction to create sub-categories of neural synchrony.

Quali-quanti researchers might then generalize from the qualitative observations. Namely, if sharp rises in right TPJ synchrony align with moments of high mutual engagement, we could hypothesize that such increases are carried by bodily synchrony. In other words, we predict that the degree to which right TPJ rises sharply relates to interpersonal bodily synchrony across the two conversations.

To test this, we examined the moments where there are sharp rises in right TPJ synchrony (peaks in the 1st time derivative of right TPJ, with a minimal peak threshold of 0.2 * SD), and regress for those salient moments the slope of that rise in the right TPJ against the head and bodily synchrony value at those moments (Fig.3). While bodily synchrony did not relate to the rise in right TPJ (r = .0349, *p* = .550), we found higher head synchrony values related to higher magnitude of the right TPJ rise (r = 0.131, *p* = 0.024) (Fig. 4). Thus, rises in the right TPJ seem to quantitatively relate to moments of high head motion synchrony, which we qualitatively related to moments of high multimodal mutual engagement.

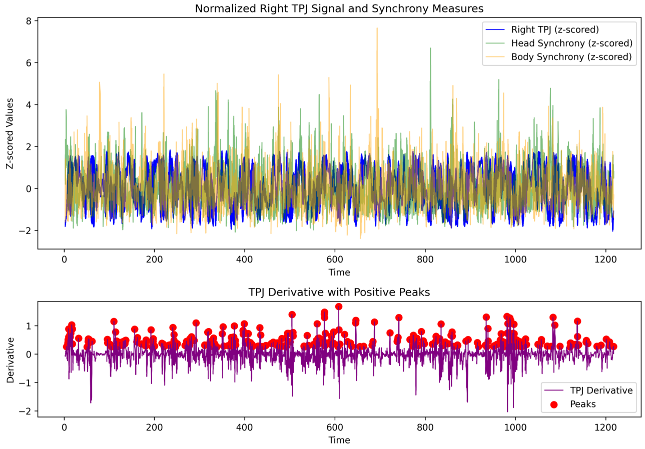


Figure 3. Salient moments in the right TPJ with steep rises in the values obtained by isolating positive peaks (peak height threshold, 0.2 * SD) in the 1st time derivative of TPJ. The magnitude of these peaks is regressed against concurrently observed values of head and bodily synchrony.

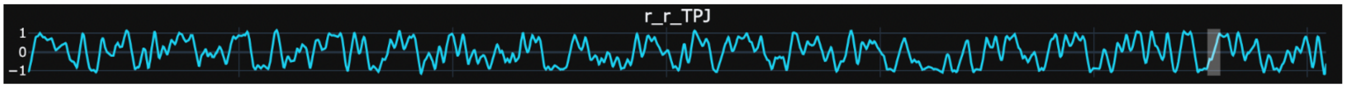


Figure 2. Time series screenshot of neural synchrony at the right temporoparietal junction (r_r_TPJ).

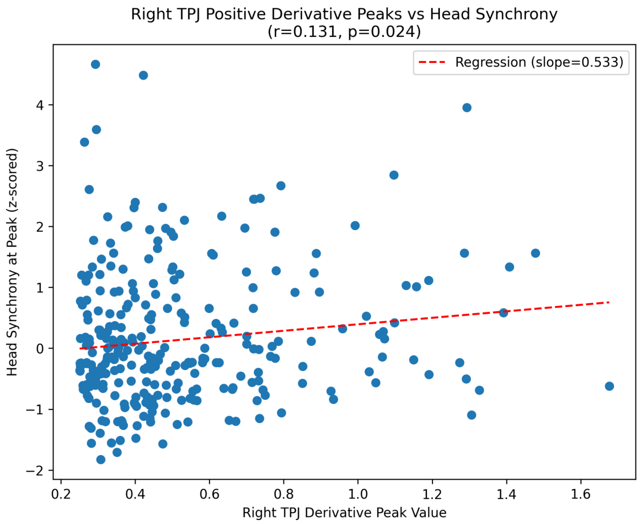


Figure 4. Relationship between the magnitude of the rise in Right TPJ synchrony and the magnitude of head synchrony, at moments where there are steep positive rises in the Right TPJ synchrony (positive peaks at threshold = 0.2 * *SD*).

Discussion

We introduced the potential of dynamic data dashboards to supplement the research toolkit of qualitative and quantitative researchers alike. Dynamic data dashboards are interfaces that connect raw and rich audiovisual data sources with the derived and summarized data and analysis. Here we specifically focused on connecting the raw audiovisual data with multimodal bodily and neural synchrony measures. As a proof-of-concept that dynamic data dashboards can productively kick-start a quali-quanti research cycle, we qualitatively observed that at moments of sharp increases in neural synchrony in the right temporal parietal juncture, there is often a marked moment of mutual multimodal engagement. These qualitative observations can then prompt attempts for further generalization, wherein we predicted that quantitatively moments of higher bodily motion synchrony would predict the magnitude of rises in TPJ synchrony. Indeed, head motion synchrony over the entire two conversations seemed to be related to rise in the TPJ synchrony during salient high-synchronous moments.

