Doctoral Dissertation

博士論文

Information Theoretical Approaches to Dyadic Human Interactions

（情報理論的アプローチによるヒトの二者間相互作用の解析）

A Dissertation Submitted for the Degree of Doctor of Philosophy

March 2021

令和 3 年 3 月博士 (理学) 申請

Department of Physics, Graduate School of Science,
The University of Tokyo

東京大学大学院理学系研究科
物理学専攻

Hiroki Kojima
小島　大樹

Abstract

It is not yet well understood how we become aware of the presence of other people as being other subjects. In this thesis, we focused on the dyadic interaction between two subjects and aimed to characterize the dynamics which account for the feeling of the other's presence. For this purpose, we adopted two experimental settings, Perceptual Crossing Experiment, and TypeTrace Messenger.

Perceptual crossing experiment was designed as the minimal experimental setup to investigate the dyadic interaction between two subjects. In this experiment, two subjects were asked to identify the other in the 1D virtual space. The dynamics of each trial was fully captured by 4 time series data, which were each agent's movement trajectory, and the time course of the sensory feedback by vibrating device. We characterized these time series data quantitatively using several different measures and compared these values with the perceptual awareness scale to investigate the feature of dynamics that was related to the feeling of other's presence. First, we measured the movement synchronization between two subjects and found that these were related to the perceptual awareness scale, and the trial with the larger movement synchronization was associated with the stronger feeling of other's presence. Next, we used local transfer entropy to quantify directions of influence in the dyadic interaction, and compared this value before and after the click, at which subjects were convinced that they found the other subject. Before the click, we found the significant increase of passive reception of the stimulation from the other subject, which was characterized by the local transfer entropy from the other subject's movement to the stimulation. After the click, this was gradually switched to the active touching, which was characterized by the local transfer entropy from the subject's own movement to the stimulation. This transition from passive to active was evident in the trial with the high perceptual awareness scale.

The experiment of TypeTrace Messengers is designed to investigate the effect of medium during dyadic interactions. Social presence, or the subjective experience of being present with another existing person, varies with the medium used for the interaction. Early theories argued that social presence depends on the richness of information mediated through communication system. Later, counters to this idea argued that even computer-mediated communication systems, however, deprived of social cues compared to face-to-face conversation, can generate as much social presence. Until now, social presence researches in general have mainly focused on uni-directional aspects of each exchanged message. On the other hand, researches in social cognition have studied the importance of bi-directional interaction in understanding dyadic interactions. Our primary purpose is to quantify the degree of social presence among the participants of a realtime online chat system with a few statistical measures. To this end, we developed a software called "TypeTrace" that records all keystrokes of online chat interactants and reenacts their typing actions. Our results show that when we just increase the richness of information by presenting the typing process during the chat, the subjective ratings of how strongly the subjects felt the other's presence does not significantly increase. When the information concurrency of the chat is augmented, we found that transfer entropy between the interactants becomes higher, and the social presence, as well as, emotional arousal, intimacy increased. This result shows that the mere augmentation of information richness does not necessarily lead to an increase in the social presence, and concurrent communication is another important factor for fostering vivid conversation in virtual communication.

Contents

List of Figures

Acronyms

1D one-dimensional. 15

AU Action Unit. 50

CC Cross Correlation. 23

CMC Computer Mediated Communication. 11

CRQA Cross Recurrence Quantification Analysis. 10

DWT Discrete Wavelet Transform. 51

EEG Electroencephalogram. 7

FtF Face to Face. 12

GSR Galvanic Skin Resonance. 49

IKSI Inter KeyStroke Interval. 56

IOS Inclusion of Other in the Self. 49

JS-divergence Jensen-Shannon Divergence. 56

MENT Mentalizing networks. 3

MNS Mirror Neuron System. 3

PAS Perceptual Awareness Scale. 17

PCE Perceptual Crossing Experiment. 15

RQA Recurrence Quantification Analysis. 10

SIPT Social Information Processing Theory. 12

TT Turn-Taking. 21

TT Messenger Type Trace Messenger. 47

WCLR Windowed Cross-Lagged Regression. 23

Part I

Preliminaries

Chapter 1

Introduction

In nature, many complex behaviors arise from the interaction between some components. As a part of these behaviors, dyadic human interactions seemingly constitute a class of unique complex dynamics, and these social encounters are also commonly observed and playing important role in our everyday life. In this thesis, we are aiming at understanding the characteristics of the dynamics and exploring the underlying principle of it.

As an introduction, we first describe conventional approaches for the understanding of social interactions, where social encounters are usually characterized by the use of the particular type of cognition, called social cognition (Frith and Frith, 2012). Social cognition is the cognitive ability of each agent that is used to understand the internal state of the other agents. In this context, as we will see later, dyadic human interaction is understood as the combination of the two decomposed cognitive processes, in which each agent tries to understand the other's state from the observation of the other's behavior.

In contrast to these approaches, some researchers have recently claimed that the interaction dynamics itself should be investigated rather than the individual cognitive ability. We will explain this approach to the social interactions, "2nd person psychology", putting more emphasis on the dyadic interactions itself and claiming that these interactions provide the basis of the cognition and not the other way round. We review the previous experiments and simulations in favor of this approach and the analysis tools for the characterization of the interaction dynamics. Lastly, we provide the overview of the present thesis, which mainly consists of two experiments that were in line with these researches and designed to investigate the characteristics of dyadic interaction dynamics.

1.1 Conventional Approaches in Social Cognition

Social cognition is referred to as the cognitive process of how we understand the other's internal state or "mind". This attribution of the "mind" to the observation of the other's behavior is called mentalization. To infer the internal state, agents are assumed to have a prototypical model, "Theory of Mind", as a basis for the inference (Premack and Woodruff, 1978). To explain the mechanism of how we obtain the prototypical model of the agents, two paradigms were mainly proposed in this field, "Theory Theory" and "Simulation Theory" (Frith and Frith, 2012; Wiltshire et al., 2015).

"Theory Theory" is the paradigm that agents have already had known the "theory" of how the internal state of the agent works and how these appear in their behaviors. Agents observe the others' behaviors and infer the internal state of the others by applying this "theory" to observed behaviors. In neuroscience researches, this mechanism is usually attributed to mentalizing networks (MENT) (Frith and Frith, 2006). In contrast, "Simulation Theory" claims that agents understand others by simulating how they feel when the agents do the same behaviors as the others. This mechanism is attributed to the mirror neuron system (MNS) in neuroscience (Gallese and Goldman, 1998). In this case, agents use internal simulation of others' behaviors to infer others' internal states, so this can be regarded as an approach from 1st person perspective. Until now many behavioral and neuroimaging studies to investigate these systems and have confirmed the role of both systems (Van Overwalle and Baetens, 2009). The inference in "Theory Theory" uses the objective characterization of the other's behavior, which is based on the "third-person perspective", and the inference in "Simulation Theory" uses internal simulations, which is based on the first-person perspective. These approaches will be contrasted to the "second person" approach, which will be explained later.

Another important issue is that when these cognitive systems are triggered. This can be rephrased as what kind of things were regarded as objects with animacy, or agents. This problem is called agency detection. Many experiments have been conducted to find out the condition to be judged as agents. These experiments have revealed that to be judged as agents, it does not need much detail in appearance. For example, Johansson (1973) conducted experiments using point-light displays. In this experiment, the appearance of the human was abstracted to points placed at major joints, and only the motion of the points, which was named "biological motions", were displayed to the subjects, and they found that the subjects perceived these moving points as a human. This experiment showed that motion cue was effective in the perception of agency and the detailed body appearance is not required. Furthermore, Kozlowski and Cutting (1977); Dittrich et al. (1996) showed that the subjects also succeeded to recognize the age and gender only from the movement of the points.

A single object can also be perceived as an agent when its movement has certain characteristics (Scholl and Tremoulet, 2000). This was first shown by Heider and Simmel (1944). In this experiment, they prepared the animation films in which three geometrical objects were moving and interacting with each other in a certain way, and they showed that the subjects who watched the film perceived each geometrical object as an agent. This result was further confirmed in the follow-up experiments (Kassin, 1981). After these experiments, some studies attempted to systematically investigate the characteristics of movement which induce the feeling of agency or animacy. For example, Bassili (1976); Dittrich and Lea (1994) prepared several animations of two moving objects changing the condition of their interactive movements, such as contingency or the type of trajectories, and systematically estimated which aspects of movements were responsible for the feeling of interaction, intention, and animacy. Tremoulet and Feldman (2000) showed that, unlike the previous experiments which displayed movement of several objects, just displaying the movement of single objects could create the feeling of animacy, and they concluded that the major factor was based on the magnitude of the speed change, and the angular magnitude of the direction change in their experimental settings. Some investigations also targeted infants and reported that self-propelled movements (Luo and Baillargeon, 2005) and contingent movements (Johnson, 2003) were treated as agents by infants.

1.2 Interaction in Social Cognition

The two explanation paradigms for understanding others, "Theory Theory" and "Simulation Theory" differ from each other, but they both implicitly assumed the same paradigm, which is a “detached observer”. They assume the problem of cognition of others as the problem of how to passively process the information (others’ behaviors), and interactive processes between two agents were completely missing in the picture (Froese and Gallagher, 2012). The experiments of agency detection introduced above also share the same assumption. In these experimental settings, the presented stimulus moved usually independent of the movement of the observer, and no realtime interaction was involved in the experimental settings. They regarded that the cause of the perception of an agency should be explained as the characteristics of the movements of only the objects being observed, and the agency detection mechanism should be understood as the individual mechanism to detect such characteristics (Frith, 2008).

As an opponent to these detached observer views, some philosophers started to propose “Interaction Theory” as an alternative paradigm to “Theory Theory” and “Simulation Theory”, which claimed that the interaction was central to understanding others (Froese and Gallagher, 2012). For example, De Jaegher et al. (2010) claims that, based on developmental studies and dynamical systems, interactions are not the contextual factor, but an enabling factor for social cognition. The background of these theories is based on the field of embodied cognition. In cognitive science, the traditional way of thinking cognition as a symbol manipulation or an information processor has been criticized by the approach called embodied cognition (Varela et al., 1991; Clark, 1998). Embodied cognition claims that cognition is an active process in the environment, not a passive information processing, and regards the interaction to the environment itself constitute the cognitive process. In this way, interaction theory claims that the interaction dynamics itself contribute to the social cognitive process.

1.3 The 2nd Person Psychology

After the proposal by the philosophers that interaction should be central to the study of social cognition, some researchers in psychology started to pay attention to the interactive aspect of cognition (Sebanz et al., 2006; Gallotti and Frith, 2013).

Recently, Schilbach et al. (2013) claimed that the investigation on social cognition had not fully investigated the social encounters in a truly interactive manner, and it was representing the “dark matter” of social science.

They claimed that the study on social cognition depended on spectatorial accounts, which assume that when people make sense of other people, each agent is a detached observer and intellectually infer the state of the other from the information obtained by observation. For example, the "simulation theory" account of social cognition provides the understanding of the other by simulating myself, or the "first-person grasp", and the "Theory Theory" account regards the understanding of the other as the inference based on the innate knowledge, or "third-person grasp", but these two describe the way how we process the information of the other and these are explanations from spectatorial accounts. Also, the experiments on agency detection such as the Heidler's moving objects assumed the detached observer, not as the possible interactant.

To overcome this paradigm, they proposed the "second-person" approach (Schilbach et al., 2013; Redcay and Schilbach, 2019), which set social interaction and emotional engagement as central constituents. Emotional engagements are feelings of engagement with and emotional responses to the other, and social interactions are reciprocal interactions with others. Three proposals were presented towards the second-person approach, (i) a person being a detached observer as compared to experiencing a social situation with an attitude of emotional engagement ("experience"), (ii) experimental paradigms used to investigate social cognition allowing or not allowing for interaction ("participation"), and (iii) data collection and analysis taking place at the level of a single or two (or more) individuals ("data collection & analysis"). In this approach, mutual interaction is indispensable and cannot be simply decomposed into each agent's inference process, therefore this approach is regarded as a 2nd person perspective approach. We will review experimental and theoretical studies of this approach in the next section.

1.4 Experimental Study

Below, I will present examples of the experimental paradigms of second-person psychology that include realtime dyadic interactions.

1.4.1 Synchronized tapping

Synchronized tapping is an experimental paradigm using finger tappings in which subjects are asked to maintain a given beat and also synchronize to a signal from the other subject's tapping. This experiment aims to analyze what kind of dyadic interaction pattern emerges as a result of the coordination of the pair. For example, Konvalinka et al. (2010) found that if only one of the subjects can hear the other's sound, i.e. unidirectional coupling, "leader-follower relations", in which one subject follow the movement of the other, emerged. When both subjects can hear each other, then there are no "leader-follower relations", and bidirectional coupling occurred, based on the analysis using windowed cross correlations. Also, Tognoli et al. (2007) used brain-imaging and found that these components were suggested to belong to the human MNS, hence inhibiting and enhancing the MNS. Despite the simultaneous EEG recordings, however, this study did not look at inter-brain interactions between interacting partners. As a variation of this study, synchronized aiming Skewes et al. (2015) showed that the asymmetry in the task difficulty in each subject causes leader-follower relations.

Movement synchronization can happen without any intention to couple with each other and this phenomenon has been reported in many experiments, such as pendulum (Schmidt et al., 1990; Schmidt and O'Brien, 1997; Richardson et al., 2005), postural swaying (Shockley et al., 2003), rocking chairs (Richardson et al., 2007) and so on. This synchronization happens even the subjects are instructed not to synchronize (Schmidt and O'Brien, 1997; Richardson et al., 2007). Also, the synchronization to the other can increase the affiliation to the other (Hove and Risen, 2009).

1.4.2 Gaze Interaction

Gaze has been known to be an important component in social interactions (Emery, 2000), and some experiments using realtime gaze interaction have been reported (Schilbach, 2015). Bayliss et al. (2013) used a gaze-contingent eye-tracking system and showed that presenting a picture of a face with congruent and incongruent gaze affected the gaze behavior and affective behaviors. Using this type of system, Pfeiffer et al. (2011) constructed an interactive system that presented a virtual character whose gaze was congruent or contingent to the subject's gaze and showed that the congruent movement enhance the ascription of humanness to the avatar, and if the subject believed that the other was cooperative, the contingent movement drives the ascription.

Hirsch et al. (2017) recorded the brain activities of two subjects during eye-to-eye contact and found cross-brain coherence in signals originating within the left superior temporal, middle temporal, and supramarginal gyri as well as the pre and supplementary motor cortices.

1.4.3 Conversation

During conversations, postural coordination has been reported to happen spontaneously in many studies (Shockley et al., 2009), and unconscious mimicry occurred when they had a desire to create rapport (Lakin and Chartrand, 2003). Suda et al. (2010) recorded brain activity during face-to-face conversation and found a robust activation over the frontal and superior temporal regions. Also, some reported that the use in linguistic expression was aligned during a conversation (Pickering and Garrod, 2004).

In the context of developmental studies, Murray and Trevarthen (1985) investigated the dyadic interaction between infants and their mothers, and compared the reaction of infants when communicating through live video chats, and recorded videos. In this study, they showed that 2-month old infants can distinguish the live video and the recorded video of their mother. However, some concerns were raised about this classic research. For example, the infants may have become simply fussy over time, or the transition from the live videos to the recorded videos can be noticed by infants because these were not smoothly connected by the limitation of the technology at that time. Nadel et al. (1999) replicated this experiment with a more controlled setup to answer these concerns and confirmed that the babies showed different behavior between live and recorded conditions.

Perceptual Crossing Experiment (Auvray et al., 2009), which we study in this thesis and will be explained in the next chapter, was designed to make the minimalistic parallel of these experiments to investigate what is happening during the dyadic interactions.

1.5 Simulation Study using Dynamical Systems

Interactionists claimed that the interaction dynamics itself can explain various experimental results that were previously considered impossible without the inference of others. To claim this, constructing the example that these phenomena can be constructed using simple dynamics is important. Below we will show the example of these models using numerical simulations.

The simplest case is interpersonal synchronizations, which were observed in, for example, pendulums, rocking chairs, and postures. For the explanation of alignments, coupled oscillator models were widely used (Haken et al., 1985). For example, Richardson et al. (2007) analyzed the synchronization in rocking chairs based on coupled-oscillator models.

For the modeling of more complex behaviors, for example, Iizuka and Ikegami (2004), and Ikegami and Iizuka (2007) focused on the particular dyadic behavior called "turn-taking", which indicate that the two agents behave alternately. They simulated this behavior as a tag-game with the role of the chaser and the evader switches spontaneously in a 2D environment. Each agent possessed a recurrent neural network and sensed another agent's relative position, velocity, and angle. Agents were evolved using a genetic algorithm based on the ability of turn-taking. They found there emerged mainly two types of agents, regular turn-takers and chaotic turn-takers, and chaotic turn-takers were more adaptive to different generations of agents, and also sensitive to live agents and recorded agents. Also, their simulations showed that during turn-taking the predictability of the other agent's behavior decreased.

Di Paolo et al. (2008) and Iizuka and Di Paolo (2007) simulated the perceptual crossing experiment, which we will explain later, by evolutionary robotic techniques, which will be described in the next chapter.

1.6 Characterization of Dyadic Interactions

In this section, we will review the methods to quantify the dyadic interactions.

