Model of Coordination Flow in Remote Collaborative Interaction

Dari Trendafilov[1,2], Daniel Polani[1]
[1]*School of Computer Science*
University of Hertfordshire
Hatfield, UK
{d.trendafilov, d.polani}@herts.ac.uk

Roderick Murray-Smith[2]
[2]*School of Computing Science*
University of Glasgow
Glasgow, Scotland, UK
roderick.murray-smith@glasgow.ac.uk

We present an information-theoretic approach for modelling coordination in human-human interaction and measuring *coordination flows* in a remote collaborative tracking task. Building on Shannon's mutual information, *coordination flow* measures, for stochastic collaborative systems, how much influence, the environment has on the joint control of collaborating parties. We demonstrate the application of the approach on interactive human data recorded in a user study and reveal the amount of effort required for creating rigorous models. Our initial results suggest the potential *coordination flow* has as an objective, task-independent measure for supporting designers of human collaborative systems and for providing better theoretical foundations for the science of Human-Computer Interaction.

Index terms— Coordination, Collaboration, Uncertainty, Human Factors, Information theory.

1 Introduction

A fluid engaging collaboration between people connected remotely via a computer has long been a goal of technology mediated interaction. Modern hardware has the sensing, processing and feedback potential for a more exciting range of high-bandwidth, tightly-coupled human–human interaction styles. However most current systems fall short of that, with stilted, discrete exchanges of information the norm.

Cooperation in the real world emerges as a distinct combination of innate and learned behaviour according Tomasello [18], of which a key element is the use of language, as the joint action of speakers and listeners performed in ensemble, that embodies both individual and social processes [3]. To identify the underlying psychological processes supporting human collaboration and understand how humans perceive, intend, learn, control, and coordinate complex behaviours we need a general framework connecting brain, mind and behaviour, and extending the physical concepts of self-organization [5]. Recent advances in the development of cognition and action, unifying dynamic systems theory with neuroscience, show how by processes of exploration and selection, multimodal experiences form the bases for self-organizing perception-action categories [16]. Human–human interaction is a high bandwidth, multimodal and highly complex process, but this is not characteristic for technology mediated solutions developed by researchers interested in the exploration of such systems. To design better interactive systems, we argue that, we need to draw on sound principles and formal models. However, as argued in [17], this has often been lacking in the HCI community, in part due to the perceived gap between the complexity of human behaviour and our ability to capture and model it.

One interesting current challenge is the development of a formal measure, quantifying the level of coordination between participants of computer-mediated environments. In this paper we adopt an information-theoretic approach, and explore the potential of mutual information as an objective measure, reflecting how much influence the environment has on users' joint control in the course of interaction. A rigorous measure of *coordination flow*, could provide an analytical tool, revealing the trends and the gradients in interactive models, and could give direct insight into the underlying properties and provide confidence regions for the system's parameters. It could help provide a firm foundation for designers to treat and evaluate human–human interactive systems in a general fashion.

2 Background

Mediated environments utilise a range of digital devices, connected in networks, which creates various sources of disturbances affecting the quality of interaction. The key contributor to uncertainty is delay in the feedback loop. Lag is recognised as a major bottleneck for usability in human–computer interaction [9]. In human–human mediated interaction, however, the challenge is even more pronounced, as we have to account also for the variable response time of a human decision maker, which – unlike machines – varies across individuals and depends on many internal and external factors, making it highly unpredictable. Other sources of disturbances are different types of noise – digital sensor imprecision, human sensorimotor inaccuracy, transmission noise, etc. The quality of control depends on feedback that must reflect the uncertainty of system beliefs. Interfaces should work with the uncertainty, not just filter it out [11], as appropriate use of uncertain feedback could regularise user behaviour and lead to smoother interaction [8, 24]. Uncertainty poses challenges for designers when evaluating modern mobile interactive systems and raises the need for theoretical frameworks for modelling and inference as vital aspects of system analysis [20].

Information theory provides important quantities for the characterization of systems in the physical world [23, 2]. The use of Shannon's mutual information is ubiquitous in this context. A particular interest lies in the identification of the "flow of information" in a given system, for which typically variants of mutual information measures of correlative character are used, where the joint information stems from a common past [10, 15]. Recent work shows the utility of having a measure for a "flow of information" [22, 1, 7, 6]. Building on work of Pearl [13], Ay and Polani [2] introduced a concept of information flow for discrete worlds, formalised on causal Bayesian networks. Schreiber introduced "transfer entropy" to quantify information transfer in Markov chains [15].