At this phase, the research remains entirely in the exploratory rather than confirmatory mode of inquiry (Wagenmakers et al., 2012), aiming to begin defining and stabilizing the potential phenomenon of interest. The quali-quanti cycle should therefore not end here, because relying solely on this approach is statistically prone to generating false-positive inferences and requires more statistical scrutiny (e.g., assumption checking), given that we are deliberately examining the data before testing patterns on a larger scale.

Indeed, we argue that this stage should mark the beginning of a more iterative quali-quanti cycle. For example, the current findings could further invite characterizations of what exactly prompts bodily synchronies, or what precedes and follows—speech, brain activity, or bodily synchrony itself? Or how are these neural-bodily patterns embedded within the broader context of the conversation? Answering these questions with the DIMS Dashboard contributes to a more thoroughly developed understanding of the phenomenon, integrating multimodal processes during social interactions.

As such, rather than a purely data-driven approach, the quali-quanti research we envision is a *phenomenon*-driven approach that can serve, for example, as a hypothesis-generating engine. This approach aims to bridge the micro-scale (qualitative) with a more meso-scale level of description of social interaction. Depending on the goals of the researcher, this cycle can lead to a fully confirmatory research project, where hypotheses that have been refined and operationalized from the quali-quanti cycle are tested in a generalizable way over a much larger dataset—one that may not be meaningfully analyzed at the micro level. Alternatively, the research project may evolve into a deeper qualitative inquiry to investigate cases that clearly do not conform to the broader patterns observed. In either case, we believe this iterative process allows researchers to develop a more nuanced and comprehensive understanding of the phenomenon of interest.

The benefits of data visualizers like the DIMS Dashboard extend beyond the ease and efficiency with which researchers can combine original with derived data. A key advantage is that dashboards like these can be custom designed for specific research projects. Thus, data dashboards like DIMS go beyond traditional annotation tools that offer limited time series visualization with video (e.g., ELAN; Wittenburg et al., 2006) by enabling flexible, dynamic visualizations that draw from the full range of analytic and visualization techniques in modern cognitive science. Unlike ELAN, which provides relatively fixed visualization formats and requires users to work within a specialized, less widely adopted scripting environment, DIMS is built on open-source, Python-based infrastructure—making it easier to adapt and extend for diverse needs. Most importantly, the current approach allows for flexibility of matching primary data with a variety of data visualization techniques, and we are working on future expansions through other modular plugins, such as recurrence plots and cross-wavelet spectral change plots. This accessibility facilitates broader interdisciplinary collaboration and literacy as it allows researchers to explore a wide range of patterns within and across various levels of measurement. In general, connecting different measurement scales enables a more comprehensive mapping of the dynamic systems underlying social interactions.

Fostering a Quali-Quanti Integration

The integration of qualitative and quantitative approaches benefits researchers on both sides. Quantitative researchers can learn to explore their data at a more micro level and generate testable hypotheses based on their observations. Having a visual overview of data contextualized with the original video also allows researchers to check assumptions and can help with the interpretation of complex analyses, which might otherwise become abstracted away from the original behaviors being studied.

A good example of this abstraction pitfall is the concept of synchrony itself. The precise interpersonal or cognitive functions of bodily and/or neural synchrony have been widely debated (Dale et al., 2013; Schilbach & Redcay, 2024). Reported synchrony measures are often based on averages generated over time across experimental conditions (e.g., Hale et al., 2020; Lin et al., 2023; Shockley et al., 2003). Seldom do (we), quantitative researchers, then also provide a deeper qualitative analysis of what actually happens during the highly synchronous moments that have boosted their averages. This is a process of abstraction, where researchers risk becoming too comfortable with the outcome metrics while neglecting the phenomenon for which those metrics are a proxy (i.e., confusing the map from the territory).

In line with philosophies of science which emphasize action and experimentation over passive observation and pure theorizing, technologies like microscopes require skilled wielding to stabilize and reproduce a certain phenomenon (Hacking, 1983). Similarly, we argue that innovations such as synchrony measurements in cognitive science require researchers to develop expertise not only in using these tools but also in understanding the nature of the measurement apparatus—whether for neural or bodily synchrony. As such, we believe that quantitative researchers can become more observant rather than mere observers with visualization tools like the DIMS dashboard: "*observation, in the philosophy-of-science usage of the term, plays a relatively small role in experimental science… Another kind of observation is what counts: the uncanny ability to pick out what is odd, wrong, instructive or distorted in the antics of one's equipment. The experimenter is not the ''observer'' of traditional philosophy of science, but rather the alert and observant person.* (Hacking, 1983, p. 230)".