1.6.1 Cross Correlation

Cross correlation between two timeseries X(t) and Y(t) with lag τ is defined as,

$$R_{XY}(\tau) = \frac{1}{T}\sum_{t}\frac{X(t) * Y(t-\tau)}{\sigma_X \sigma_Y},$$

where T is the total time steps, σ_X and σ_Y were standard deviation of X(t) and Y(t), respectively. Cross correlation is widely used to quantify the degree of synchronization between two time series data. Konvalinka et al. (2010) used cross correlation to analyze the synchronized tapping experiments, and measured the degree of synchronization between pairs. Also, they identified the leader-follower relationship by calculating cross correlation changing the lag τ.

1.6.2 Recurrence Quantification Analysis (RQA)

Recurrence quantification analysis (RQA) is a method to characterize the dynamical system using the structure of recurrence in its phase space (Marwan et al., 2007). RQA uses a recurrence matrix $\boldsymbol{R_{i,j}}$ calculated as follows,

$$\boldsymbol{R_{i,j}} = \begin{cases} 1 & (\vec{x_i} \approx \vec{x_j}) \\ 0 & (\vec{x_i} \not\approx \vec{x_j}) \end{cases},$$

where $\{\vec{x_i}\}$ is a trajectory in its phase space and $\vec{x_i} \approx \vec{x_j}$ means equality up to an error ε.

Fusaroli and Tylén (2016) applied this method to analyze the lexical, prosodic, and speech/pause data during conversations and quantified the structure by the average length of the recurrent trajectory of the system, which was used as a measure of the regular patterns, and the entropy of the length distribution, which was used as a measure of the presence of a plurality of patterns. They used RQA in three ways. First, they applied RQA to the time series data from each subject to check the presence of structure in each subject's speech. They claimed this as the test for "self-consistency". Second, they used cross recurrence quantification analysis (CRQA), which quantifies not the recurrence in single time series, but calculates the similarity between two time series, to quantify "interactive alignment". Third, they used RQA to the time series data which was constructed by merging two time series data from each pair, and they claimed that this enabled to quantify the "interpersonal synergy".

1.6.3 Allan Factor

Allan factor was originally proposed in Allan (1966) and has been used to characterize fluctuations of point processes in different time scales, especially to identify the flicker noise in the process. Allan factor $A(T)$ is defined as follows,

$$A(T) = \frac{\langle (N_i(T) - N_{i+1}(T))^2 \rangle^2}{2\langle N_i(T) \rangle},$$

where $N_i(T)$ is an event count in the ith window segmented in size T.

Kello et al. (2017) calculated Allan Factor from different categories of sound such as speech, music from different genres, and vocalization of non-human animals, and compared the results across all categories. From the results, they classify the sound data based on the degree of the presence of a hierarchical temporal structure. Abney et al. (2014) applied Allan Factor to acoustic onset times of the audio of conversations, and by calculating the similarity of Allan Factor between each dyad, "complexity matching", they found that, when the conversation was affiliative, Allan Factor resembled each other.

1.6.4 Transfer Entropy

Transfer entropy was developed by Schreiber (2000) to identify the informational flow between two time series, and defined as follows,

$$T_{Y \to X} = I(\mathbf{Y}_{n+1}^{(l)}; X_{n+1} | \mathbf{X}_n^{(k)}),$$

where $\mathbf{X}_n^{(k)} = \{X_{n-k+1}, ..., X_n\}$, $\mathbf{Y}_n^{(l)} = \{Y_{n-l+1}, ..., Y_n\}$ (k: target history length, l: source history length).

The use of transfer entropy for the identification of directional influence between dyads was proposed (Hasson and Frith, 2016), but at present, in very few researches transfer entropy is used to analyze the dyadic behavioral data (Trendafilov et al., 2020). Granger-Causality is another method to identify the information flow, and in some researches in neuroscience, Granger-Causality is used to analyze information flow between brain activities of dyads (Schippers et al., 2010) .

1.7 Studies in Computer Mediated Communications

Schilbach et al. (2013) claimed that the aspect missing in conventional social cognition research is emotional engagements, which are feelings of engagement and emotional responses to the other, in addition to realtime interactions. The different line of research, computer mediated communication (CMC), is related to these aspects, and the study of CMC has been interested in how CMC influences interpersonal relations. Here, I will briefly review some theories from CMC researches, which will be reviewed in detail in chapter 6.

Short et al. (1976) argued that various communication media differed in their capacity to transmit classes of nonverbal communication in addition to verbal content. The fewer the number of cue systems a system supported, the less warmth and involvement users experienced with one another. Culnan and Markus (1987) called this type of theory as "cue-filtered out theory". This line of research continued to gain support in the 1980s, where Daft and Lengel conducted seminal analyses of the richness of information and media (Daft and Lengel, 1984, 1986). In the 1990s, Walther argued in opposition to media richness theories, stating that if enough time is spent on CMC, interactants can achieve a level of interpersonal relationship as high as face-to-face (FtF) communications (Walther, 1992). Walther later argued that communication means specific to CMC could even create a more robust social bond compared to FtF (Walther, 1996) because it stimulates more self-disclosure than FtF and thus can lead to higher social attraction. This school of thought, called the social information processing theory (SIPT) (Walther, 2015), has helped researchers to transcend the simple dichotomy of rich and poor media and to scrutinize the social phenomena in CMCs more in-depth. These researches, however, investigated the characteristics of the messages and not the interaction dynamics between the agents, which will be the focus of our second experiments.

1.8 Overview of the Thesis

The present thesis consists of mainly two parts. In the first part, we will present the result from the reanalysis of perceptual crossing experiments (PCE). This part is aimed at showing how we can quantitatively characterize the dyadic dynamics especially using the information theoretical value, Transfer Entropy, and how these values are related to the presence of others in this minimal interactive experiment. In this part, we will start by reviewing the previous studies on PCE, and its interpretations. Next, we will explain the experimental data that we used in our analysis, and the analytical methods we used here. Then we will show the result of our analysis, and discuss the results by especially emphasizing the comparison to the previous researches of PCE.

In the next part, we will present the experimental research on the effect of communication medium using our newly developed text chat system "Type Trace Messengers". The main focus of this part is how the difference in settings of our text chat system alters the subjective reports, physiological responses, and typing patterns of the subjects. After a brief review of the background research of CMC, we will explain the chat system that we developed for this experiment, the experimental setups, and methods for our analysis. Then we will report the results of our experiments, which include the comparison to the phone call dataset.

Part II

Perceptual Crossing Experiment

Chapter 2

Review of Previous Experiments

Perceptual crossing experiment (PCE) was designed as simplified parallel of Murray and Trevarthen (1985)'s "double TV experiment". In this experiment, two subjects were asked to identify the other in the one-dimensional (1D) virtual space.

Two subjects were placed in different space, and each had a computer mouse and a tactile stimulation device. By moving the mouse, subjects can move their avatars in virtual 1D space. If both avatars were in the same position in the 1D space, both subjects received feedback from the tactile stimulation device.

In the virtual space, in addition to avatar objects two kinds of different objects existed. One was "static object" which stayed at the same position in the 1D space, and the other object was "shadow" which located at the certain distance from the other avatar. If the avatar and these objects are in the same position, the subject also received tactile feedback. The task of each subjects was to click during the tactile stimulation if they thought this stimulation came from the other avatar.

2.1 Result from Auvray (2009)

Auvray et al. (2009) did this experiment to 20 subjects (10 pairs) and found that the number of correct click (clicking when the avatar touched the other avatar) was significantly larger than the number of wrong click (clicking when the avatar touched the fixed object or the shadow).

However, they found that the probability of click (the number of click divided by the number of encounter with the objects) was not significantly different between the avatar and the shadow, which meant that the subject cannot distinguish between the avatar and the shadow.

They concluded when two avatars interacted with each other the dynamics were stabilized, so the duration of the tactile stimulation from the avatar increased and as a result they tended to click correctly.

They also reported that the subject tended to reverse the direction of the movement when they crossed a source of stimulation, which might be the mechanism to stabilize the interaction between the agents.

2.2 Simulation study on the PCE

Di Paolo et al. (2008) and Iizuka and Di Paolo (2007) simulated the perceptual crossing experiment by agent simulation using recurrent neural networks. They used the same setup with the original perceptual crossing experiment, and agents were evolved to stay close with each other as long as possible.

They found that the agents were evolved to be able to achieve the task, and this evolved agents show the behavior that after crossing the source of the simulation the agents tended to reverse the direction of movement, which was also observed in Auvray et al. (2009).

2.3 Lenay's variation

After the Auvray et al. (2009)'s experiment, some variations of the experiment have been conducted. One example was Lenay and Stewart (2012). In this variation, subjects received feedback not as tactile stimulation, but 3 kinds of sounds. When agents encounter different kinds of object, subjects received different sounds and the subjects were asked to identify which sound is assigned to the other agent.

In this setup, they found that subjects can successfully detect the avatar from the shadow. They explained this success because each object is accompanied with always the same sound, so the subjects can simply compare the intrinsic properties for the three objects.

2.4 Froese's variation

Froese et al. (2014a) and Froese et al. (2014b) also did the variation of the PCE. In this experiment, the scores are evaluated on the pair not on the single subject and promote the subjects to cooperate with each other, and also each subject was allowed to click only once in a single trial to enhance attention on the stimulation.

In this case, they found that the subjects significantly succeeded to distinguish between the agent and the shadow.

They called the situation that two subjects click correctly in the same trial, "Joint success", and only one subject click correctly "single success". They found that the number of joint success was significantly larger than the single success.

Also they found characteristic dynamics "turn-taking", which indicated the behavior that two agents moved alternately, and this was related to the subjective scale of the other's presence "Perceptual Awareness Scale" (PAS).

2.5 Aim of the following analysis

In the following chapters, I will present the re-analysis of Froese et al. (2014a)'s experiment.

The aim of this analysis was to clarify the characteristics of the dynamics which caused the feeling of the other's presence. In the previous papers, we only had turn-taking as a measure to investigate the dynamics, and this can only capture the one aspect of it. Here, we will use synchronization measure, such as cross correlation, and directed information measures to investigate the dyadic dynamics from different point of view.

Chapter 3

Methods

In this study, we re-analyzed the experimental data originally reported by Froese et al. (2014a) by characterizing the time series data quantitatively using different measures from the previous works.

In this chapter, first I will describe the detail of the experiment of Froese et al. (2014a), and then I will introduce the measures which we used to re-analyze the experimental data. The main part of chapter 3 and 4 was published in Kojima et al. (2017).

3.1 Participants

Participants were healthy volunteers recruited from acquaintances at the University of Tokyo and at Osaka University (N = 34). There were 25 Japanese nationals, the rest were from various countries. Six were female. The mean age was 29 years. Teams of participants were created as volunteers became available.

The study protocol was approved by the local ethics research committee of the Graduate School of Information Science and Technology, Osaka University, and by the local ethics research committee of the Graduate School of Arts and Sciences, the University of Tokyo, and has been performed in accordance with the ethical standards laid down in the Declaration of Helsinki. All of the participants gave their written informed consent before taking part in the study.

3.2 Experimental Setup

In Froese et al. (2014a)'s version of the perceptual crossing paradigm, two adults are placed in distinct locations such that they cannot perceive each other; their sight is blocked and they wear noise-cancelling headphones (see Figure 3.1). Their only manner of making contact is via a simple interface consisting of a trackball that records horizontal movements and a hand-held vibration motor that is either on or off. The trackball is operated with the dominant hand while the motor is held in the other hand. Their movements control the motions of an avatar located in an invisible 1D virtual environment (see Figure 3.2). This 1D virtual environment was 600 unit long with periodic boundaries, and each object was embodied as 4 unit long. The motor continuously vibrates whenever their avatar overlaps with another object in the virtual space. Position and sensor data was recorded every 10 ms (100 Hz).

This version of perceptual crossing experiment was designed to enhance the cooperativity between the subjects, which was the main difference from the Auvray et al. (2009)'s original perceptual crossing experiment. In order to do this, the score of the task was evaluated on the pairs of players, not on the score of the single player, and also explicitly instructed to help each other to find each other in the virtual space.

They are to click once using the trackball (and only once per trial) in order to signal to the experimenters when they become aware of interacting with the other player; the other player is not aware of the click. No feedback is provided during the experiment, which meant that each participant was not told whether the click was correct or not after each trial. Each pair can interact in a sequence of 15 trials, each with a duration of 60 seconds.

3.3 Subjective Reports

After each trial the experience of the players is evaluated in several ways if they happened to click in that trial.

In particular, they were asked to rate the clarity of their experience of the other's presence at the moment of their click on the basis of a social version of the perceptual awareness scale (PAS), based on the PAS that was proposed by Ramsøy and Overgaard (2004).

This scale was constructed by studying which scale participants prefer to use for reporting conscious experiences if allowed to construct the scale themselves(Ramsøy and Overgaard, 2004), and these scales was shown to be related to the particular neural response, the visual awareness negativity in the case of visual conscious experiences by a magnetoencephalographic study (Andersen et al., 2015).

This scale is a four-point scale, and 1 means having had no experience, 2 means having had an ambiguous experience, 3 means having had an almost clear experience, and 4 means having had a clear experience (table 3.1).

They were also asked to give a short free-text description of that experience and their strategy. (Figure 3.3)

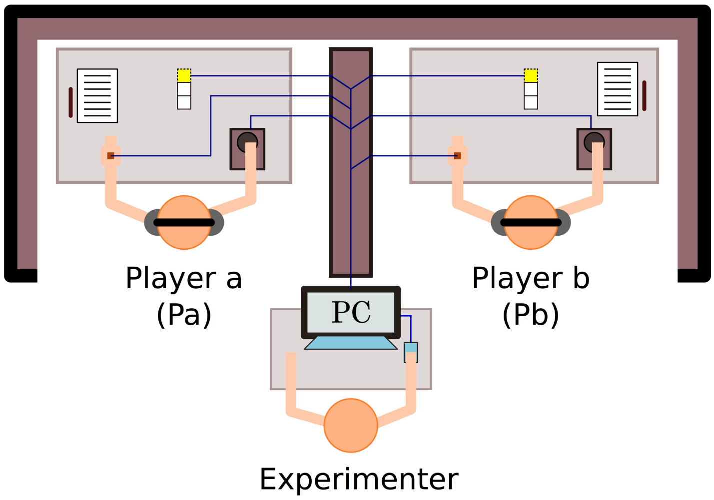

Figure 3.1: Experimental setup of perceptual crossing paradigm. The two participants can only engage with each other via a human-computer interface that reduces their scope for embodied interaction to a minimum of translational movement and binary tactile sensation. Each player's interface consists of two parts: a trackball that controls the linear displacement of their virtual avatar, and a hand-held haptic feedback device that vibrates at a constant frequency for as long as a player's avatar overlaps with another virtual object and remains off otherwise. Three small lights on each desk signal the start, halftime (30 seconds), and completion of each 60-second trial. Figure originally published in Froese et al. (2014a).

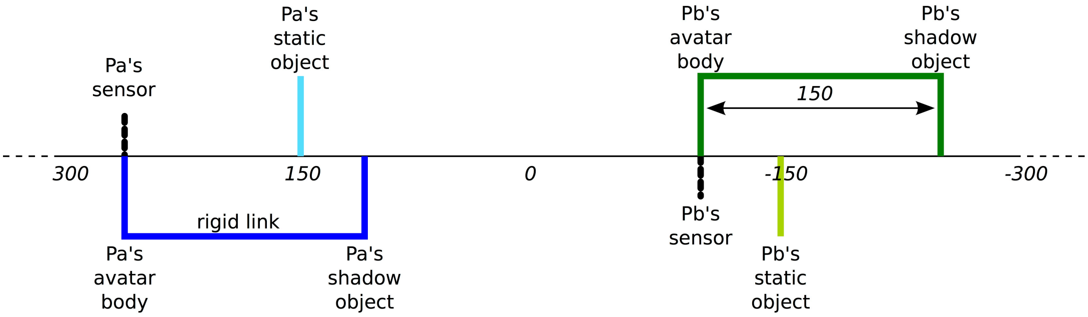

Figure 3.2: Virtual environment of perceptual crossing paradigm. Players Pa and Pb are virtually embodied as “avatars” on a line that wraps around after 600 units of space. This virtual space is invisible to the participants. Each avatar consists of a binary contact sensor and a body object. Unbeknownst to the players a “shadow” object is attached to each avatar body at a fixed distance of 150 units. There are also two static objects, one for each player. All objects are four units long and can therefore only be distinguished interactively in terms of their qualitatively different affordances for tactile engagement. No other forms of interaction were possible. Figure originally published in Froese et al. (2014a).

3.4 Analysis of Sensorimotor Trajectories

The dynamics of each trial was fully captured by 4 time series data, which were each agent's movement trajectory, and the time course of the sensory feedback by vibrating device. Here, I will introduce the methods that we used to characterize the data, turn-taking, cross correlation, windowed cross-lagged regression and transfer entropy.

3.4.1 Turn-Taking

First measure was the degree of turn-taking (TT). Turn-taking corresponds to the behavior that two agents moves alternately, and this measure was designed to capture this behavior by Froese et al. (2014a). Basically, this measure is high if two agents don't move simultaneously, and the movement duration of each agent is close.