Sato and Ay [14] explored scenarios with a dynamic component, in which players adapt their strategies over time in order to achieve desired cooperative behaviour. Ay and Polani [2] suggest that "information flow", with its causal character, could measure a player's contribution for the emergence of a particular cooperative strategy. Recent predictive Bayesian concepts of sensorimotor control and low-level decision making increasingly gain momentum [8].

Galantucci explored the complexity of human behaviour in the absence of pre-established human communication systems in order to elucidate how these systems emerge and develop in the context of joint human activities [4]. A study on the interaction of motion and conversational behaviour, show the emergence of spontaneous synchronisation in walking patterns during mobile phone conversations, and suggest the benefits of a gait alignment measure [12].

In recent studies we investigate embodied remote human collaboration, exploring the emergence of cooperative strategies using limited modes of communication [19, 21]. Building on this work, here we propose a model and a measure to quantify coordination in data collected earlier.

To characterise coordination we explore the notion of conditional mutual information[1], which is defined for random variables X_A and X_B given X_S as follows,

$$I_p(X_A : X_B|X_S) = \sum_{x_S} p(x_S) I_p(X_A : X_B|x_S), \qquad (1)$$

where

$$I_p(X_A : X_B|x_S) = \\ = \sum_{x_A} p(x_A|x_S) \sum_{x_B} p(x_B|x_A, x_S) \log \frac{p(x_B|x_A, x_S)}{p(x_B|x_S)}.$$

3 Model

3.1 Experiment

In our earlier study [19], participants performed a simple collaborative target acquisition task in pairs via shared mediated environment, while sitting in separate rooms. The only available mode of communication was scrolling a finger on a touch-sensitive digital device. Achieving a good performance, in terms of number of targets acquired, depended on participants cooperation, which in turn required a certain level of coordination. The aim was to explore the strategies executed by different pairs and to get an insight into the level of coordination that was achieved using the imposed minimalistic mode of communication.

The interaction concept consists of two subjects simultaneously exploring a virtual membrane from their respective side, trying to find a hole and touch each other. The feedback mechanism allows users to sense each other whenever their fingers meet on the shared membrane and to sense the holes in their side of the membrane. The membrane is shown in a section as a vertical gray strip (Figure 1). A bell-shaped marker represents the finger position and a black square – a hole. Using the input device the user can probe the membrane up and down and search for holes and for the remote partner. Holes and the remote partner can only be seen in their close proximity (Figure 1), otherwise they are hidden. The user can obtain information only by sensing for impact events, i.e. whenever their pointer collides with objects

[1] Populating the densities of the model with the right content is based on the assumption of how we think people coordinate, and compensates for various delays characteristic for human behaviour.

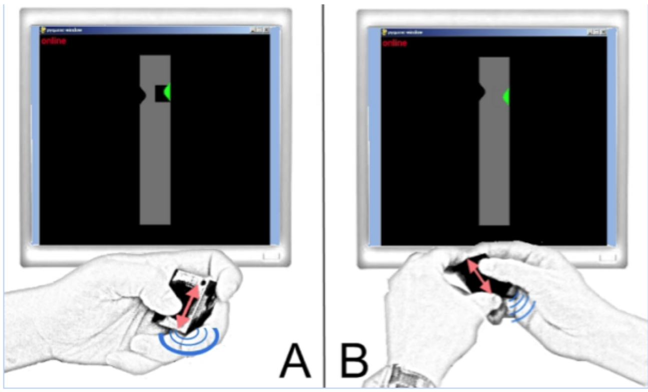

Figure 1: Player A (in green on his display) has found a hole, shown as black square, and his partner (in black) on the other side of the membrane. Player B (in green on his display) can only see his partner (in black), as there is no hole in the vicinity on his side.

in the shared environment. Each side of the membrane has three holes, one of which is shared. The task requires users' active exploration of the membrane in order to locate and acquire simultaneously the shared hole. Each experimental session lasted five minutes and user input was recorded.