A current limitation of the DIMS Dashboard is that, while it supports dataset-specific customization, it lacks a flexible interface for easily integrating new datasets. Addressing this limitation is a key focus of our next development phase, as we aim to build a more user-friendly and modular system that can accommodate a wider range of datasets with minimal configuration.

A critical caveat to the widespread adoption of dynamic data visualizers like DIMS Dashboard should also be considered. Researchers sometimes compute deliberately derived measures that do not necessarily have an immediate or directly perceivable link to the interactions in situ. Indeed, some measures, such as complexity matching or multidimensional cross-recurrence analysis, are specifically designed to capture dynamics that unfold over longer timescales. It would be a categorical mistake to seek "burstiness" or "complexity" within a conversation in a short bout of interaction (Abney et al., 2014). Future updates to tools like DIMS could offer more detailed guidance and interface options to help users interpret long-timescale metrics in context. These improvements would support more accurate and theory-driven use of multimodal data.

Nevertheless, we argue that by bridging measurements with a phenomenon, qualitative researchers will generate better intuitions for the level of description their theorized measurements are meant to capture. Similarly, qualitative researchers may benefit from applying their qualitative analysis to computationally derived time series. By making computational measures more accessible, both in terms of their generation and their contextualization within the original video data, qualitative researchers can broaden the depth and scope of data at their disposal.

Concluding Remarks

While research traditions have often separated qualitative and quantitative researchers, we believe that innovations in research reporting like this DIMS Dashboard may help promote a truly integrated quali-quanti approach. Quantitative measurements enable abstraction from the here and now, allowing researchers to describe systems within a much higher dimensional space. Some of the stable patterns emerged in this higher dimensional space—when identified through quantitative observations—may manifest as highly recognizable moments in social interaction. For example, we may find that when high brain desynchronization occurs during moments with high bodily synchrony, these instances correspond to communication failures that interaction partners attempt to avoid. A quali-quanti research team, equipped with the right quantitative tools to detect statistically reliable high dimensional recurrences (e.g. low brain synchrony paired with high bodily synchrony) and the right qualitative skills, may uncover that bodily regulation during cognitive opposition represents a particular pragmatic mode that operates along a body-brain-conversational axis (Alviar et al., 2023). Such pragmatic modes are difficult to identify. Just as how software like PRAAT (Boersma, 2007) has helped qualitative researchers interpret high-dimensional space of speech acoustics in terms of pragmatic functions, we believe that more integrative tools like this DIMS Dashboard are needed to achieve the same for dynamic, multimodal, multidimensional interactions.

Open Data and Code

Code for building a basic dynamic visualization dashboard is provided here: https://envisionbox.org/embedded_dynamicvisualizer.html. Since we are still fine tuning, we will share the complete code supporting the DIMS Dashboard along with full journal publication of the project.

Acknowledgements and Funding

The DIMS Dashboard emerged from the Dynamic Interactions and Methodologies Symposium at UCLA (https://gracemiao.com/dims/) funded by the Summit View Research Foundation. GM & RD were supported in part by NSF grant BCS-2120932 to RD. WP is funded by NWO VENI grant (VI.Veni 0.201G.047), a DFG VICOM grant (PO 2841/1-1), and a NWO XS grant (406.XS.24.01.043).

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General, 143*(6), 2304.

Alviar, C., Kello, C. T., & Dale, R. (2023). Multimodal coordination and pragmatic modes in conversation. *Language Sciences, 97*, 101524.

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). WhisperX: Time-accurate speech transcription of long-form audio. In *arXiv [cs.SD]. arXiv.* http://arxiv.org/abs/2303.00747

Boersma, P. (2007). Praat: doing phonetics by computer. http://www. praat. org/.

Cheong, J. H., Molani, Z., Sadhukha, S., & Chang, L. J. (2023). Synchronized affect in shared experiences strengthens social connection. *Communications Biology, 6*(1), 1099.

Clark, H. H. (1996). *Using language.* Cambridge University Press.

Cooney, G., & Wheatley, T. (in press). On conversation. In D. T. Gilbert, S. Fiske, E. Finkel, & W. Mendes (Eds.), *Handbook of social psychology* (6th ed.).

Dale, R., Fusaroli, R., Døjbak Håkonsson, D. D., Healey, P., Mønster, D., McGraw, J., ... & Tylén, K. (2013). Beyond synchrony: Complementarity and asynchrony in joint action. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).

Dale, R., Warlaumont, A. S., & Johnson, K. L. (2023). The fundamental importance of method to theory. *Nature Reviews Psychology, 2*(1), 55-66.

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language*. MIT press.

Giulianelli, M., & Fernández, R. (2021). Analysing human strategies of information transmission as a function of discourse context. In *Proceedings of the 25th Conference on Computational Natural Language Learning* (pp. 647-660).