At each time step we classified the state of each player's behavior in binary terms as either moving (1) or non-moving (0) by evaluating his or her

PAS	Experience of other's presence
1	No experience
2	Ambiguous experience
3	Almost clear experience
4	Clear experience

Table 3.1: Froese et al. (2014a)'s social version of the perceptual awareness scale (PAS) adapted from the PAS by Ramsøy and Overgaard (2004).

trackball movement (I will refer to these binary movement time series as B1 and B2 for participants Pa and Pb, respectively). Movement was considered to have taken place whenever the change in avatar position dx from one time step to the next was bigger than an 8th of the avatar's length (i.e. 4/8 = 0.5 so that if $dx > 0.5$ unit long, 1, else 0). Since avatar positions tend to fluctuate during a player's "turn" We chose to set a lower limit to the duration of movement pauses so as not to accidentally end up with a turn being divided into micro-turns. Thus, we only set movements to 0 if there was no motion over at least 50 consecutive time steps (500 ms), otherwise they remain set to 1.

In order to determine the differences between players' activity we applied the logical "Not-And" operator to their movement time series, which resulted in a time series of activity differences D (i.e. D = B1 Not-And B2). Then, we assigned to each participant their active contribution of this exchange by applying the logical "And" operator and summing the result (i.e. C1 = sum(B1 And D); C2 = sum(B2 And D)). The overall TT performance for a given time period was then calculated by multiplying the player's active contributions. This multiplication means that one-sided situations, in which one player is continuously active while the other is continuously passive, get low TT scores. Finally, we normalized the outcome such that the turn-taking score TT = (4 * C1 * C2) / T2, where T is the number of time steps. The range of TT is therefore [0, 1], with 0 representing a complete absence of TT interactions and 1 representing a perfect exchange of periods of activity and passivity between the subjects. We analyzed the TT in the 10s preceding a click.

3.4.2 Movement Synchronization: Cross Correlation and WCLR

The measure of turn-taking (TT) interaction that was proposed by Froese et al. (2014a) can tell us whether players were exchanging periods of activity and passivity in an orderly manner, but it does not say much about the similarity of the patterns of activity that were being exchanged. Given that interpersonal synchrony is widely considered to reflect psychological connectedness, we applied measures of movement synchrony, namely cross correlation (CC) and windowed cross-lagged regression (WCLR)(Altmann, 2011).

Cross correlation (CC) between two timeseries X(t) and Y(t) with lag τ is calculated as,

$$\text{(CC)} = \sum_t X(t) * Y(t-\tau)/\sigma_X \sigma_Y$$

σ_X and σ_Y were standard deviation of X(t) and Y(t), respectively. Here we divided by these standard deviation in order to take into account of individual difference in the motion.

Windowed cross-lagged regression (WCLR) between two timeseries X(t) and Y(t) with lag τ is calculated in the following way. First, we models one time series $X(t)$ in two different ways.

Model1:

$$X(t) = \beta_{10} + \beta_{11} X(t-\tau) + \epsilon_{1t}$$

Model2:

$$X(t) = \beta_{20} + \beta_{21} X(t-\tau) + \beta_{22} Y(t-\tau) + \epsilon_{2t}$$

Both models are linear regression using lagged time series, but model1 only use only own past time series data and model2 use not only own past data but also time series data from the other time series data.

Then using the coefficients of determination (R^2_{Model1} and R^2_{Model2}), windowed cross-lagged correlation (WCLR) is quantified as below.

$$\text{(WCLR)} = R^2_{\text{Model2}} - R^2_{\text{Model1}}$$

CC is a common measure of movement synchrony but can be confounded by auto-correlation, which may lead to inflated measures of interpersonal synchrony, a problem which is avoided by WCLR. Therefore, we decided to use WCLR in addition to CC.

We calculated CC and WCLR between two agents' velocity data in the periods of 10s before each click with a time lag in the range of [-5, 5] seconds, which means that clicks occurring during the first 15s of a trial were excluded from CC and WCLR analysis. In total 28 clicks had to be excluded.

We also used WCLR to calculate the windowed time delay yielding the largest WCLR value, which gives an indication of the most relevant timescales in which synchrony can be measured.

3.4.3 Local Transfer Entropy

Lastly, we looked more specifically at the influence of players' movements on each other's tactile sensations preceding a click using a measure known as local transfer entropy (Lizier et al., 2008).

Transfer entropy was originally proposed by Schreiber (2000), and this measure can capture the directional influence from one time series to the other time series. This transfer entropy can be also formulated in temporally local form (Lizier et al., 2008), which allows us to calculate transfer entropy for specific segments of a time series.

First, transfer entropy $T_{Y \to X}$ is defined as

$$T_{Y \to X} = \sum_{u_n} p(u_n) \log \frac{p(x_{n+1} | x_n^{(k)}, y_n^{(l)})}{p(x_{n+1} | x_n^{(k)})}$$

Here, n is a time index, u_n represents the state transition tuple $(x_{n+1}, x^{(k)}, y^{(l)})$, $x^{(k)}$ and $y^{(l)}$ represent the k and l past values of x and y up to and including time n. In the following analysis, we used $k = l = 4$.

This can also be formulated using mutual information.

$$T_{Y \to X}(k, l) = I(\mathbf{Y}_n^{(l)}; X_{n+1} | \boldsymbol{X}_n^{(k)})$$

When this value is calculated from the experiment, the probability of u_n is calculated as

$$p(u_n) = \frac{c(u_n)}{N} = \frac{\sum_{a=1}^{c(u_n)} 1}{N}$$

Here, N is the total number of the observation, and $c(u_n)$ is the count of the u_n.

If we substitute this to the definition of transfer entropy, we can derive the following equations.

$$
\begin{aligned}
T_{Y \to X} &= \frac{1}{N} \sum_{u_n} (\sum_{a=1}^{c(u_n)} 1) \log \frac{p(x_{n+1}|x_n^{(k)}, y_n^{(l)})}{p(x_{n+1}|x_n^{(k)})} \\
&= \frac{1}{N} \sum_{n=1}^{N} \log \frac{p(x_{n+1}|x_n^{(k)}, y_n^{(l)})}{p(x_{n+1}|x_n^{(k)})}
\end{aligned}
$$

Therefore, if we define $t_{Y \to X}(n+1)$ as

$$
t_{Y \to X}(n+1) = \log \frac{p(x_{n+1}|x_n^{(k)}, y_n^{(l)})}{p(x_{n+1}|x_n^{(k)})}
$$

transfer entropy can be described as the average of $t_{Y \to X}$.

$$
T_{Y \to X} = \langle t_{Y \to X}(n+1) \rangle
$$

This value $t_{Y \to X}(n+1)$ depends on the time index and this is temporally local, so Lizier et al. (2008) called this local transfer entropy and this can provide the spatiotemporal information transfer profiles from the time series data.

We applied this measures to the time series data from PCE experiment. This analysis of the whole sensorimotor loop is a significant methodological advance because previous time series analyses of perceptual crossing have only focused on movement by itself, thus leaving the interdependency between movement and sensation underlying meaningful perception (Mossio and Taraborelli (2008); Noë (2004)) unexamined.

Transfer entropy uses a joint probability distribution from two time series, and to calculate this distribution function empirically from the time series, we need to discretize the time series. The simplest way is to convert the movement time series to a binary sequence is in terms of the directional change of movement, i.e. we label the point in time as a state 1 when it changes, otherwise we turn it into a state 0. The sensory time series was also converted to a binary sequence, i.e. when the haptic feedback turns on or off we label the point in time as a state 1, otherwise it set to 0 to mark the absence of a change in sensor state. In order to determine the most important time scale of the time series, first we calculated transfer entropy by utilizing the whole trial while adjusting the down-sampling rate. We found there was a peak of transfer entropy around 50ms from the movement data to the sensory input data. We therefore took 50ms as the characteristic time scale and used it for the further analysis of the local transfer entropy.

Given that we are interested in determining the sensorimotor signature of social awareness, we related these objective measures with the subjective PAS ratings of the clarity of the other's presence. In particular, we excluded clicks that were not reported to have been associated with an experience (i.e. PAS 1 or no PAS report), and restricted the data to ambiguous, almost clear, and clear experiences (i.e. PAS 2, 3, and 4, respectively). In addition, we did not further discriminate between the clicks that are associated with these conscious reports in terms of their objective correctness since we were interested in studying the general conditions of the transition to a social experience rather than to a veridical social experience per se. The final dataset consisted of 101, 122, and 143 clicks associated with reports of a PAS score of 2, 3, and 4, respectively. Out of these 366 clicks 321 correctly identified the other's avatar.

3.5 Statistical Analysis

In order to analyze the relationship between the perceptual awareness scale (PAS) and the movement coordination measures (TT, CC, WCLR), first we averaged those movement measures with the same PAS, and applied one-way ANOVA followed by post hoc Bonferroni test.

In the analysis of local transfer entropy, we used Welch's t-test to examine whether the average transfer entropy were different before and after a click.

Doc. v1.6

Trial #1/15:

Q1. *If you clicked during the trial, how clearly did you experience that you were interacting with your teammate just before the time of the click?*

Please select a category to describe the experience of your partner:

1. No experience
2. Vague impression
3. Almost clear experience
4. Clear experience

Q2. *If you chose category 2, 3 or 4, please describe the sensation (or the experience, feeling, etc.) of your partner's presence more precisely in your own words.*

Q3. *If you clicked during the trial, how confident are you now that you correctly identified a moment during which you were interacting with your partner?*

Please select a category to describe your confidence in your click's accuracy:

1. No confidence
2. Low confidence
3. Medium confidence
4. High confidence

Q4. *Please briefly describe your strategy for helping your team in this trial. For example:*

- *How did you help your partner to recognize which of the objects was your avatar?*
- *How did you know when you were interacting with your partner's avatar?*
- *How did you know when you were interacting with a static or a prerecorded object?*

4

Figure 3.3: The questionnaire after each trial.

Chapter 4

Results

Players' activity during a typical trial is shown in Figure 4.1. Note the extended period of interpersonal interaction in the first half of the trial, followed by nearly instantaneous clicks by both players. This is followed by disengagement, then a short interaction with their respective static objects, and finally re-engagement just before the end of the trial.

4.1 Qualitative Analysis of Movement Coordination

We can use the example trial shown in Figure 4.1 to illustrate the cross correlation (CC) and windowed cross-lagged regression (WCLR) measures (Figure 4.2). It can be seen that CC greatly overestimates the amount of movement synchrony, while WCLR picks out only a few temporal regions. For example, there is a bright blue patch from x = 15 s to 20 s for time lags of around -3 s. This tells us that during the preceding 10 s periods, starting from 5 s to 15 s and ending during the period of 10 s to 20 s of that trial, player Pb leads Pa with a delay of around 3s. If we check this result against what is happening during that time in Figure 3, we can see that indeed player Pb (blue line) leads Pa (red line) beginning around 8 s by inducing the latter to also start oscillating. In the 10 s preceding the clicks, Figure 4.1 shows that the direction of influence has become reversed, with Pa staring to oscillate from around 25 s and then pausing, while Pb continues to oscillate until pausing around 30 s. If we compare this with Figure 4.2, we see some bright blue bands following the clicks for time lags of around 2 s,

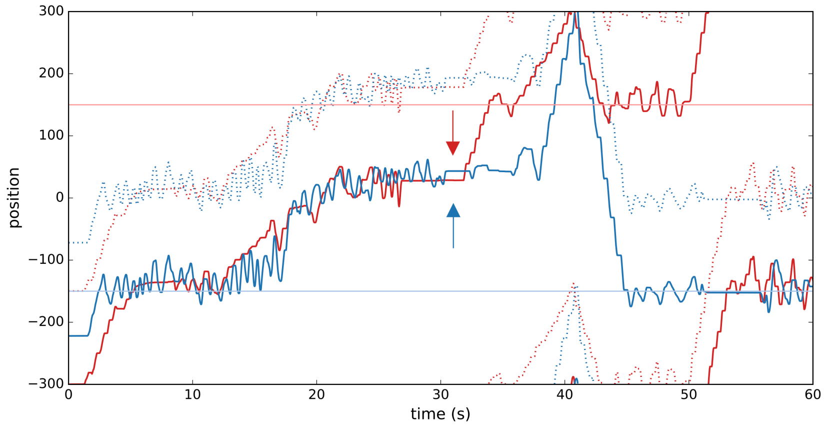

Figure 4.1: Example of time course of an illustrative trial (E1T1). Thick red and blue lines show the change in position of the avatars of players Pa and Pb, and red and blue dotted lines trace their "shadow" objects. Horizontal lines at y = -150, 150 correspond to the position of the player-specific static objects. Red and blue arrows indicate the time of clicking by Pa and Pb, which occurred practically instantaneously (within 0.05 seconds). After the trial players Pa and Pb reported that their experience of the other's presence at the moment of the click consisted in a "vague impression" (PAS 2) and a "clear experience" (PAS 4), respectively.

which suggests that in the seconds preceding the clicks Pa's behavior leads Pb's behavior.

Nevertheless, it can also be observed that the WCLR method may be confounded when the players happen to move similarly but without interacting directly. For example, it turns out that the highest values in Figure 4.2 are given for the period from around x = 42 s to 50 s for lag times of around 0.5 s, even though Figure 3 reveals that in the corresponding period starting from 32 s onwards the players had already separated and just happened to move in roughly similar ways, with Pa slightly leading Pb but without direction interaction. However, we do not arbitrarily want to exclude such cases of behavioral coordination because they may still tell us something meaningful about the quality of the interaction. After all, one possible reason why these two players continued to move similarly even after spatially disengaging is that they had already become entrained during the first half of the trial.

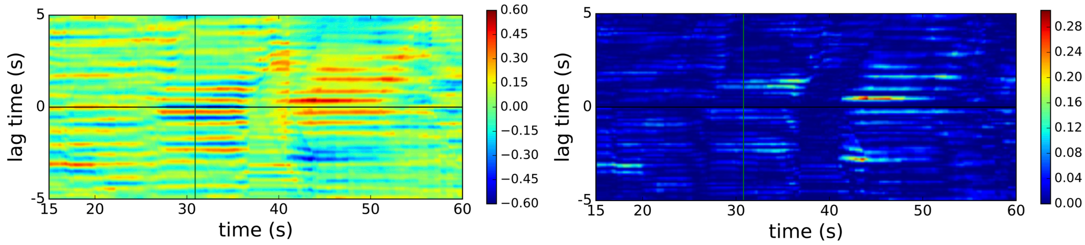

Figure 4.2: Heat maps of cross correlation (CC, left) and windowed cross-lagged regression (WCLR, right) for an illustrative trial (E1T1). Measures are applied to periods of 10 seconds. The x-axis corresponds to the end point of the time window. The y-axis corresponds to the length of the windowed time lag; a positive sign means that behavior of participant Pa can explain that of Pb after the given delay (conversely, a negative sign means that the direction of influence instead goes from Pb to Pa). Thus, it starts at 15 s because a lag time of 5 s means that we compare Pa's activity from 0 s to 10 s with Pb's activity from 5 s to 15 s (or vice versa for a lag time of -5 s). The vertical lines after 30 seconds represent the nearly instantaneous moment of clicking by both players.

We note that turn-taking interaction and synchrony can both give high values for a trial when the players exchange periods of activity and passivity whereby that activity is similar in form, too. But they can also be mutually dissociated in other cases. As illustrated in Figure 4.3, players can exchange periods of activity and passivity whereby that activity itself does not have much resemblance (high TT and low WCLR), and players can greatly overlap in their activity but still share a lot of similarity in their movements (low TT and high WCLR). Here we applied the measures to the 10 seconds preceding a click, and we calculated the WCLR value to be the maximum value from a range of window time lags [-5 s, 5 s].

4.2 Quantitative Analysis of Movement Coordination

Turn-taking interaction and movement synchrony could spontaneously emerge from the interaction dynamics without necessitating any explicit intention to

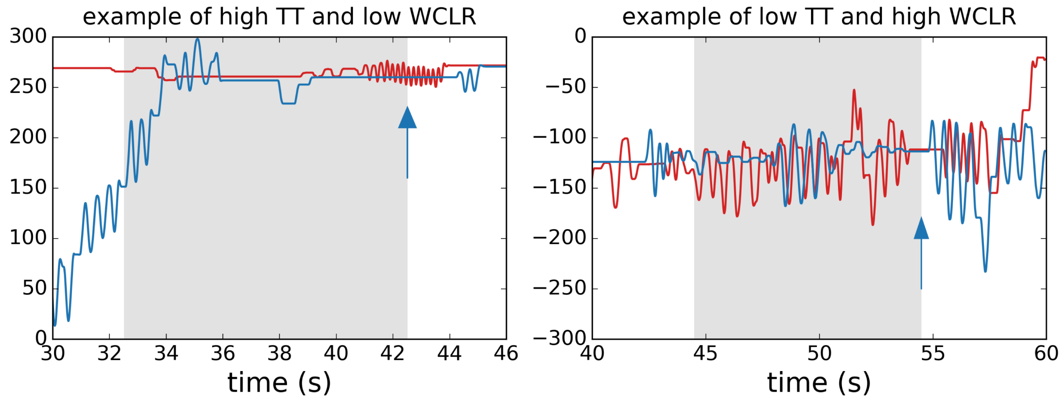

Figure 4.3: Illustrative comparison between turn-taking (TT) and windowed cross-lagged regression (WCLR) measures. For clarity we only plot the change in position of the avatars of players Pa and Pb (blue and red lines). Arrows mark the moment of player Pa's click. Taking turns by exchanging periods of movement and passivity does not entail a similarity between players' movement patterns (left, E7T6), and a similarity between players' movement patterns does not entail a turn-taking interaction (right, E17T15). The values for TT and WCLR for the 10 seconds before Pa's click (shaded regions) are 0.49 and 0.015 for the example on the left and 0.0 and 0.18 for the example on the right, respectively. The lag times that yielded these maximum WCLR values were -4.4 s (left) and -2.8 s (right).

coordinate behaviors or awareness that this is in fact occurring (Froese and Gallagher, 2012). In other words, conscious experience of social interaction cannot be reduced to objective measures of coordination; both subjective and objective aspects must be taken into account in an integrated manner.