3.2 Dyad Perception-Action Loop

To formalize our experimental model we use the causal Bayesian network representation of the perception-action loop – section of which is shown in Figure 2 – unrolled over time. This network specifies the causal relationships between both users sensor states (visual stimuli s^a and s^b) which are influenced by their actions (finger movements a and b) through the environment (virtual membrane R). Both players observe their current position, keep an estimate of the distance from their partner (which we call the error), and make decisions of choosing among three distinct actions – stay in the current position or move up/down.

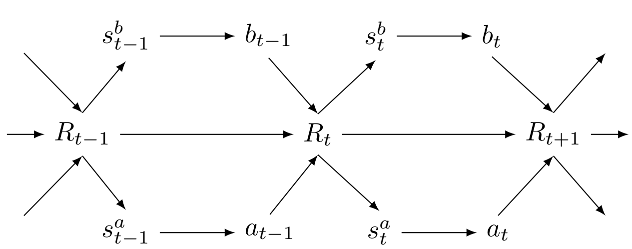

Figure 2: Section of the Bayesian network representing the perception-action loop of a dyad $(A - B)$ interacting through the environment (R) by applying actions (a and b) in response to sensor stimuli (s^a and s^b).

3.3 Stochastic Model

To analyse the level of coordination in the collected experimental data we use Equation 1. In order to apply that measure, however, we had to discretise the continuous set of actions. Furthermore, we had to ensure that the limited amount of data will be sufficient for a good approximation of the conditional probabilities. For that purpose we defined low resolution action and state spaces of three elements each, corresponding to the sign functions below, which could potentially reveal the underlying trends. Higher resolution spaces would require larger amounts of data to provide a reliable empirical density, otherwise data sparsity could bias the results. For simplicity, we assume that subjects actions are influenced only by the error (distance between them), denoted with the random variable X_S, and the direction of their motion, denoted with the random variables X_A and X_B, respectively.

We define the random variables, corresponding to the action and the error states, with the sign functions in Equations 2, 3 and 4. The operators $\approx$, $\ll$ and $\gg$ reflect the close proximity range, used in the experiment, within which the players can see each other, and refer to relations with a distance threshold of 20 pixels. Action velocity is computed from positions at the end points of a sliding window of 2.5 seconds. Using this simplified model our aim was to capture the relationship between the three random variables. To estimate the joint distribution of X_A, X_B and X_S ($p(x_A, x_B, x_S)$) we count the occurrences of the joint events in the data set. From $p(x_A, x_B, x_S)$ we derive the marginal densities $p(x_A)$, $p(x_B)$ and $p(x_S)$, and after applying the chain rule we obtain the conditional densities $p(x_A|x_S)$, $p(x_B|x_S)$ and $p(x_B|x_A, x_S)$. That is all we need to apply Equation 1.

$$X_S = \begin{cases} -1, & X_A \gg X_B \\ 0, & X_A \approx X_B \\ 1, & X_A \ll X_B \end{cases} \tag{2}$$

$$X_A = \begin{cases} -1, & X_A^t \gg X_A^{t+1} \\ 0, & X_A^t \approx X_A^{t+1} \\ 1, & X_A^t \ll X_A^{t+1} \end{cases} \tag{3}$$

$$X_B = \begin{cases} -1, & X_B^t \gg X_B^{t+1} \\ 0, & X_B^t \approx X_B^{t+1} \\ 1, & X_B^t \ll X_B^{t+1} \end{cases} \tag{4}$$

4 Results

4.1 Empirical densities

Time series recorded in our earlier study (Figure 4) suggest that the interaction consisted of a series of discrete messages. This poses a challenge for our stochastic model to infer from the raw data set. To cancel the effect of such discrete patterns we applied a moving average filter using a sliding window of 2 seconds to smooth the data (filtered data set). In addition, we applied a simple dynamic time warping algorithm, using a sliding window of 2.5 seconds, compensating for reaction time delay (delayed data set). The value of maximum delay, characteristic for close tracking performance, was inferred over all data.

From the recorded experimental data, consisting of 5000 data points per trial session, and applying the above post-processing methods, we computed three sets (raw, filtered and delayed) of empirical probability density functions ($p(x_A, x_B, x_S)$), using the proposed stochastic model.