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science.* Cambridge university press.

Hale, J., Ward, J. A., Buccheri, F., Oliver, D., & Hamilton, A. F. D. C. (2020). Are you on my wavelength? Interpersonal coordination in dyadic conversations. *Journal of nonverbal behavior*, *44*, 63-83.

Heritage, J. (2012). Epistemics in action: Action formation and territories of knowledge. *Research on Language & Social Interaction, 45*(1), 1-29.

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science, 321*(5890), 806–810.

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., ... & Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function, 220*, 587-604.

Lieberman, M. D. (2022). Seeing minds, matter, and meaning: The CEEing model of pre-reflective subjective construal. *Psychological Review, 129*(4), 830.

Lin, L., Feldman, M. J., Tudder, A., Gresham, A. M., Peters, B. J., & Dodell-Feder, D. (2023). Friends in sync? Examining the relationship between the degree of nonverbal synchrony, friendship satisfaction, and support. *Journal of Nonverbal Behavior, 47*(3), 361–384.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172.*

Miao, G. Q., Dale, R., & Galati, A. (2023). (Mis) align: a simple dynamic framework for modeling interpersonal coordination. *Scientific Reports, 13*(1), 18325.

Miao, G.Q., Jiang, Y.J., Binnquist, A., Pluta, A., Steen, F.F., Dale, R., & Lieberman, M.D. (2024). A Deep Neural Network Approach for Integrating Neural and Behavioral Signals: Multimodal Investigation with fNIRS Hyperscanning and Facial Expressions. In L. K. Samuelson, S. Frank, M. Toneva, A. Mackey & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Meeting of the Cognitive Science Society* (pp. 5630-5638). Austin, TX: Cognitive Science Society.

Moreno, P. J., Joerg, C., Van Thong, J.-M., & Glickman, O. (1998). A recursive algorithm for the forced alignment of very long audio segments. *5th International Conference on Spoken Language Processing (ICSLP 1998).*

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications, 9*(1), 332.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences, 1464*(1), 5-29.

Pinti, P., Aichelburg, C., Gilbert, S., Hamilton, A., Hirsch, J., Burgess, P., & Tachtsidis, I. (2018). A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. *Japanese Psychological Research*, *60*(4), 347-373.

Pearson, L., Nuttall, T., & Pouw, W. (2024). *Landscapes of coarticulation: The co-structuring of gesture-vocal dynamics in Karnatak vocal performance.* OSF Preprints.

Pouw, W. (2024). Wim Pouw's EnvisionBOX modules for social signal processing (Version 1.0.0) [Computer software]. https://github.com/WimPouw/envisionBOX_modulesWP

Pouw, W., & Dixon, J. A. (2020). Gesture networks: Introducing dynamic time warping and network analysis for the kinematic study of gesture ensembles. *Discourse Processes*, *57*(4), 301-319.

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *International conference on machine learning* (pp. 28492-28518). PMLR.

Ravi, N., Gabeur, V., Hu, Y. T., Hu, R., Ryali, C., Ma, T., ... & Feichtenhofer, C. (2024). Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience, 20*(8), 495-505.

Rosen, Z. P., & Dale, R. (2024). BERTs of a feather: Studying inter-and intra-group communication via information theory and language models. *Behavior Research Methods, 56*(4), 3140-3160.

Schilbach, L., & Redcay, E. (2024). Synchrony across brains. *Annual Review of Psychology, 76*.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience1. *Behavioral and brain sciences, 36*(4), 393-414.

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326.

Sidnell, J., & Stivers, T. (Eds.). (2012). *The handbook of conversation analysis*. John Wiley & Sons.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*(26), 10587-10592.

Tesler, F., Linne, ML. & Destexhe, A. Modeling the relationship between neuronal activity and the BOLD signal: contributions from astrocyte calcium dynamics. *Scientific Reports, 13*, 6451 (2023).

Thibault, P. J. (2020). *Distributed languaging, affective dynamics, and the human ecology volume I: The sense-making body*. Routledge.

Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes*, *54*(5-6), 382-405.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632-638.

Wheatley, T., Thornton, M. A., Stolk, A., & Chang, L. J. (2024). The emerging science of interacting minds. *Perspectives on Psychological Science, 19*(2), 355-373.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *5th international conference on language resources and evaluation (LREC 2006)* (pp. 1556-1559)

Wohltjen, S., & Wheatley, T. (2021). Eye contact marks the rise and fall of shared attention in conversation. *Proceedings of the National Academy of Sciences, USA, 118*(37), Article e2106645118.

Yazdian, P. J., Chen, M., & Lim, A. (2022). Gesture2Vec: Clustering gestures using representation learning methods for co-speech gesture generation. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 3100-3107). IEEE.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106*, 1125–1165.