Here, we compared each movement coordination measures with different PAS ratings (CC: 0.27 $\pm$ 0.01, 0.30 $\pm$ 0.01, 0.31 $\pm$ 0.01, WCLR: 0.084 $\pm$ 0.005, 0.095 $\pm$ 0.005, 0.11 $\pm$ 0.01, TT: 0.15 $\pm$ 0.01, 0.18 $\pm$ 0.02, 0.23 $\pm$ 0.01, with PAS 2, 3, 4, respectively), and we found that there was difference among different PAS ratings for all the objective measures. (ANOVA, F(2,337) = 4.969, $p < .01$, F(2,337) = 3.792, $p = 0.024$, F(2,364) = 8.178, $p < .001$ for CC,WCLR, and TT, respectively). Especially we found that all movement coordination measures accompanying with PAS 4 were higher than with PAS 2. (t(225) = 3.160, $p <$ .0.01, t(225) = 2.728, $p = 0.021$, t(242) =4.055,

$p < 0.001$ for CC, WCLR, and TT, respectively, with Bonferroni correction.) (Figure 4.4)

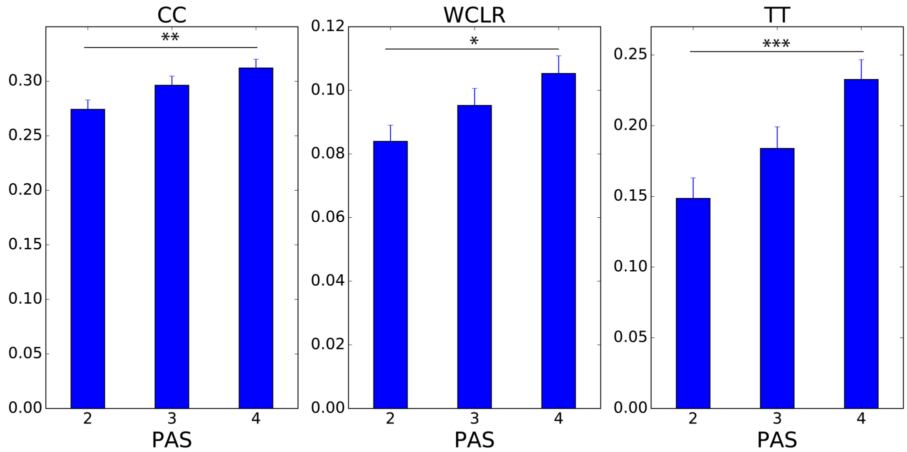

Figure 4.4: Three measures of interpersonal movement coordination and the clarity of the other's presence (PAS). We averaged the three measures of interpersonal movement coordination, cross-correlation (CC, left), windowed cross-lagged regression (WCLR, middle), and turn-taking (TT, right) scores for each PAS rating. The measures were applied to the 10 seconds preceding a click. $*$: $p < 0.05$, $**$: $p < 0.01$, $***$: $p < 0.001$ (With Bonferroni correction.)

This is an indication that these measures are characterizing some part of the real-time sensorimotor interaction signature of a clear experience of the other's presence. This seems to suggest that elevated levels of turn-taking and movement synchrony are a common feature of the transition to social awareness during genuine interactions.

4.3 Timescales of Movement Coordination

In order to learn more about the timescales in which synchrony of movements is most pronounced, we used WCLR to calculate the delay giving the highest cross correlation value for each trial. Since here we were not interested in which of the two players was leading the interaction, we took the absolute value of the lag times. We related these values with players' PAS ratings to determine whether some timescales are more relevant for explaining a clearer experience of the other's presence (Figure 4.5). Averaged lag time with each PAS rating were evaluated as 2.7 ± 0.1s, 2.5 ± 0.1 s, 3.0 ± 0.1 s. We found that the average lag time was significantly different among different PAS ratings($F(2,337) = 4.395$, $p = 0.013$), and especially we found that the lag time with PAS4 was significantly longer than that with PAS3 ($t(240) = 2.915$, $p = 0.012$, with Bonferroni correction). Those findings suggest that phenomenologically more salient forms of movement synchrony are based on a longer timescale of interaction.

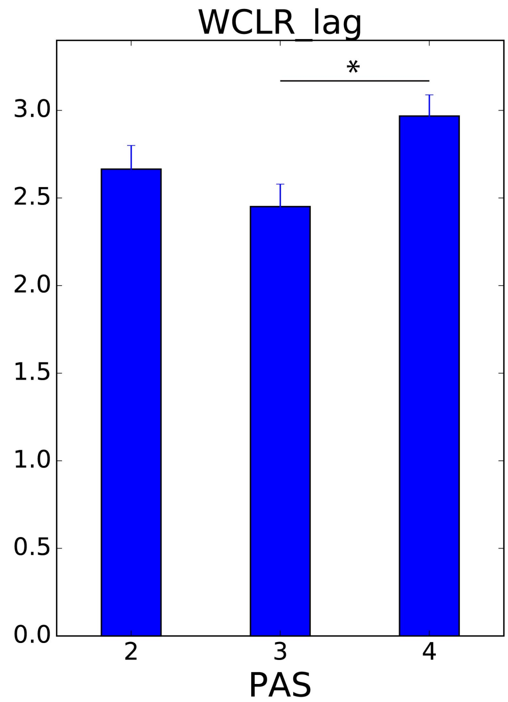

Figure 4.5: The timescale at which movement imitation is most pronounced (WCLR lag) and the clarity of other's presence (PAS). We calculated the size of the windowed time lag yielding the maximum value for windowed cross-lagged regression (WCLR) and averaged it over the same ratings of clarity of other's presence (PAS). * : $p < 0.05$ (With Bonferroni correction.)

4.4 Analysis of Direction of Influence

We used local transfer entropy (TE) to quantify directions of influence before and after a click at different timescales (a period of 10s, 5s, and 1 s before a click and 1 s, 5 s, and 10 s after a click), which meant that we first calculated local transfer entropy and take average over each period. The periods of 1 s and 10 s were chosen to coincide with the two cognitive scales of Varela's (1999) three scales of duration of the temporal horizon: (1) basic or elementary neural events (the '1/10' scale); (2) relaxation time for large-scale neural integration of cognitive or perceptual acts (the '1' scale); (3) descriptive-narrative assessments of the situation (the '10' scale). The 5 s scale was chosen as an intermediate scale that is consistent with the lag times used for the synchrony analyses. It is also of interest as an expression of cognitive events taking place at the '1' scale: spontaneous speech in many languages is organized such that utterances last 2-3 seconds and short intentional movements (such as self-initiated arm motion) are embedded within windows of this duration (Varela, 1999). I return to this point in the discussion. We denote S1 as the "self's" sensor time series and S2 as the "other's" sensor time series. Self and other are determined relative to the player who made the click. Who clicks first was not considered here. When one player touches the other, both sensors get activated at the same time (i.e. S1 = S2). They are only different (i.e. S1 $\neq$ S2) when either player touches the static objects or the shadows. Yet even though this means that Figure 4.6 and 4.7 are expected to return similar values for situations of perceptual crossing, we separate M1/2 $\rightarrow$ S1 (Figure 4.6) and M1/2 $\rightarrow$ S2 (Figure 4.7) for the sake of clarifying active/passive touch differences.

In support of the hypothesis of passive touch we found that the transition to perception of the other's presence was characterized by passively received tactile stimulation (Figure 4.6). In general, there tends to be more influence from the other's movements (M2) on the self's sensations (S1) compared with the influence of the self's movements (M1). However, we did not find a continuous period of passive touch (period >= 5s), a result that is consistent with our finding that significant lag times are < 3 s. A statistically significant heightened influence of M2 on S1 was only observed in the second immediately before the click and only for high PAS ratings (t(214.0) = 2.83 $p < .01$, t(279.9) = 3.54 $p < .001$ for PAS 3 and 4, correspondingly.).

This pattern is reversed after the click: moments of awareness rated as PAS 3 and 4 are followed by heightened influence of the self's movements on the self's sensations (i.e. comparatively more transfer entropy from M1 to S1 compared to M2 to S1). Moreover, this difference in influence only becomes significant after a few seconds and remains so until at least 10 s. (t(274.8) = 5.16 $p < 10^{-6}$, t(263.2) = 4.47 $p < 10^{-4}$, for 5 s and 10 s with PAS 4, respectively.)

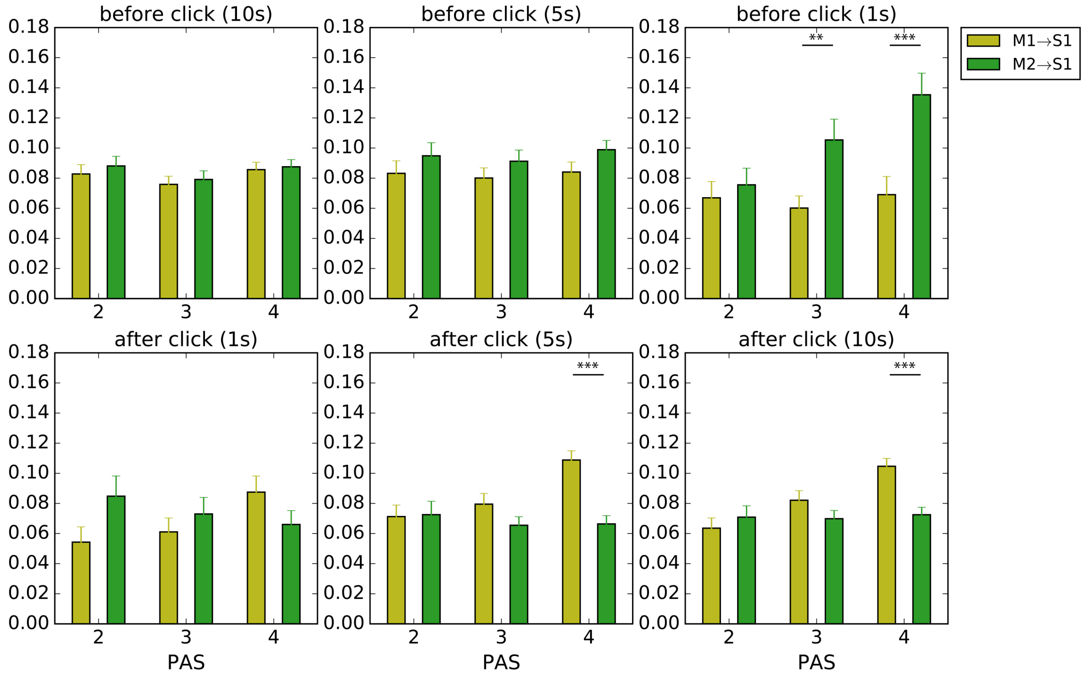

Figure 4.6: Average local transfer entropy to the self's sensations before and after clicks. We analyzed periods of 10 s, 5 s, and 1 s before a click and 1 s, 5 s, and 10 s after a click. Yellow and green bars indicate how much the movements made by the self (M1), i.e. the player who made the click, and by the other player (M2) contributed to the self's tactile stimulation (S1), respectively. Error bars represent standard error; significance was calculated using the Welch's t-test. * : $p < 0.05$, ** : $p < 0.01$, *** : $p < 0.001$

4.5 Analysis of Switch from Passive to Active touch

This post-click reversal in the flow of influences was unexpected. We considered two plausible explanations. On the one hand, this period of self-generated activity could be an example of reciprocity, in which the self now tries to make its presence clear to the other by providing them with an opportunity for undergoing passive stimulation in return, which would incidentally also stimulate the self's own sensor due to the situation of perceptual crossing. However, an alternative possibility, which is more in line with the illustrative trial shown Figure 4.1, is that a click marks the end of a bout of close interaction followed by a period of temporary disengagement, in which the movements of the other participant play only a diminished role for the self's sensations. We therefore redid the analysis shown in Figure 4.6, but this time focusing on the transfer entropy from the self to the other's sensations. The aim is to verify if the self's movements comparatively increase their influence on the other's sensations or not (Figure 4.7).

We found that in the periods leading up the self's click the other's movements dominated the other's own sensations (M2 $\rightarrow$ S2), which is to be expected if the self is mostly passive during the transition to social awareness (M1) and the other is actively moving (M2). After a click the situation becomes more complex. Following moments of clear awareness (PAS 4) the other's sensations are more influenced by the self's movements (M1 $\rightarrow$ S2), and significantly so in the longer post-click periods (t(276.8) = 3.93 $p < .001$, t(263.6) = 3.26 $p < .01$, for 5 s and 10 s with PAS 4, respectively.). This is consistent with the idea that the self returns the feeling of passive touch to the other, which from the self's perspective involves a transition from passive to active touch, but this possibility is more typical for clear awareness. After less clear experiences (PAS 2 and 3) there tends to be a stronger influence from the other's movements to other's own sensations, a trend especially notable for the immediate post-click period (1 s and 5s) and for the least clear experience (PAS 2). This is consistent with the idea that the self disengages from its interaction with the other after making a click, thereby leaving the other alone to generate their own sensations, but this decoupling is more typical for when the other's presence was not experienced sufficiently clearly.

Both of these situations can be confirmed in Figure 4.1, where the player that first disengages after the clicks (Pa) was also the one who gave a PAS

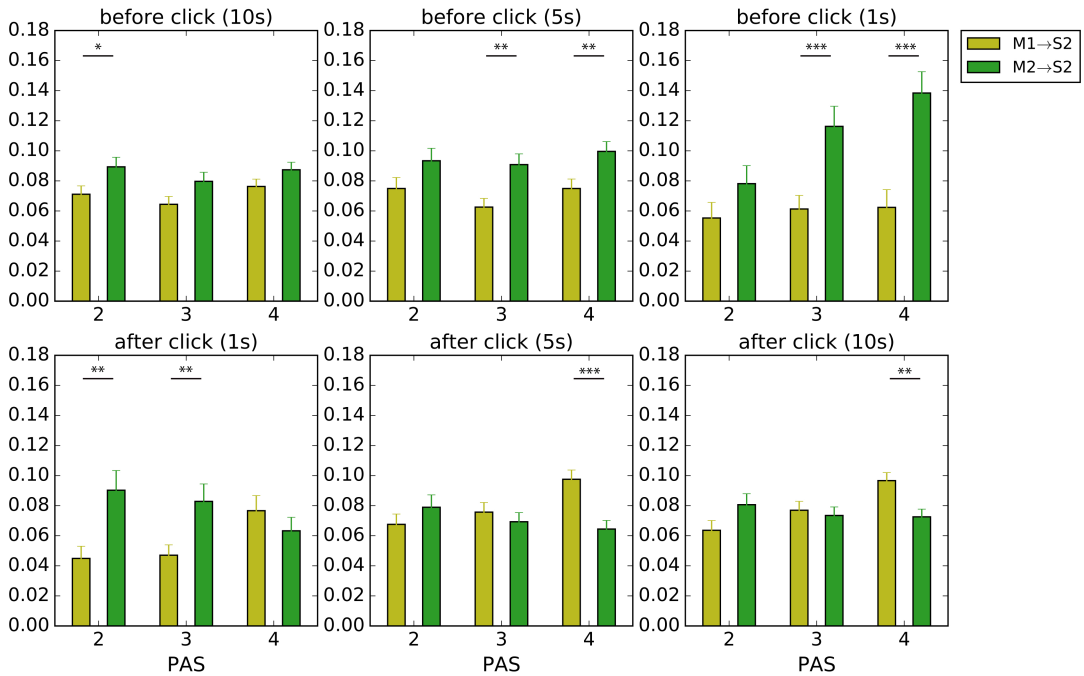

Figure 4.7: Average local transfer entropy to the other's sensations before and after clicks. We analyzed periods of 10 s, 5 s, and 1 s before a click and 1 s, 5 s, and 10 s after a click. Yellow and green bars indicate how much the movements made by the self (M1), i.e. the player who made the click, and by the other player (M2) contributed to the other's tactile stimulation (S2), respectively. Error bars represent standard error; significance was calculated using the Welch's t-test. $* : p < 0.05$, $** : p < 0.01$, $*** : p < 0.001$

score of only 2, while the other player who apparently would have continued interacting gave a PAS score of 4. These differences in the self's awareness-dependent post-click behavior, namely the transition from passive to active touch compared with relative disengagement, deserve attention in future studies.

Chapter 5

Discussion

In summary, we found that the feeling of the presence of the other, PAS, was related to the degree of synchronization, CC and WCLR, and the degree of turn-taking. Also, we characterized the information flow around the time of click using local transfer entropy and found the transition from passive to active around the click. This transition from passive to active was evident in the trial with the high PAS. In this chapter, we compare the present results to previous studies and discuss the implication of passive touch in wider contexts.

5.1 Stability of the Dynamics and Social Cognition

Murray and Trevarthen (1985) showed that infants realize the difference between live and recorded video images, and only in the live condition the interaction between infants and mothers is maintained. Auvray et al. (2009) designed perceptual crossing experiment as the minimal parallel of Murray and Trevarthen (1985)'s experiment and showed that also in this setup subjects tended to click more with the other agent than with the shadow object, and the subjects spent more time with the other agent than with the other objects.