4.2 Coordination Flow

Following Equation 1, we calculated the conditional mutual information on the three empirical densities Figure 3. The 'raw' set provided low values of mutual information, as expected, due to various types of noise and delays, diminishing the correlations visible in Figure 4. Smoothing data provided increased levels of mutual information in the 'filtered' set. However, it turns out that the key factor preventing our model from capturing the correlation in the time series is the inherent delay associated with human motor control and decision making. High sampling rates used for data collection further magnify this effect. Delay compensation resulted in a further increase in mutual information in the 'delayed' set.

The results show a clear correspondence of *coordination flow* values (Figure 3) to the three characteristic types of tracking behaviour shown in Figure 4. As can be seen, pair $P1$ shows an example of performing a very tight tracking throughout the trial session, which yields 0.98 bits of *coordination flow*. Pair $P2$, on the other hand, demonstrates a different pattern, yielding *coordination flow* of 0.44 bits, which is a considerable drop, due to the loose tracking and somewhat erratic behaviour, with longer and irregular delays. On the other extreme, Pair $P3$ achieved the lowest value of *coordination flow* (0.24 bits), as it applied a random strategy exhibiting very little tracking behaviour.

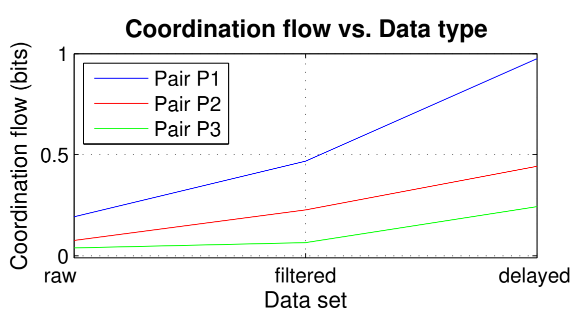

Figure 3: Levels of coordination flow, computed on P1, P2 and P3 (a) raw data; (b) moving average filtered data; and (c) delay-compensated filtered data.

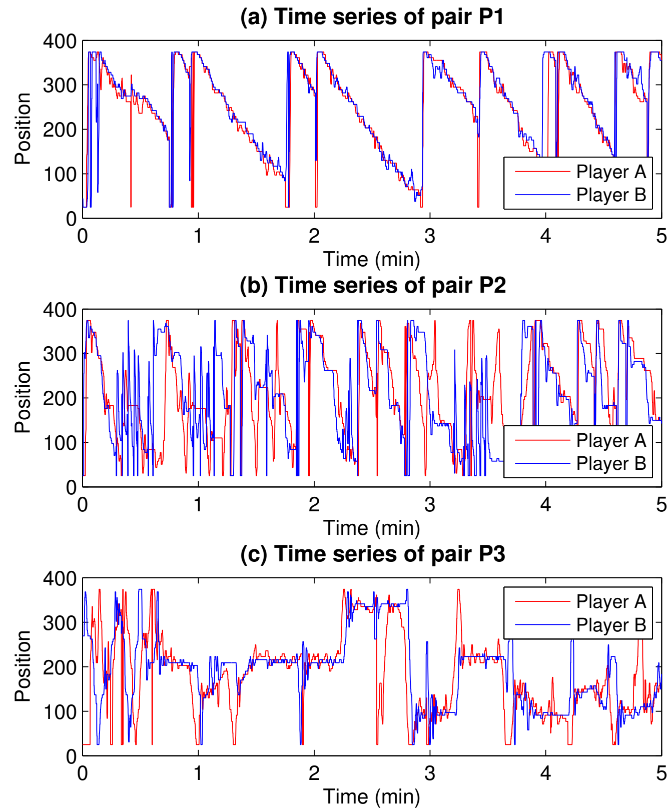

Figure 4: Time series of experimental sessions presenting examples of three distinct strategies. (a) P1 – tight tracking; (b) P2 – loose tracking and (c) P3 – random strategy.

These initial results show the potential of the utilised measure to capture the level of coordination in human tracking data. They also raise important questions associated with the application of the measure, related to the sensitivity to data smoothing and modelling of delays.