However, the probability of clicks, which was the number of clicks divided by the time duration of the stimulation, was found not qualitatively different between the avatar and the shadow. It means that subjects cannot recognize the difference from the dynamics pattern, and the reason for the larger number of clicks for avatars than for shadows was the contact between agents were tended to be maintained for a longer time. Therefore, Auvray et al. (2009) concluded that the difference between the avatar and the shadow came only from the stability of the dynamics (Auvray and Rohde, 2012). The problem here is that if we assume that the stable dynamics is the condition to be recognized as the avatar, the static object also satisfies this condition because this can also provide stable dynamics. This is conspicuous in the simulation study (Iizuka and Di Paolo, 2007; Di Paolo et al., 2008), in which the evolved agent found it hard to distinguish between the avatar and the static object rather than between the avatar and the shadow. Therefore, the stability of the dynamics only cannot fully explain the situation.

In the Froese et al. (2014a)'s version of perceptual crossing experiment, the subjects were actually able to distinguish between the agent and the shadow, which means that the probability of the click was also significantly different between these two. This clearly showed that the subjects noticed the difference between the agent and the shadow in this experiment. Also, we found that the movement of the subjects tended to synchronize with each other measured by cross correlation and WCLR, and co-create the temporal structure, turn-taking. This showed that each subject did not simply react to the input stimulus, but gradually change their movements to adapt with each other. The result of local transfer entropy further confirmed that the ability to distinguish between the agent and the shadow was not just a consequence of dynamical stability, because the time of click was characterized by the transition from passive touch to active touch, and they did not randomly click during the interaction.

These results might depend on the difference in the experimental settings between Auvray et al. (2009) and Froese et al. (2014a). One of the main differences from the original version (Auvray et al., 2009) was to encourage the subjects to cooperate to achieve the tasks, and this might enhance the synchronization of the movement and the turn-taking structures. Another main difference was that Auvray et al. (2009) allowed multiple clicks during each trial, where as Froese et al. (2014a) allowed only once. The results of Auvray et al. (2009) did not distinguish the click associated with the strong feeling of the other's presence and with the weak feeling, where as in Froese et al. (2014a), the subjects were allowed to click once, so they clicked when they were confident on the judgement. Therefore, the difference between the results from Auvray et al. (2009) and Froese et al. (2014a) might be the judgement with low confidence and high confidence. Our result that the characteristic values such as cross correlation, WCLR, TT, and local transfer entropy were less evident with less PAS, the subjective feeling of the other's presence, well accorded with this interpretation. For the rough detection of agency, the stimulus input not caused by the self might be enough, but for the strong sensation of the other's presence, the co-regulated structure with the other might be required.

5.2 Time scale in the interaction

Varela (1999) suggested the three scales of duration of the temporal horizon: (1) basic or elementary neural events (the '1/10' scale); (2) relaxation time for large-scale neural integration of cognitive or perceptual acts (the '1' scale); (3) descriptive-narrative assessments of the situation (the '10' scale). Here, we discuss the timescales appeared in our experiments based on these three scales.

The shortest time scale was the '1/10' scale, ∼100 ms. This corresponds to the window size (50 ms) used in local transfer entropy analysis. The local transfer entropy detect the influence between the two time series, so the window size corresponds to the temporal width treated as simultaneous. In our experiment, we used the number of past values as the information source of the present value as $k = l = 4$, which mean that the analysis detect the possible influence within past 200 ms. This time scale accord with the time scale in the study of sense of agency (Blakemore et al., 1999; Farrer et al., 2013). The next time scale was the '1' scale, ∼1 s. This time scale roughly corresponds to the interval of touch, the rise of local transfer entropy from the other's movement before click, and the size of turn ,which was typically around several seconds. Therefore, in this experiment, this scale correspond to the time interval of signals that one subject emits to the other. The longest time scale was the '10' scale, ∼10 s. This time scale did not explicitly appear in our result, but the exchange of several turns amounts to this scale, so this scale correspond to the time scale of the co-regulation structure between the dyads.

5.3 Passive Touch

We found that the local transfer entropy from the movement of the other to the sensation of the self was high preceding the clicks. This local transfer entropy can be interpreted as "Passive touch", in contrast to "Active touch", which corresponds to the transfer entropy from the movement of the self to the sensation of the self. Passive touch and the perception of others were also discussed in the context of phenomenology (Miyahara, 2015). He argued that passive touch constitutes the experience of the perception of others from a phenomenological point of view. In particular, he insists on the importance of passive touch in detecting agency.

The importance of the passivity was also discussed in the context of developmental studies, related to the question, when infants can be aware that someone is attending to them. From the standpoint of cognitive developmental psychology, this becomes possible after the formation of representation of others, which happens around 12 months, and representation of self, which occurs around 18 months. On the other hand, many studies have shown that infants of about 2 months emotionally react to attention to self (Muir and Hains, 1999; Nadel and Tremblay-Leveau, 1999; Reddy, 2000). From these findings, Reddy (2003) argued that the awareness of others' attention came first, and after that representation of other and self was formed as an object being attended. This is parallel to our results that the perception of the other is based on the directed touch from the other towards self, "passive touch".

These arguments suggested that passive experience is important in social cognition. Passive means something coming towards me, so this is the concept with direction. Therefore, the transfer entropy, which is an information measure with direction, is suitable to quantitatively evaluate this aspect, and we confirmed that the passive touch is qualitatively related to the subjective experience of feeling the other's presence. As far as I know, this is the first time to apply local transfer entropy analysis to social psychological experiments, and in this sense, this is the first study to quantitatively evaluate the passive and active information transfer during the social interaction.

Part III

TypeTrace Messenger

Chapter 6

Introduction

Conversations are central to our social lives. In Face-to-face (FtF) circumstances, social interaction includes not just the exchange of verbal sentences, but also interactions with non-verbal means such as body gestures, vocal cues, temporal structures in speech like turn-taking, facial expressions, and gaze exchanges. It is known that the medium of communication affects, among other aspects of social interaction, affiliative behaviors and the resulting outcomes (Sprecher, 2014).

Modern societies have become inundated by computer-mediated communication (CMC) systems. Since the early introduction of personal computers in the 1980s until the universal dissemination of smartphones in the 2010s, we have experienced a drastic influx of new CMCs. The lineage of CMC has diversified since the age of simple text-based chats, with the burgeoning of audiovisual teleconferencing systems and virtual, augmented, and mixed reality devices.

Nevertheless, text-based CMC remains one of the dominant communication channels of today. For instance, as of 2019, among the 3.2 billion people worldwide who own a smartphone (Statista, 2020), 1.6 billion use WhatsApp, 1.3 billion use Facebook Messenger, and 1.1 billion use WeChat (Clement, 2019) (all monthly active users). By contrast, in 2019, less than 10 million people worldwide owned a virtual reality device (Statista, 2018). All of these messenger applications include rich media functions such as making online video or audio calls and sending high-resolution images or audio files, but text-based chats also remain widely used among its users, albeit the exact statistic is unknown.

Short et al. initially introduced the term "social presence" in the context of telecommunication (Short et al., 1976) and conceptualized the ability of communication media to transmit social cues. Social cues consist of both verbal and non-verbal information, such as facial expressions, gestures, and physical appearance, and they serve to construct the "sense of being with another" (Biocca et al., 2003). Short et al. (1976) considered these cues the foundation of intimacy (feeling of connectedness to the partner) and immediacy (psychological distance to the partner). In the early period of research, social presence was viewed as a variable depending on the quality of media. In short, telecommunication generally was regarded as a lesser communication channel compared to the FtF communication because of its inability to transfer non-verbal cues.

This line of research continued to gain support in the 1980s, where Daft and Lengel conducted seminal analyses of the richness of information and media (Daft and Lengel, 1984, 1986). However, in the 1990s, Walther argued in opposition to media richness theories, stating that if enough time is spent on CMC, interactants can achieve a level of interpersonal relationship as high as FtF communications (Walther, 1992). Walther later argued that communication means specific to CMC could even create a more robust social bond compared to FtF (Walther, 1996) because it stimulates more self-disclosure than FtF and thus can lead to higher social attraction. This school of thought, called the social information processing theory (SIPT) (Walther, 2015), has helped researchers to transcend the simple dichotomy of rich and poor media and to scrutinize the social phenomena in CMCs more in-depth.

Since then, theoretical discussions of social presence concerning CMC faced the need to reconcile technologically mediated social interaction with unmediated interaction (Biocca et al., 2003). In remote learning environments, social presence has been measured to predict participants' learning satisfaction (Gunawardena and Zittle, 1997), and comparisons of different types of CMC have been analyzed (Tu, 2002). Antheunis et al. (Antheunis et al., 2010) conducted a thorough quantitative analysis of an Social Networking Service (SNS) based on a hypothesized model of social presence theories and found supportive evidence of SIPT. A recent systematic review of mediated social presence research by Oh et al. (Oh et al., 2018) marshaled different previously studied predictors of social presence, including new media such as virtual and augmented reality systems, while pointing out that social presence does not always lead to positive outcomes. The relationship between social presence and the valence of communication is yet to be further elaborated.

Moreover, the dominant trend in social presence research so far has been to treat only the change in the characteristics of each message, which is uni-directional. On the other hand, in the field of social cognition, researchers have argued that bi-directional interaction plays a central role in understanding dyadic interactions (Schilbach et al., 2013; Gallotti et al., 2017; Redcay and Schilbach, 2019).

In this study, we investigated these relationships by recording and analyzing the dyadic bi-directional interaction of CMC. We particularly focused on the temporal dynamics of interaction and each interlocutor's response during several types of text chat systems. In order to find evidence for potential factors that contribute to the generation of social presence in a dyadic CMC setup, we formulated the following two questions and designed our series of experiments accordingly.

First, how does the increase of informational richness affect interactions in CMC? The pre-SIPT line of theories predicted that the lack of social cues such as facial expressions would decrease social presence. However, neither SIPT nor later research rigorously measured such richness of information in text-based CMC. Secondly, how does the concurrency of information exchange between the interactants influence the dynamics of a CMC interaction? Past social presence research often mixed synchronous and asynchronous CMC such as chat, e-mail, and teleconferencing. In our study, we specialized in synchronous text chat in order to observe results varying on the difference of information concurrency. We employ transfer entropy to measure such degree of information concurrency.

Our analysis of keystroke dynamics focused on the coupling between the two subjects of text chat. To capture the bi-directional aspect of the text chat, analysis of time-series data of dyadic interaction is required. In this direction, some studies characterized temporal dynamics using some measures such as recurrence quantification analysis (Fusaroli and Tylén, 2016) and the Allan factor (Kello et al., 2017). In this study, we used transfer entropy, which is a measure in information theory used to detect information flow between two time-series data (Schreiber, 2000). In the previous sections, we analyzed the dyadic interactions in perceptual crossing experiment, which consisted of a minimal CMC that only involved a vibration device and a computer mouse, using the local form of transfer entropy (Lizier et al., 2008), and we found that passive information flow was related to the feeling of the presence of the others. Here, we measured changes in the amount of transfer entropy between the four conditions of our experiment and also in relation with the phone call data set.

Chapter 7

Methods

7.1 TypeTrace Messenger

TypeTrace is a software that records the entire typing processes of writing and replays it by varying the font size as a function of writing speed of each letter (i.e., the font size becomes larger when there is a slower writing speed). The software has also been used for a quantitative analysis of a professional creative writer's process of writing a new novel (Kudo et al., 2015). TypeTrace software has been demonstrated at several art exhibitions (e.g., Aichi Triennale 2019 exhibition).

We here developed a new TypeTrace Messenger (TT Messenger) based on the previous versions of TypeTrace. TT Messenger is a Web application that enables users to take part in dyadic chat online on PC browsers. We use Google Firebase for the backend system, and the software runs on modern Chrome browsers. We wrote the software in JavaScript and recorded typing data in the JSON format.

TT Messenger records all key typing during a chat session and is capable of precisely reenacting each typing action. This playback includes all the processes of typing, such as pauses, corrections, and deletions.

TT Messenger has four different conditions (Fig.7.1):

1. It looks like a regular online chat system. Before the partner sends a message, the recipient can only see a dotted line, which shows that the partner is typing something. When the partner sends the message, the recipient can see it as a static text.

2. The recipient sees the partner's message in a dynamic playback (dynamically presenting the playback of the other' s text message typing) as soon as she receives it. Therefore, recipient has to wait until the playback finishes in order to see the resulting final message. We designed this setup to consider our first question on the richness of information exchanged between interactants.

3. Just like in the second condition case, the messages play back as soon as they are sent, but with an additional visual effect. The software records the duration taken to type every word and changes each word' s respective font sizes as it plays them back. For instance, when a user takes three seconds to finish typing a word, that word would appear with a bigger font size than the previous word that took only one second to type. We added this effect in order to visualize the rhythm of the typing. We hypothesized that this additional social cue would have a comparable effect with facial expressions and body gestures in FtF communication. We expected the results from this condition would shed light on our first question about information richness.

4. The text chat becomes concurrent, and it works in real time. As soon as the partner starts typing, the process is transmitted to the recipient' s screen in real time, even without the partner sending the text. The subjects can send the message at any moment, but they do not have to. The two parties can simultaneously type, and each other's messages are displayed at the same time. We designed the fourth condition to examine our second question on information concurrency.

In the following sections, we explain how we used these four configurations for our experiment, in which we asked subjects to freely converse with each other and compared the results of subjective reports, physiological markers, and patterns in keystroke events among the four settings. We recruited the subjects pairs who were acquainted with each other, so our experiment can be regarded as an experiment of "Mediated social presence of (very) familiar interactions" in the classification of Biocca et al. (2003).

7.2 Participants

Participants were healthy volunteers recruited from acquaintances at Waseda University (N = 18). They were all Japanese nationals; 11 were female, and the median age was 22 years old. All pairs were already acquainted before the experiments. The study protocol was approved by the local ethics research committee of Waseda University (Ethics Review Procedures concerning Research with Human Subjects; Application Number:2018-273; Approved on 25th of January, 2019), and the methods were carried out according to the ethics committee guidelines and regulations. All of the participants gave their written informed consent before taking part in the study.

7.3 Experimental Procedures

Two participants were placed in different rooms. Each were provided with a laptop PC, and we asked them to freely converse with each other through TT Messenger. We did not set a theme for the conversations. For each trial, we asked the pairs to converse for 10 minutes and to answer questionnaires after that. The experiments consisted of two rounds of four sessions, and each session included every condition (1 to 4) of TT Messenger in random order.

During each session, we recorded the keystroke events, galvanic skin response (GSR), and facial expressions. An example of the keystroke timeseries data is shown in Fig.7.2.

7.4 Subjective Reports

We used a five-point Likert scale to estimate the subjective rating of the degree of nervousness, enjoyment, closeness, presence of the other, and time delay. Actual questionnaire items in Japanese and English translations are listed below.

- I got nervous during the conversation. (相手との会話に緊張した)
- I had an enjoyable conversation. (会話がはずんだ)
- I felt close to the partner. (相手との距離が近く感じられた)
- I strongly felt the presence of the partner. (相手の存在感を強く感じた)
- I felt the time delay when exchanging messages. (メッセージのやり取りに時間がかかったと感じた)

We also asked each subject to report the Inclusion of Other in the Self (IOS) scale before the experiment and after each trial. The IOS scale has been used to measure the subjective closeness to others and is known to correlate well to other subjective markers of interpersonal closeness (Aron et al., 1992).

7.5 Measurements

During the experiments, we recorded galvanic skin response (GSR) by Shimmer GSR sensors and facial expressions by the web camera mounted on the computer, which were later analyzed by OpenFace (Baltrušaitis et al., 2016) to extract action units (AU). We measured GSR of one subject from the pair and switched to the other subject on the second round of experiments. Facial expressions were simultaneously recorded from both subjects. Keystroke events are collected through TT Messenger.

7.6 Transfer Entropy

Transfer entropy (Schreiber, 2000) from time series process Y to X is formulated using conditional mutual information as

$$T_{Y \to X} = I(\mathbf{Y}_{n+1}^{(l)}; X_{n+1} | \mathbf{X}_n^{(k)})$$

, where $\mathbf{X}_n^{(k)} = \{X_{n-k+1}, ..., X_n\}$, $\mathbf{Y}_n^{(l)} = \{Y_{n-l+1}, ..., Y_n\}$ (k: target history length, l: source history length).

Effective transfer entropy (Marschinski and Kantz, 2002) is calculated by subtracting the mean value of null distribution of transfer entropy, which is constructed by calculating the transfer entropy with a resampled surrogate source time series, from the original transfer entropy. We calculated the effective transfer entropy between subjects' keystroke event time series (or phoneme event time series for the phone call data) downsampled to 100ms windows. We used JIDT (Lizier, 2014) for the calculation, and we set $k = l = 2$.

7.7 Phone call dataset

In order to compare the keystroke dynamics with the dyadic dynamics that have different modalities, we used a conversation corpus, CallFriend (Yaeger-Dror, 2004), which consists of telephone conversations data in Japanese.

From this corpus, we used the audio data and age information of 10 conversations (file id were ja_4044, ja_4261, ja_6221, ja_6228, ja_6688, ja_6698, ja_6700, ja_6707, ja_6717, ja_6739). The median age was 32 years old, and we used the first 10 minutes from each audio file.

For the extraction of phoneme events from audio data, we applied the phoneme segmentation method by Ziółko et al. (2006). This method is based on a six-level discrete wavelet transform (DWT) analysis, and it detects the boundary of phonemes as the time of rapid change in each subband power. We used the sym6 wavelet and set a minimal threshold of subband DWT power, $p_{\min}$, to 0.005. The other parameters were kept the same as in the original paper.

We used the boundaries of the phoneme segmentation as the phoneme events' time series, comparable to the keystroke events for TT Messenger data, and applied the same analysis to the event sequences.

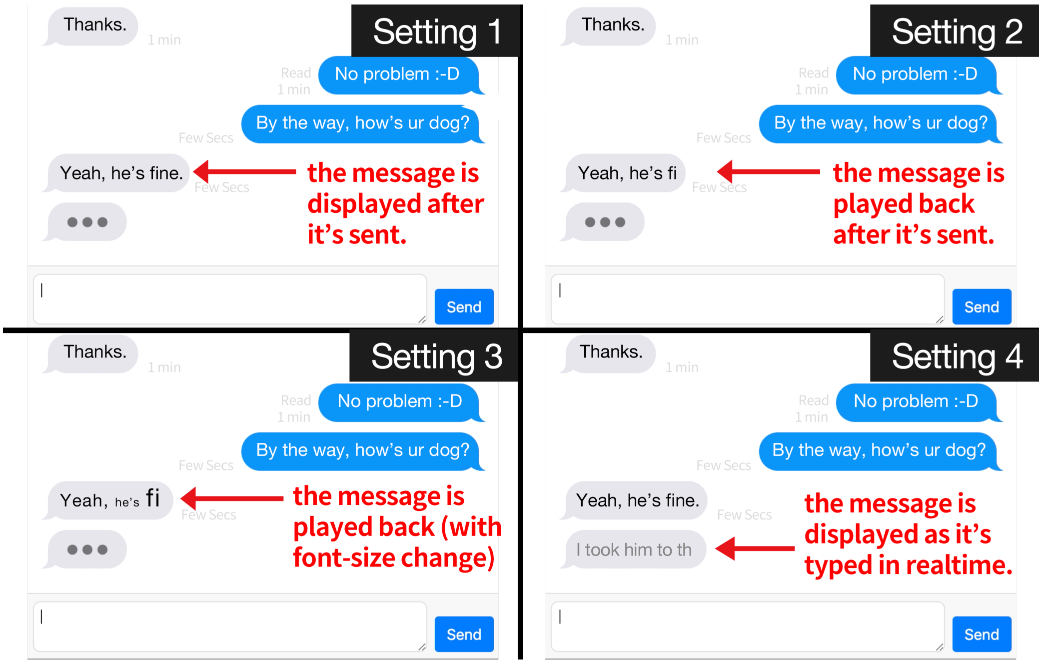

Figure 7.1: Screenshots of the 4 different conditions of our chat system, TypeTrace Messenger. The actual chats in our experiments were conducted in Japanese, but we created this figure with English texts for explanation purposes. In condition 1, the messages are displayed statically, which corresponds to a regular online chat system. In condition 2 the whole process of typing is dynamically displayed, not just the static messages. In conditions 2 and 3, the font size of the messages changed according to the time to type that message. In condition 4, the content the subjects are typing is simultaneously shown in the other's display.

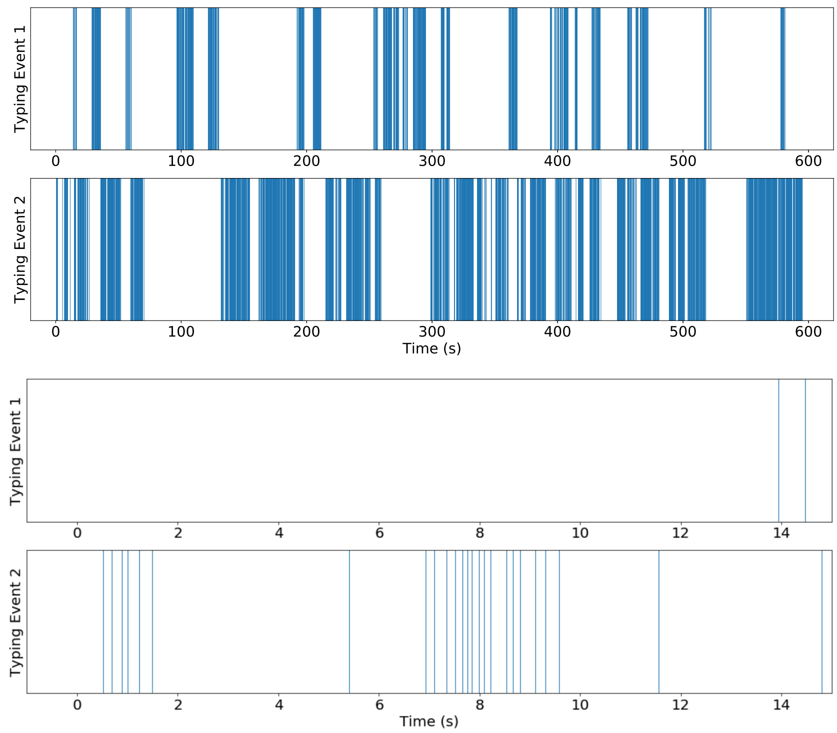

Figure 7.2: Example of the keystroke data from two subjects in one trial. Each vertical line corresponds to one keystroke event. Top: The keystroke events of whole trial (10 min). Bottom: The keystroke events of first 15 s.

Chapter 8

Results

Below, we report the results from subjective reports, physiological markers, and keystroke dynamics, and compare among different conditions of the chat system and telephone conversation data. If not otherwise stated, we used the Friedman test for statistical testing and the Nemenyi test for post-hoc testing.

8.1 Subjective Reports

First, we investigated the subjective reports after each session. The histogram of ratings for each item in different conditions of TT Messenger is shown in Fig.8.1.

We found that in condition 4, the rating of Enjoyment was significantly higher than it was in condition 2 ($p < 0.05$) and condition 3 ($p < 0.05$), the rating of Closeness was significantly higher than it was in condition 3 ($p < 0.05$), and the rating of Presence was significantly higher than it was in condition 1 ($p < 0.05$). The rating of Time delay was significantly smaller in condition 4 than it was in conditions 1 ($p < 0.05$), 2 ($p < 0.05$), and 3 ($p < 0.05$). No significant difference was found in the rating of Nervousness.

Also, we investigated the change in IOS before and after each trial. We found that the percentage of positive change was 22%, 14%, 22%, and 44% in conditions 1, 2, 3, and 4 respectively, but there was no significant difference among these conditions ($p = 0.06$).

8.2 Physiological Markers

In order to confirm the result from subjective reports, we also recorded physiological markers. Here, we used GSR and facial expressions extracted by OpenFace (Baltrušaitis et al., 2016) to recognize the emotional state of the subjects.

8.2.1 GSR

We recorded GSR, which is related to states of arousal (Dawson et al., 2017), during each trial. We calculated the median value from the time series and subtracted the initial value to characterize the amount of increase of GSR during each trial.

The median values of GSR from all trials were $4.7 \times 10^{-3} \mu$S, $4.7 \times 10^{-2} \mu$S, $-3.0 \times 10^{-3} \mu$S, 0.16μS for conditions 1, 2, 3, 4, respectively (Fig.8.2). Also, we found the GSR values from condition 4 were significantly higher than the values from conditions 1 ($p < 0.005$) and 3 ($p < 0.05$).

8.2.2 Cheek Raiser (AU6)

We recorded facial expressions with web cameras during each trial and analyzed using OpenFace (Baltrušaitis et al., 2016). OpenFace extracted the elementary facial motion unit, action unit (AU). We used AU6 (cheek raiser), which is related to the feeling of happiness (Ekman, 1997; Sato et al., 2019).

The median values of AU6 from all trials were 0.13, 0.23, 0.15, 0.21 for conditions 1, 2, 3, 4, respectively (Fig.8.2). Also, we found that the values of AU6 from condition 4 were significantly higher than the values from condition 1 ($p < 0.005$) and 3 ($p < 0.05$).

8.3 Keystroke Dynamics

Dyadic conversations were characterized by synchrony of utterances and the turn-taking patterns. We quantified them in each of the four conditions to study the differences.

8.3.1 Synchronization in Typing Patterns

In order to quantify the synchrony in typing patterns, we used two measures, Jensen-Shannon divergence (JS-divergence) between histograms of inter keystroke intervals (IKSIs) and correlation coefficient in the medians of IKSIs.

First, we used JS-divergence of the IKSI histograms between subjects in pairs to measure the dissimilarity in typing patterns in each trial and compare them among different TT Messenger conditions . The median values of JS-divergence were 0.015, 0.034, 0.026, and 0.018 for conditions 1, 2, 3, and 4 respectively (Fig.8.3), and there was no significant difference among these conditions ($p = 0.5$).

Secondly, in order to measure the degree of synchronization in typing speed during each trial, we split each trial into 1-min windows, calculated the median values of IKSIs of each subject for all ten windows, and calculated the correlation coefficient of the median values between the two subjects. The median values of results from all pairs were -0.07, -0.1, -0.04, and 0.07 for conditions 1, 2, 3, and 4 respectively (Fig.8.3), and there was no significant difference among these conditions ($p = 0.4$).

8.3.2 Pattern in Turns

In order to characterize the global typing patterns, we analyzed the pattern in the chunk of keystroke events (which we call turns) as follows. We identified each turn by chunking a keystroke event within the threshold interval, which we set to 2 s. We used median size of turns (sec), number of turns, total time of turns (sec), and overlapping ratio between two subjects in each trial to characterize the turn structure (Fig8.4).

We found no significant difference among the median size of turns (the median values were 3.0 s, 2.6 s, 2.8 s, and 2.5 s for conditions 1, 2, 3, and 4 respectively $p = 0.9$) and overlapping ratios (the median values were 0.24, 0.22, 0.25, and 0.20 for conditions 1, 2, 3, and 4 respectively, $p = 0.2$).

On the other hand, the median values of the number of turns were significantly different for those conditions. They were 48.5, 44.0, 44.5, and 57.0, for conditions 1, 2, 3, and 4 respectively, and the numbers in condition 4 were significantly higher than those in conditions 2 ($p < 0.05$) and 3 ($p < 0.001$). Also, the median values of total time for typing were 2.1×10^2s, 1.7×10^2s, 1.7×10^2s, and 2.0×10^2s for conditions 1, 2, 3, and 4 respectively. The numbers in condition 4 were significantly higher than those in conditions 2 ($p < 0.01$) and 3 ($p < 0.01$).

8.4 Information flow between keystrokes of partners

A second remarkable aspect of dyadic communication is the direct perception of the other's presence. We assume that when a subject's utterance is more driven by the other, the sense of presence increases. In such moment, the subject becomes less autonomous and more passive. The sense of passive awareness becomes the source of producing the presence of others (Kojima et al., 2017). This point will be revisited later.

We used effective transfer entropy (Schreiber, 2000; Marschinski and Kantz, 2002) to measure the information flow between subjects' keystroke events. We downsampled the keystroke event time series to a 100ms window and calculated effective transfer entropy with $k = l = 2$.

The median values of effective transfer entropy in each condition were 1.2×10^{-3}, 3.7×10^{-4}, 5.5×10^{-4}, and 2.3×10^{-3} for conditions 1, 2, 3, and 4 respectively (Fig.8.5), and the values in condition 4 were significantly higher than those in condition 1 ($p < 0.001$), condition 2 ($p < 0.001$), and condition 3 ($p < 0.001$).

We also calculated the transfer entropy using the keystroke events excluding the events of pressing return keys, which were used to confirm or send messages and not to produce the sentences. (Fig.8.6) We found that the effective transfer entropy tended to decrease without return key events, but the transfer entropy in condition 4 was still significantly higher than condition 1 ($p < 0.05$), condition 2 ($p < 0.001$), and condition 3 ($p < 0.001$). This implies that the timely responses to the sent messages were partly responsible for the value of transfer entropy, but also other types of interactions occurred especially in condition 4.

8.5 Comparison to phone call dynamics

So far, we have analyzed the chat data obtained from our chat system, TypeTrace Messenger. In order to compare these results with different types of dyadic interactions, we also used publicly distributed telephone conversation data from CallFriend corpus (Yaeger-Dror, 2004) and analyzed the data in the same way as we did to our chat data.

A phone call is not a CMC per se, but it is still omnipresent in modern societies and is a common feature included in many CMC applications. At the same time, although the modality differs radically between voice and text, a phone call resembles our experimental settings of text chat where participants are separated in different locations and converse without non-verbal social cues such as facial expressions and eye gazes. In both phone calls and text chats, participants spontaneously take turns, with overlaps in their utterances. However, we did not compare with FtF conversation because the structure of interaction differs even more substantially between FtF and text-chat.

CallFriend consists of sound data of actual telephone conversations and their scripts. We analyzed phone call sound data of 20 Japanese individuals (Yaeger-Dror, 2004). For pre-processing, we first extracted phoneme events from audio file (Ziółko et al., 2006) and analyzed these phoneme events' time series in the same manner as above.

First, we analyzed the turn structure of the telephone conversation (Fig.8.4). The median values of the median size of turns, number of turns, total time of turns, and overlapping ratio were 1.3s, 121.5, 2.9×10^2s, and 0.25 respectively. For statistical testing, we performed a Kruskal-Wallis test with a Mann-Whitney U test as post-hoc, and we found that median size of turns was significantly smaller than that of our chat data with every condition ($p < 0.001$), and the number of turns and total time of turns were significantly longer than that of our chat data ($p < 0.001$, except between total time of turns in with condition 4 and the telephone conversation, $p < 0.01$).

Secondly, we calculated effective transfer entropy between phoneme events timeseries from dyads. We downsampled the phoneme event time series to a 100ms window and calculated effective transfer entropy with $k = l = 2$, in the same way as in the chat analysis. The median value of the effective transfer entropy of the phoneme events from a telephone conversation was 0.014 (Fig.8.4). For statistical testing, we performed a Kruskal-Wallis test with a Mann-Whitney U test as post-hoc and found that the transfer entropy of the telephone conversation was significantly higher than the transfer entropy of our chat data with every condition ($p < 0.001$).

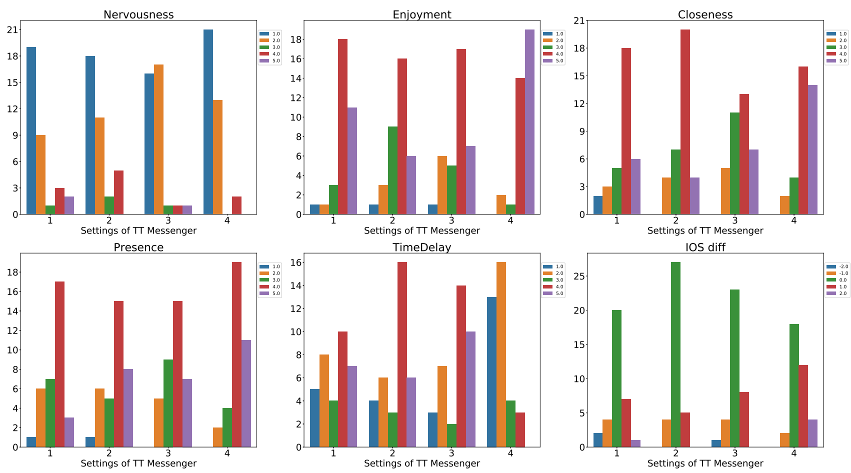

Figure 8.1: Histograms of subjective ratings in each condition of TT Messenger for Nervousness, Enjoyment, Closeness, Presence, Time Delay and IOS Change (from top left to bottom right). Enjoyment was significantly higher in condition 4 than 2 ($p < 0.05$), and 3 ($p < 0.05$), Closeness was significantly higher in condition 4 than 3 ($p < 0.05$), Presence was significantly higher in condition 4 than 1($p < 0.05$), and Time Delay was significantly smaller in condition 4 than 1 ($p < 0.05$), 2 ($p < 0.05$), and 3 ($p < 0.05$). No significant difference was found in the rating of Nervousness ($p = 0.4$) and IOS ($p = 0.06$).