4.3 Simulations

To explore the properties of the proposed measure and investigate its behaviour at the extremes of our stochastic model we performed series of simulations, using the three empirical $P1 - P3$ (provided by the delayed set) and five extreme densities $P4 - P8$, defined on the same set of random variables (X_A, X_B, X_S). $P5$ corresponds to tightly-coupled controllers, $P6 - P8$ to different types of non-coordinated behaviour, and $P4$ is a mixture of the above. Using these 8 models, we define a basis in (x, y) and generate 10000 alternating models in a two dimensional grid of 100×100 resolution, with the following linear interpolation. Given four joint densities $q1$, $q2$, $q3$ and $q4$ over (X_A, X_B, X_S), we define a new density $q(x_A, x_B, x_S)$ following Equation 7, with corresponding value of *coordination flow* presented in Figure 5.

$$r1 = x * q1 + (1 - x) * q2, \qquad x \in [0, 1] \qquad (5)$$

$$r2 = x * q3 + (1 - x) * q4, \qquad x \in [0, 1] \qquad (6)$$

$$q = y * r1 + (1 - y) * r2, \qquad y \in [0, 1] \qquad (7)$$

In order to get further insights into the sensitivity of *coordination flow* to changes in the underlying distributions we computed the Jensen-Shannon divergences[2] of the three empirical $(P1 - P3)$ vs. all densities across the grid. In Figure 6 the red lines connect points of zero divergence on (b), (c) and (d) respectively, to the corresponding empirical densities $P1$, $P2$ and $P3$ in the *coordination flow* plot (a). This perspective could provide additional help in visualising the trends on the *coordination flow* curve.

[2] $JSD(P\|Q) = (D(P\|M) + D(Q\|M))/2$, where $M = (P+Q)/2$, $D(P\|Q) = \sum_x P(x) log(P(x)/Q(x))$ is Kullback-Leibler divergence.

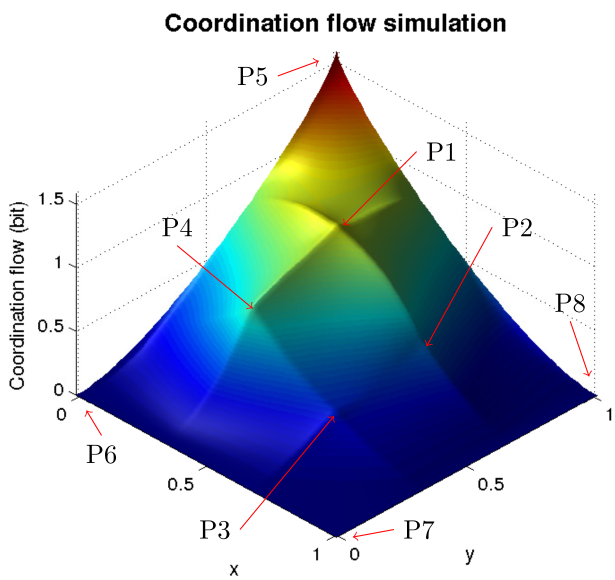

Figure 5: *Coordination flow* simulation using linear interpolation on the three empirical $(P1 - P3)$ and five extreme $(P4 - P8)$ densities.