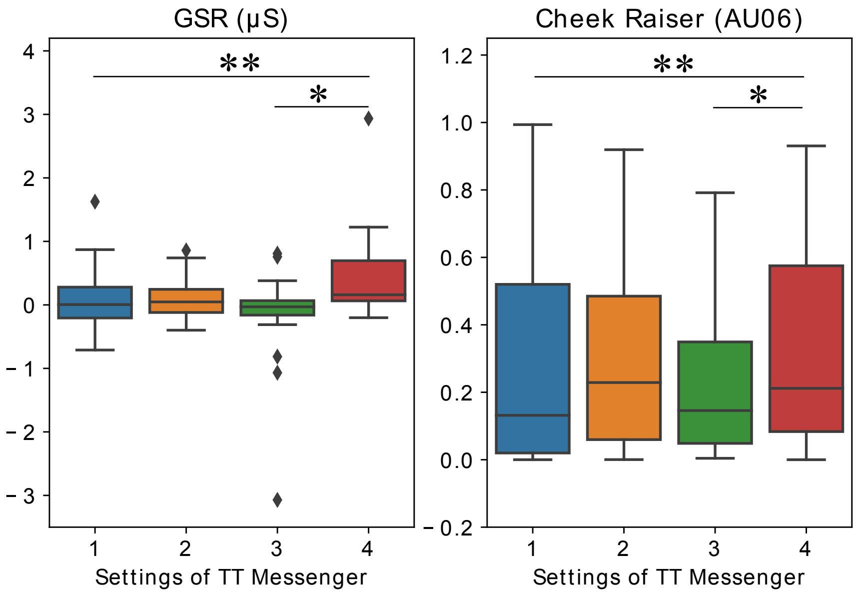

Figure 8.2: Results of physiological markers. Left: The median value of GSR during each trial for different conditions of TT Messenger. Right: The median value of AU6 (Cheek Raiser) during each trial for different conditions of TT Messenger. ($* : p < 0.05$, $** : p < 0.01$)

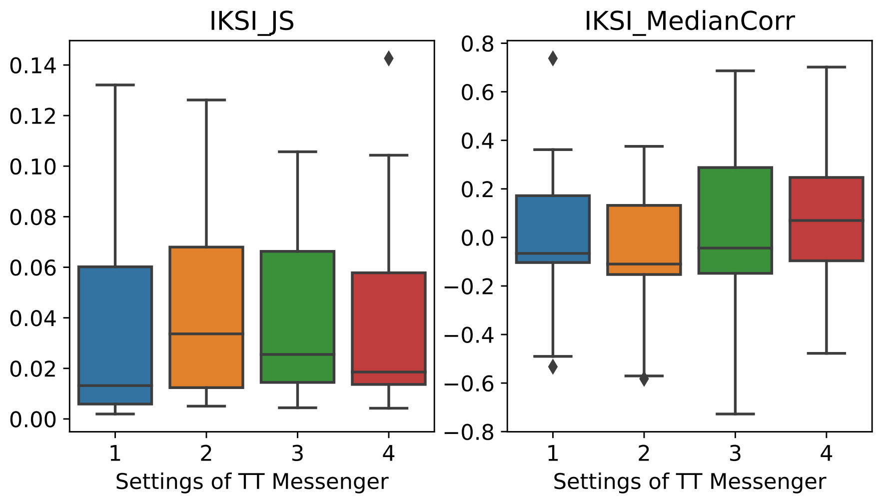

Figure 8.3: Two synchronization measures between subjects' keystroke event timeseries. Left: JS-divergence between two subjects' IKSIs histograms of each trial for different TT Messenger conditions. Right: Correlation coefficient between IKSIs among 1-min segments in different TT Messenger conditions.

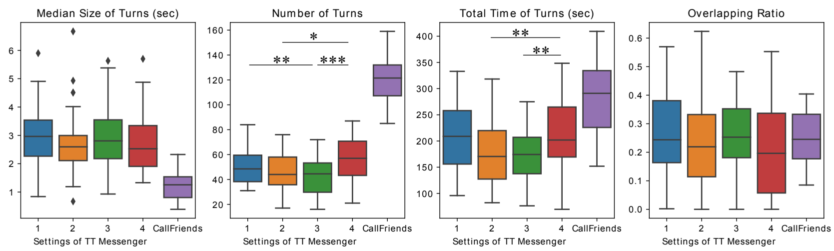

Figure 8.4: Four measures to characterize the structure of turns, which are identified from chunking keystroke event timeseries, median size of turns (sec), number of turns, total time of turns (sec), and overlapping ratio between dyads. Each measure was calculated using different TT Messenger conditions and phoneme timeseries data obtained from the telephone conversation dataset, using CallFriend.($* : p < 0.05$, $** : p < 0.01$, $*** : p < 0.001$)

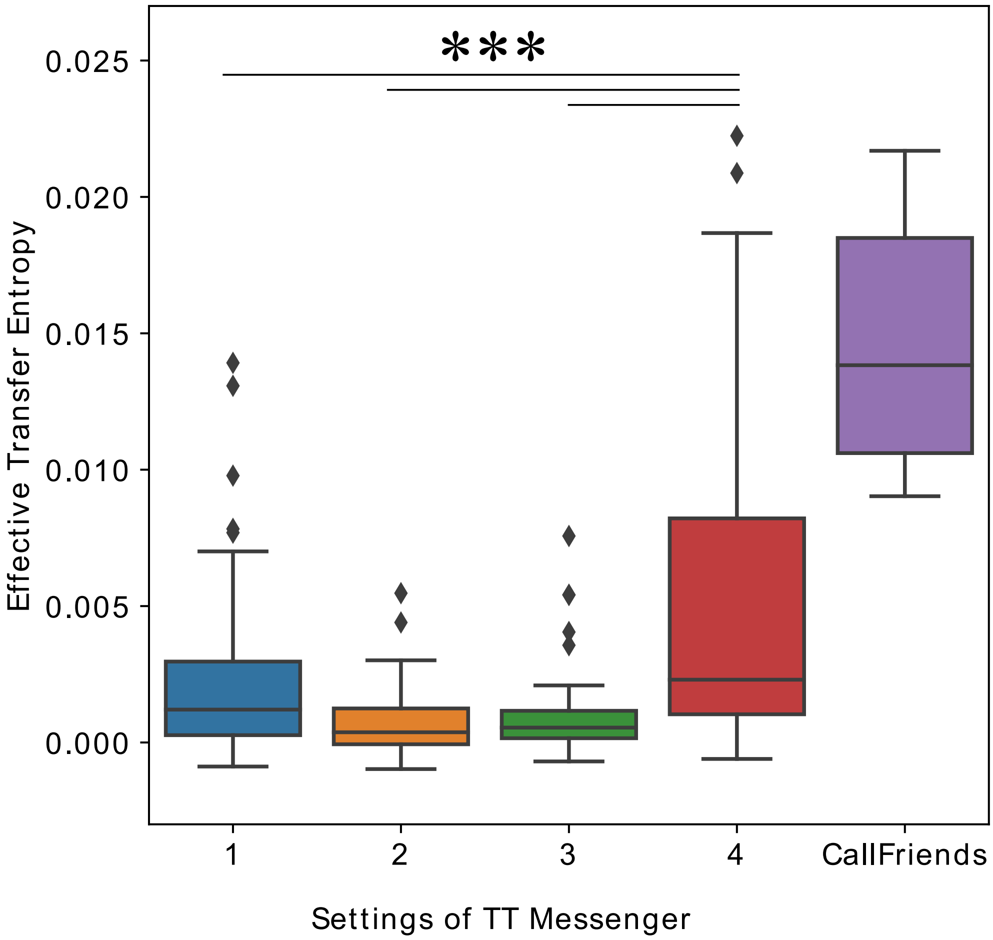

Figure 8.5: Effective transfer entropy between subjects' keystroke events timeseries data or phoneme events timeseries data from the telephone conversation data. Effective transfer entropy was calculated between two timeseries downsampled to 100ms windows, and $k = l = 2$. (*** : $p < 0.001$)

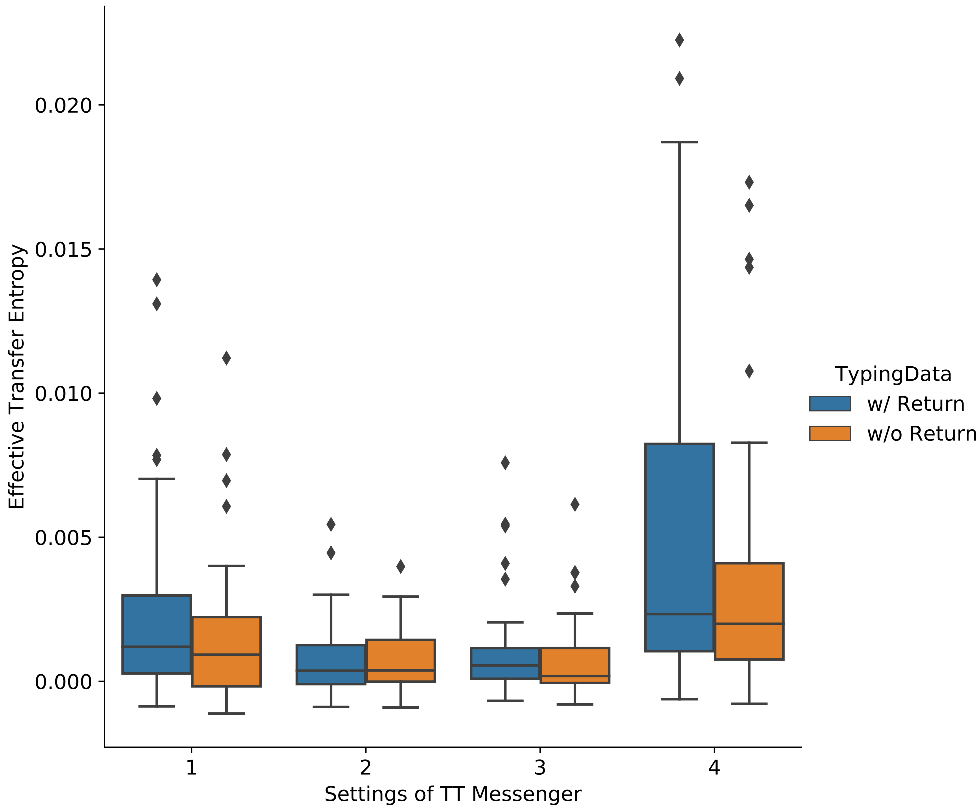

Figure 8.6: Comparison of Effective transfer entropy calculated from subjects' keystroke events including return keys and without return keys. Effective transfer entropy was calculated between two timeseries downsampled to 100ms windows, and $k = l = 2$.

Chapter 9

Discussion

9.1 Summary of our findings

With the aim to increase the social presence in our text chat system, we escalated the measures of richness and concurrency by introducing 4 different steps.

Richness of conversation designates the excess amount of information conveyed with communication. For example, in the case of a dyadic conversation, the richness increases by introducing environmental sounds, bodily gestures, facial expressions, eye directions, and so forth. We formally introduced the richness of communication in our experiment in a systematic way. Concurrency signifies multiple events happening simultaneously. For example, while in a dyadic conversation, people often look away, unconsciously touch things, and some unexpected disturbances (e.g., coffee is served by a waiter or suddenly a dog barks) await. In this paper, TypeTrace emphasizes this concurrency effect.

First, we can increase information richness by presenting the playbacks of the typing process in TypeTrace chatting (in the case of conditions 2 and 3). Transfer entropy between the interactants becomes lower, and the cognition of the presence of others does not increase. We discuss the interpretation of this result below. Secondly, we can increase the concurrency of interaction, namely the concurrency of information flow, by adding "redundant" elements to the main body messages (which is exemplified in condition 4). Emotional arousal and intimacy increase as the result of condition 4 and transfer entropy between the interactants becomes higher. We interpret the increase of transfer entropy from the other to self as the sign of increasing the sense of presence (e.g., Kojima et al. (2017)). Together with the subjective reports, we affirm that the concurrency of information is an important factor for fostering vivid conversation in CMC.

As for the comparison to the phone call data set, the transfer entropy of the phone conversation revealed to be close to that of condition 4. Additionally, the number of turns and the total time of turns are significantly greater in condition 4 when compared to conditions 2 and 3. And although we have only found a tendential increase of the number of turns in condition 4 than in condition 1, we argue that the increase of concurrency of condition 4 makes its dynamics closer to a phone call conversation.

Sherman et al. (2013) investigated the influence of the communication medium on the interaction between friends and found that the subjective report of the bonding to the other was lower during text chats than during FtF, video chats, and audio chats. This result was in line with our result in the sense that the media with high concurrency induced the higher subjective feeling of the presence of the other. As a different experimental setting, Sprecher (2014) investigated the influence of the communication medium on relation development by investigating how well pairs of unacquainted subjects can get acquainted with each other through different communication mediums. They found that the same tendency that the dyads who got acquainted with each other through text chats reported lower scores on affiliative outcomes compared to FtF, video chats, and audio chats. This result might suggest that the concurrency might also contribute to the development of relationships when getting acquainted with each other.

9.2 Characteristics of keystroke patterns

In this study, we characterized the dyadic interaction during our text chat system using several measures. First, we investigated whether the keystroke dynamics of two subjects were synchronized by comparing the distribution of the key stroke intervals, and the median keystroke interval in 1 min segments. By both measures, we did not find significant synchronization in the key stroke patterns. The movement synchronization was widely observed during the dyadic interaction especially during conversations (Shockley et al., 2009), but in the case of keystrokes, the pattern was quite unique to each individual and even used to biomerics (Teh et al., 2013), so the typing patterns might be less likely to be affected by the other's behavior.

We also analyzed the pattern in the turns, and found that the size of the turn was around several seconds, which was comparable to the time scale found in the WCLR analysis of PCE study. The window size used in the transfer entropy analysis was 100 ms and this was also comparable to the time scale used in PCE analysis. These commonality in the time scale might suggest the shared mechanism behind these two experiments.

9.3 Social Presence

In the first part of our thesis, results from the perceptual crossing experiment (PCE) suggest that the feeling of the presence of the partner, or social presence, significantly correlates with the sense of being touched by the other (passive touch). This is supported by our analysis of the transfer entropy of the two interactants' inputs. A high transfer entropy from A to B means that the information that A possesses contributes more to determining the future states of B. Another way to put it is that B's actions are not self-determined, but are determined by A. We adapted this interpretation to the results of the calculations of this current study. Our subjective reports, physiological measurements, and informational analysis confirm that social presence correlates with intimacy (social attraction), immediacy (psychological distance), and interactivity among CMC participants. Our results also suggest that it is possible to augment the level of social presence evoked by a text-based CMC by increasing the concurrency of information flow between participants. Based on our results, we believe that transfer entropy can be a measure of the social presence in a CMC environment and could serve as an important design principle for such communication systems.

Our experiments gradually manipulated the granularity of the incoming partner's message. Our initial prediction was that the social presence could be augmented by showing the typing process of the received messages (condition 2) and the automatic changes of font sizes (condition 3). However, neither transfer entropy nor subjective reports were higher in these conditions than in the standard chat setting. We speculate the reason is that the typing playback itself causes a delay in synchronous chat communication. The receiver has to wait until the playback finishes to understand the message entirely. This time delay causes a non-negligible effect on the perception of social presence and transfer entropy. Indeed, scores of subjective reports and physiological data show that positive emotions in those circumstances were lower than in the standard chat.

Early researches of CMC argued that their lack of non-verbal cues lowers the social presence of their participants (Short et al., 1976; Daft and Lengel, 1984). Richness of information and media in CMC was considered the major predictor for satisfactory communication. Later, the Social Information Processing (SIP) theory (Berger et al., 2016) suggested that accustomed users find and use alternative cues specific to CMC systems in order develop interpersonal relationships, and rejected the idea that the quality of CMC is merely determined by the richness of the media involved (Walther, 1992). Since then, researchers have been pursuing the difference in the levels of social presence depending on the richness of the medium involved in CMC (Oh et al., 2018), but some researches suggest that the richness of media can sometimes have a negative impact on the communication (Dinakar et al., 2015).

We consider our current study contributes to the Social Presence literature, and more specifically, in relation to the field of Human-Computer Interaction (HCI), by introducing transfer entropy, an informationally quantitative measurement that is congruous with psychological reports and physiological markers. These results emphasize the significance of information concurrency, which could be used for analyzing social presence in addition to the richness of media. Further research is needed to evaluate the impact of concurrency and social presence in a longitudinal setup, to understand its benefits and drawbacks on the mind of CMC users. Finally, the fact that our cognition of social presence and emotions are affected by the CMC system we use suggests both social responsibility and further possibility for designing better CMC systems to improve their users' well-being (Liu et al., 2019).

Part IV

Conclusions

Chapter 10

Discussion

This thesis aimed to understand the properties of dyadic dynamics and especially to identify what constitutes the actuality of communication. In the first part, we analyzed the result from the perceptual crossing experiment, which is a minimal experiment that includes realtime dyadic interaction, and found that the movement of two interactants synchronized with each other and organized the long range temporal pattern like turn-takings, and the time of identifying the other was associated with "passive touch".

In the second part, we expanded the scope of our study to computer-mediated communication systems and introduced the new text chat system with different settings which modulate the richness of information and concurrency. The results showed that the increasing the richness of information did not significantly increase the presence of the others, but rather the concurrency enhanced the presence and the dynamics were characterized by the increase in transfer entropy between two interactants' keystroke patterns.

In this section, based on these findings, we outlined a possible theory to characterize the dyadic interaction and the dynamics which yield the presence of the others. We start with the definition of social interaction from phenomenological studies. We compare this to our findings and aim to clarify the statement by formulating in information theoretical terms. We will explore the limitation of the present frameworks, and propose possible directions.

10.1 Characterization of Social Interaction

In the context of phenomenological studies, De Jaegher et al. (2010) defined the social interaction as a "co-regulated coupling between at least two autonomous agents, where: (i) the co-regulation and the coupling mutually affect each other, constituting an autonomous self-sustaining organization in the domain of relational dynamics and (ii) the autonomy of the agents involved is not destroyed (although its scope can be augmented or reduced)". From this definition, we especially discuss the three aspects, which are a mutually engaged and co-regulated interaction, the autonomy of the agents, and an autonomous self-sustaining organization of the interaction.