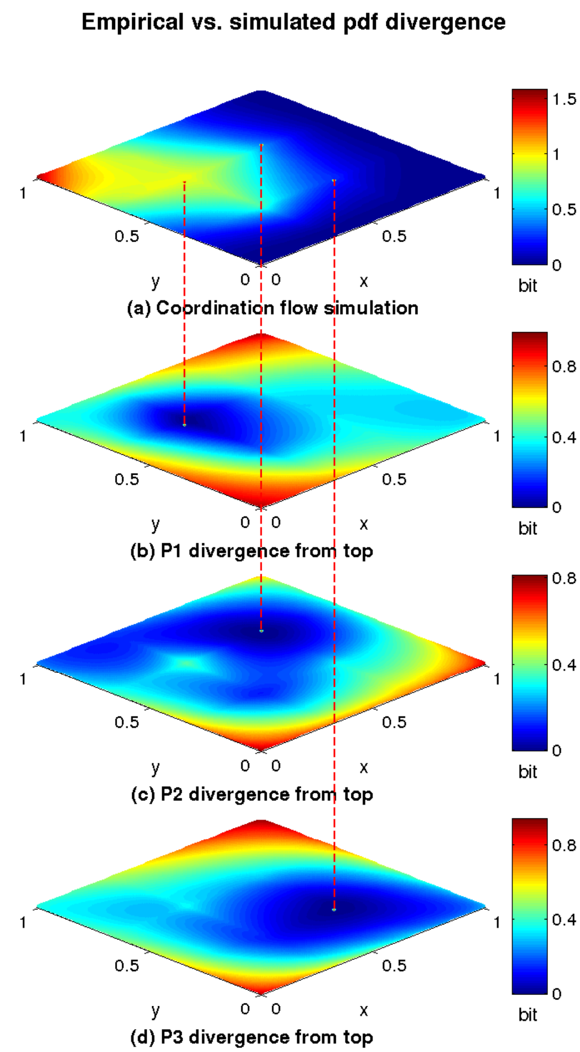

Figure 6: Jensen–Shannon divergence of the three empirical $(P1 - P3)$ vs. all densities used in the simulation in Figure 5. (a) *Coordination flow* (as in Figure 5); (b), (c) and (d) J–S divergence between $P1, P2$ and $P3$ respectively and the corresponding densities in (a). The red lines connect $P1, P2$ and $P3$ from (a) with the corresponding points of minimum (J–S=0 bit) divergence on (b), (c) and (d).

5 Discussion

Results show that levels of *coordination flow* are higher for partners performing a closely engaged tracking, as pair $P1$ (Figure 4a), and drop significantly for pairs who occasionally or more frequently disengage from tracking each other as $P2$ and $P3$ (Figure 4b and 4c). This suggests that the proposed measure could capture salient stochastic properties of the experimental data and could help infer the potential level of coordination. Our results were consistent throughout the data set of 13 pairs, who participated in the experiment, however, for brevity we only present three characteristic types of human behaviour, which help us introduce the approach. Furthermore, we refrained from presenting results related to standard performance metrics, such as success rate (i.e. number of targets acquired) or completion time (i.e. time to acquire a target), as they can be found in [19], and since the focus of the current paper is on the characterisation of tracking behaviour, which is not necessarily directly associated with performance oriented measures. For example, a pair like $P1$, delivering very smooth and consistent tracking performance throughout the session, might be too slow in locating and acquiring targets, as the partners are extremely careful to not lose contact with each other, and thus get a lower total score in the end. Others, like $P2$, might have a strategy of systematic jumping from one target to the next, without being afraid of losing contact with their partner, resulting in a less consistent tracking behaviour, but a faster target acquisition, leading to a higher total score. Others, like $P3$, might have no strategy at all, and jump randomly from one target to another, which is not qualified as tracking behaviour in our analysis, but could still achieve a relatively high total score. In the follow-up questionnaire pair $P1$ admitted having a working strategy of moving together from the top down, which confirms that their tracking behaviour was intentional. In this paper we investigated the *coordination flow* induced by the error to the joint control of both participants. However, another interesting topic for future research is to decouple the dyad and measure the flow of influence between both players. The time series of pair $P2$ (Figure 4b) reveal patterns of leader–follower behaviour and those of $P1$ (Figure 4a) – of turn-taking leadership, where the leader and the follower roles were not clearly defined – the latter resembling more to a smooth dance than the former command-and-control behaviour, an interesting topic for future research.

The relation between empirical data and the corresponding levels of *coordination flow* suggests monotonic and expected trends. This work explored the proposed measure on a particular set of data, however, further work is required to expand and validate the approach in other domains. These results, along with the theoretical coherence of *coordination flow*, suggest its potential as part of a future toolset for understanding interactive systems.

Applying the *coordination flow* measure, however, requires prior theoretical modelling, which, depending on the particular system, may become too costly. There is a trade-off between the accuracy of the theoretical models and the reliability of the *coordination flow* measure – the more accurate the models, the more costly they are to create, but the more reliable the measure they imply.

Furthermore, the quality of data may affect to a great extent the modelling process, as suggested in this initial work. Sampling rates, noise, delays, sensing and feedback resolution all contribute to quality of data and in most real-world cases advanced data smoothing, dynamic time warping and other pre-processing methods might be required prior to applying the *coordination flow* measure. In our prototype system, for example, we used two Bluetooth enabled devices, paired to two laptops, connected over WiFi, and the resulting purely communication delay is added on top of the sensorimotor, decision making, software overhead and other sources of lags. This work focused on developing a low-level perceptual model of tracking behaviour, however, an interesting topic for future research is the design of higher level mental models of coordination. Traces of the latter appear in the time series (Figure 4a, 4b) – when subjects reach the end of the membrane and jump to the other end, interrupting the tracking patterns and resulting in discontinuities. These traces are, however, ignored by our model as artefacts in the filtering process.

6 Conclusion

The present work was motivated by the need for a systematic quantification of coordination in computer-mediated environments. We presented an evaluation of how people collaborate, and proposed a model, applying standard mutual information to quantify coordination, based on purely observational quantities. In developing this model, we desired to capture essential properties of a Shannon-type quantity, measurable in bits. We have shown that, the proposed model and measure capture the correlation in the observed data, which suggests the potential of the approach, in providing an analytical tool to support system designers. The results reveal the amount of effort, required for rigorous modelling – the more accurate the models, the more costly they are to create, but the more reliable the measure they imply. Future experiments, using simulated agents and human users, would give us more control of the activity levels and firmer ground for observing the detailed interactions, that evolve, as people engage and disengage from remote contact with each other.

Acknowledgment

The authors would like to thank Saija Lemmelä, Martin Biehl and acknowledge support by H2020-641321 socSMCs FET Proactive project, Nokia Research Finland and Demola/Suuntaamo Labs Tampere.

References

[1] N. Ay and D. Krakauer. Geometric robustness theory and biological networks. *Theory in biosciences*, 125(2):93–121, 2007.

[2] N. Ay and D. Polani. Information flows in causal networks. *Advances in Complex Systems*, 2008.

[3] H. H. Clark. *Using Language*. Cambridge University Press, 1996.

[4] B. Galantucci. An experimental study of the emergence of human communication systems. *Cognitive Science*, 29(5):737767, 2005.

[5] S. Kelso. *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Complex Adaptive Systems, 1997.

[6] A. Klyubin, D. Polani, and C. Nehaniv. *Empowerment*: A universal agent-centric measure of control. In *Proc. IEEE congress on Evolutionary Computation*, pages 128–135, 2005.

[7] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Organization of the information flow in the perception-action loop of evolved agents. *NASA/DoD Conference on Evolvable Hardware (IEEE Computer Society)*, page 177180, 2004.

[8] K. P. Körding and D. M. Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427:244–247, 2004.

[9] I. S. MacKenzie and C. Ware. Lag as a determinant of human performance in interactive systems. In *INTERCHI*, pages 488–493, 1993.

[10] K. Matsumoto and I. Tsuda. Calculation of information flow rate from mutual information. *J. Phys. A: Math. Gen.*, 21(6), 1988.

[11] R. Murray-Smith. Empowering people rather than connecting them. *International Journal of Mobile HCI*, 1(3), 2009.

[12] R. Murray-Smith, A. Ramsay, S. Garrod, M. Jackson, and B. Musizza. Gait alignment in mobile phone conversations. *Proc. MobileHCI*, pages 214–221, 2007.

[13] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[14] Y. Sato and N. Ay. Adaptive dynamics of interacting markovian processes. *Working Paper 06-12-051 (Santa Fe Institute*, 2006.

[15] T. Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, (85):461–464, 2000.

[16] E. Thelen and L. B. Smith. *A Dynamic Systems Approach to the Development of Cognition and Action*. Bradford Book, 1996.

[17] H. Thimbleby. *Press On Principles of Interaction Programming*. MIT Press, 2007.

[18] M. Tomasello. *Why We Cooperate*. Boston Review Books, 2009.

[19] D. Trendafilov, S. Lemmelä, and R. Murray-Smith. Negotiation models for mobile tactile interaction. *Mobile Social Signal Processing*, pages 64–73, 2014.

[20] D. Trendafilov and R. Murray-Smith. Information-theoretic characterization of uncertainty in manual control. *Proc. IEEE Systems, Man, and Cybernetics*, pages 4913–4918, 2013.

[21] D. Trendafilov, Y. Vazquez-Alvarez, S. Lemmelä, and R. Murray-Smith. Can we work this out?: an evaluation of remote collaborative interaction in a mobile shared environment. *Proc. MobileHCI*, pages 499–502, 2011.

[22] T. Wennekers and N. Ay. Finite state automata resulting from temporal information maximization. *Neural computation*, 17(10):2258–2290, 2005.

[23] J. A. Wheeler. Complexity, entropy and the physics of information. *Santa Fe Studies in the Sciences of Complexity*, page 328, 1990.

[24] D. M. Wolpert and K. P. Körding. Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10(7), 2006.