The first point is that the agents mutually affect each other, which means that neither the agent behaves independently from the other nor just follows the other's behavior. The coupling between the dyads is widely characterized as the synchronization of the movements (Schmidt et al., 1990; Shockley et al., 2003; Lakin and Chartrand, 2003; Richardson et al., 2007), and in our PCE experiments, we observed the synchrony as the cross correlation of the movements. This synchronization shows the presence of the coupling with each other, but usually does not provide the direction of the influence, so cannot estimate the mutuality of the influence. On the other hand, Transfer Entropy is a measure with direction, so it is suitable to characterize the mutuality. We observed in both experiment that the feeling of the presence of the other was associated with the high transfer entropy, and especially in the PCE experiment, we found that the transition from "passive touch" to "active touch" at the time of clicks, and this clearly showed the presence of the mutuality in the interaction.

The second point is that the autonomy of each agent. This seems incompatible with the first point because autonomy implies independence from the other's behavior, so how to reconcile these aspects is important here (Aucouturier and Ikegami, 2009). When two agents completely synchronize with each other, the autonomy of the agent is missing, but when the agents move independently with each other, then the mutual influence is missing. The turn taking structure might be one of the solutions to this in a way that the turn structure itself is co-regulated, but the content of each turn is up to the agent. The degree of autonomy of the other is observed by the other agent as the unpredictability of the input signals. The simulation study of turn-takings (Iizuka and Ikegami, 2004) reported that at the time of switching roles of two agents, the predictability of the other agent's movement dropped, so this structure enables to produce the unpredictability but the timing of the unpredicted signals to be predictable.

The third point is the autonomous self-sustaining organization of the interaction. This aspect points out that the interaction pattern itself, like turn-takings, is organized and maintained during the interaction, and this interaction pattern confines the behavior of each subject (De Jaegher and Di Paolo, 2007). Fusaroli and Tylén (2016) characterized this aspect from conversation data using "interpersonal synergy". They quantified synergy by how the pattern is more organized when combined the utterance data of two compared to each individual data using recurrence quantification analysis. We suspect that this aspect might be more clarified under the notion of top-down causation, as we discuss later.

In the following sections, we use some information measures for the attempt to integrate these aspects theoretically.

10.2 Empowerment Measure

In the previous sections, we argued three aspects of social interaction, mutual coupling, the autonomy of the agents, and autonomy of the interaction. Here, we focus on the first two aspects and sketch the possible theory based on the information theoretical measure, Empowerment (Klyubin et al., 2005; Salge et al., 2014).

In the following, we denote the sensor input of the self and the other as S^1 and S^2, and the movement of the self and the other as M^1 and M^2, respectively. We assume that the autonomy of the other, which is conceptualized as the unpredictability of the other, can be described as the entropy of the other's movement $H(M^2)$ or the resulting sensory input $H(S^1)$, and the mutual coupling can be described as the transfer entropy between M^1 and M^2. From the argument above, these values should be high during social interactions.

The information measure called Empowerment (Klyubin et al., 2005; Salge et al., 2014) has been proposed as the values to be maximized by biological agents, and defined as,

$$\mathfrak{E} = \max_{p(a_t)} I(S^1_{t+1}; M^1_t).$$

The relation of this to the values introduced for social interactions is as follows. First, the value related to the autonomy $H(S^1)$ is related to the mutual information as $I(S^1_{t+1}; M^1_t) = H(S^1_{t+1}) - H(S^1_{t+1}|M^1_t)$. Also, when the S^1 is mostly produced by the other's movement, like in the turn-takings and $S^1 \simeq M^2$, then the mutual information can be approximated as $I(M^2_{t+1}; M^1_t)$, whereas the transfer entropy can be written as $I(M^2_{t+1}; M^1_t|M^2_t)$. In this formulation, the timestep is important, because when it is small, $I(S^1_{t+1}; M^1_t)$ just corresponds to active touch, so the time step for these calculations should be around sec as we discuss in the following section.

At this moment we have not succeeded to directly model the social interaction using this framework, mainly because the timescale of interaction can be varied and co-regulated by the partners whereas the calculation of Empowerment requires a fixed timestep value. We believe that our approach aiming to formulate the dyadic interaction in terms of information theoretical measures should contribute to the understanding of social interactions by making the hypothesis testable and clarifying the underlying principles.

10.3 Top-down causation and Autonomy of interaction

The third aspect, the autonomy of the interaction might be grounded by the research in the context of top-down causation. For example, Rosas et al. (2020) quantified the top-down causation from two aspects, causal decoupling, and top-down causation. They used partial information decomposition (Williams and Beer, 2010) and Integrated Information decomposition (Mediano et al., 2019) to quantify the information gained by combining the several information sources, not by evaluated separately, and they call this quantity informational synergy. Based on these measures, they claim that the causal decoupling is quantified as the mutual information between the mutual information of the collective properties of different time steps, and the degree of top-down causation is quantified as the mutual information between the collective properties and individual information source. The degree of causal decoupling of the dyadic interaction might well quantify the degree of autonomy of social interaction, and the degree of top-down causation might be used to identify how much each agent is affected by the dyadic dynamics.

10.4 Timescale of interaction

We have developed possible theories on dyadic interaction based on information theoretical measures, but these theories should also be constrained by some physiological aspects such as time scales.

In both experiments of the present thesis, we found there were two characteristic time scales, which are ∼1 s and ∼100 ms. The first time scale ∼s roughly corresponds to the size of turns. In PCE, this was captured in the time lag which maximizes the WCLR, and about 2 s. On the other hand, in TypeTrace experiments, this was captured in the median size of turns, and about 3s for TypeTrace Messengers, and 1 s for the phone call data. A similar time scale was reported in the literature, and for example, Abney et al. (2014) analyzed conversation audio data and found around 4s as a characteristic time in Allan Factor analysis. Despite that the medium of the dyadic interaction was different from each other among these experiments, they all share a similar time scale implies that we all share similar order of time scales in dyadic interaction, and we might use this for turn-takings. However, the actual size of the turns can be varied within a similar order of time scales depending on the interaction medium, as we found that the median size of turns of phone call data was smaller than the size of TypeTrace Messengers.

The other time scale is ∼100 ms. This time scale is found in our experiments as window size for the calculation of Transfer Entropy, which corresponds to the temporal width treated as simultaneous. This time scale has also been reported in the studies of sense of agency (Blakemore et al., 1999, 2000; Farrer et al., 2013) as the size of the time delay between the action and the following sensory consequences which the subject fully felt the sense of agency.

10.5 Multi-level Alignment and Predictive Codings

So far, we only investigated superficial dynamics of interactions, but when we talk with each other, the content of each utterance also matters. In the analysis of TypeTrace Messenger, we only used the keystroke pattern and did not use the actual contents of the text. Some studies showed that the alignment also present in the texts (Garrod and Pickering, 2004), and propose that the multiple level alignment exists and influences with each other (Pickering and Garrod, 2004). To study the existence of multiple levels and the influence among different levels, the modeling of dyadic interaction between the agent with internal models seems to be required, and predictive coding might be a good candidate for the theoretical frameworks.

From the perspective of active inference (Friston et al., 2016), agents have their own generative models to predict the incoming sensory inputs and try to minimize these prediction errors by their actions or by updating their models. Their models predict the future sensory inputs using hierarchical architecture, so the information in a low level will be propagated to the higher level layers, resulting in changing the model parameters in the top layers. So far, the complete model which includes the bodily movement to the internal thought has not been constructed, but if we construct these in single hierarchical generative models, the mutual influence between the coordination in the behavior and the alignment in the representation during the conversation might be clarified.

Another important issue is that the prediction of the other agent usually cannot be converged because the agent always behaves unpredictably beyond our imaginations. Ikegami and Morimoto (2003); Taiji and Ikegami (1999) used the term "hot prediction" to indicate the prediction based on an unstable internal model, concerning the undeterministic nature of living systems. Hot prediction is opposite to cold prediction, which refers to the prediction of a physical cause-effect relationship, and the more we experience the event the prediction will get more stable. On the other hand, the hot prediction is a prediction of the behavior of agents, and this is always unpredictable because it depends on the hidden context such as internal dynamics, so the prediction will never become stable. They studied the dyadic interaction based on hot prediction. In those studies, two agents with recurrent neural networks are trained to get a high score at a coalition game. Also, each agent generates an internal model of the other agent that best imitates the other's past behavior pattern, this corresponds to hot prediction. When they simulate a three-person coalition game using this, they found itinerant phenomena that the agents exchange coalition pairs from time to time.

One of the characteristics of dyadic human interaction is that it usually does not converge, but rather possesses nonstationarity and produces novelties. This novelty might be originated from the internal complexity in the generative models of each agent. In this case, the ability to create novelties seems to be attributed to each agent and irrelevant to the dyadic interactions. However, we argue that the internal complexity just implies that each agent can potentially generate various outputs, but not ensure that all potential outputs can be internally generated. We suspect that the presence of the other can enhance the generation from the unexplored region in the internal model. Empowerment might be related to this aspect because it corresponds to how many various responses can be generated in response to the other agent's action. The other route to the novelty might be in the dynamical instabilities in mutual predictions, the hot prediction. In this setting, two agents are assumed to be coupled strongly and the interaction dynamics are central to the novelty production. In this case, it is hard to predict the behavior of each agent separately, so we expect that the causal decoupling and the top-down causation should be observed. In both cases, we argue that the dyadic interaction and underlying information theoretical structures are the basis for the novelty production in social encounters and these might provide the explanation why we are always longing for interactions with others.

Acknowledgements

I would like to thank everyone who supported me.

Especially, I am deeply grateful to Prof. Takashi Ikegami for supporting my research in many aspects, and co-authors of our paper, Dr. Tom Froese, Dr. Dominique Chen, Dr. Mizuki Oka, and Dr. Hiroyuki Iizuka. I would like to thank Prof. Akira Suyama and Prof. Hideo Higuchi for many valuable advice to my research, and Prof. Munehito Arai for the support to this thesis.

Also I would like to thank Ikegami lab's members for many fruitful discussions.

References

Abney, D. H., Paxton, A., Dale, R., and Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, 143(6):2304–2315.

Allan, D. W. (1966). Statistics of atomic frequency standards. *Proceedings of the IEEE*, 54(2):221–230.

Altmann, U. (2011). Investigation of movement synchrony using windowed cross-lagged regression. *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues*, pages 335–345.

Andersen, L. M., Pedersen, M. N., Sandberg, K., and Overgaard, M. (2015). Occipital meg activity in the early time range ($<$ 300 ms) predicts graded changes in perceptual consciousness. *Cerebral Cortex*, 26(6):2677–2688.

Antheunis, M. L., Valkenburg, P. M., and Peter, J. (2010). Getting acquainted through social network sites: Testing a model of online uncertainty reduction and social attraction. *Computers in Human Behavior*, 26(1):100–109.

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology*, 63(4):596–612.

Aucouturier, J.-J. and Ikegami, T. (2009). The illusion of agency: Two engineering approaches to compromise autonomy and reactivity in an artificial system. *Adaptive Behavior*, 17(5):402–420.

Auvray, M., Lenay, C., and Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New ideas in psychology*, 27(1):32–47.

Auvray, M. and Rohde, M. (2012). Perceptual crossing: the simplest online paradigm. *Frontiers in human neuroscience*, 6.

Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2016). Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE.

Bassili, J. N. (1976). Temporal and spatial contingencies in the perception of social events. *Journal of Personality and Social Psychology*, 33(6):680.

Bayliss, A. P., Murphy, E., Naughtin, C. K., Kritikos, A., Schilbach, L., and Becker, S. I. (2013). "gaze leading" : Initiating simulated joint attention influences eye movements and choice behavior. *Journal of Experimental Psychology: General*, 142(1):76.

Berger, C. R., Roloff, M. E., and Walther, J. B. (2016). Social Information Processing Theory (CMC). (Cmc):1–13.

Biocca, F., Harms, C., and Burgoon, J. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators and Virtual Environments*, 12(5):456–480.

Blakemore, S.-J., Frith, C. D., and Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of cognitive neuroscience*, 11(5):551–559.

Blakemore, S.-J., Wolpert, D., and Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, 11(11):R11–R16.

Clark, A. (1998). *Being there: Putting brain, body, and world together again.* MIT press.

Clement, J. (2019). Whatsapp - statistics & facts. [Online; accessed 23-March-2020].

Culnan, M. and Markus, M. (1987). Information technologies. jablin, fm; putnam, ll; roberts, kh & porter, lw (eds.): Handbook of organizational communication: An interdisciplinary perspective.

Daft, R. L. and Lengel, R. H. (1984). Information richness: a new approach to managerial behavior and organization design. *Research in Organizational Behavior*, 6:191–233.

Daft, R. L. and Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management science*, 32(5):554–571.

Dawson, M. E., Schell, A. M., and Filion, D. L. (2017). The electrodermal system.

De Jaegher, H. and Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the cognitive sciences*, 6(4):485–507.

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in cognitive sciences*, 14(10):441–447.

Di Paolo, E. A., Rohde, M., and Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? modelling the dynamics of perceptual crossing. *New ideas in psychology*, 26(2):278–294.

Dinakar, K., Chen, J., Lieberman, H., Picard, R., and Filbin, R. (2015). Mixed-initiative real-time topic modeling & visualization for crisis counseling. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, IUI ’ 15, page 417–426, New York, NY, USA. Association for Computing Machinery.

Dittrich, W. H. and Lea, S. E. (1994). Visual perception of intentional motion. *Perception*, 23(3):253–268.

Dittrich, W. H., Troscianko, T., Lea, S. E., and Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, 25(6):727–738.

Ekman, R. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews*, 24(6):581–604.

Farrer, C., Valentin, G., and Hupé, J. (2013). The time windows of the sense of agency. *Consciousness and cognition*, 22(4):1431–1441.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Pezzulo, G., et al. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68:862–879.

Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1499):2033–2039.

Frith, C. D. and Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4):531–534.

Frith, C. D. and Frith, U. (2012). Mechanisms of social cognition. *Annual review of psychology*, 63:287–313.

Froese, T. and Gallagher, S. (2012). Getting interaction theory (it) together: integrating developmental, phenomenological, enactive, and dynamical approaches to social interaction. *Interaction Studies*, 13(3):436–468.

Froese, T., Iizuka, H., and Ikegami, T. (2014a). Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. *Scientific reports*, 4.

Froese, T., Iizuka, H., and Ikegami, T. (2014b). Using minimal human-computer interfaces for studying the interactive development of social awareness. *Frontiers in psychology*, 5.

Fusaroli, R. and Tylén, K. (2016). Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science*, 40(1):145–171.

Gallese, V. and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, 2(12):493–501.

Gallotti, M., Fairhurst, M. T., and Frith, C. D. (2017). Alignment in social interactions. *Consciousness and Cognition*, 48:253–261.

Gallotti, M. and Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, 17(4):160–165.

Garrod, S. and Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, 8(1):8–11.

Gunawardena, C. N. and Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer - mediated conferencing environment. *American Journal of Distance Education*, 11(3):8–26.

Haken, H., Kelso, J. S., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological cybernetics*, 51(5):347–356.

Hasson, U. and Frith, C. D. (2016). Mirroring and beyond: Coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693).

Heider, F. and Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology*, 57(2):243–259.

Hirsch, J., Zhang, X., Noah, J. A., and Ono, Y. (2017). Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *Neuroimage*, 157:314–330.

Hove, M. J. and Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social cognition*, 27(6):949–960.

Iizuka, H. and Di Paolo, E. (2007). Minimal agency detection of embodied agents. *Advances in Artificial Life*, pages 485–494.

Iizuka, H. and Ikegami, T. (2004). Adaptability and diversity in simulated turn-taking behavior. *Artificial Life*, 10(4):361–378.

Ikegami, T. and Iizuka, H. (2007). Turn-taking interaction as a cooperative and co-creative process. *Infant Behavior and Development*, 30(2):278–288.

Ikegami, T. and Morimoto, G. (2003). Chaotic itinerancy in coupled dynamical recognizers. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 13(3):1133–1147.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, 14(2):201–211.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431):549–559.

Kassin, S. (1981). Heider and simmel revisited: Causal attribution and the animated film technique. *Review of Personality and Social Psychology*, 3:145–169.

Kello, C. T., Bella, S. D., Médé, B., and Balasubramaniam, R. (2017). Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes. *Journal of the Royal Society Interface*, 14(135).

Klyubin, A. S., Polani, D., and Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In *2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 128–135. IEEE.

Kojima, H., Froese, T., Oka, M., Iizuka, H., and Ikegami, T. (2017). A sensorimotor signature of the transition to conscious social perception: co-regulation of active and passive touch. *Frontiers in psychology*, 8:1778.

Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Quarterly journal of experimental psychology*, 63(11):2220–2230.

Kozlowski, L. T. and Cutting, J. E. (1977). Recognizing the sex of a walker from a dynamic point-light display. *Perception & psychophysics*, 21(6):575–580.

Kudo, A., Okada, T., and Chen, D. (2015). Writing style and revision process of a novelist examined through real-time creative process data. *Cognitive Studies*, 22(4):573–590.

Lakin, J. L. and Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological science*, 14(4):334–339.

Lenay, C. and Stewart, J. (2012). Minimalist approach to perceptual interactions. *Frontiers in human neuroscience*, 6.

Liu, D., Baumeister, R. F., Yang, C.-c., and Hu, B. (2019). Digital Communication Media Use and Psychological Well-Being: A Meta-Analysis. *Journal of Computer-Mediated Communication*, 24(5):259–273.

Lizier, J. T. (2014). Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Frontiers in Robotics and AI*, 1:11.

Lizier, J. T., Prokopenko, M., and Zomaya, A. Y. (2008). Local information transfer as a spatiotemporal filter for complex systems. *Physical Review E*, 77(2):026110.

Luo, Y. and Baillargeon, R. (2005). Can a self-propelled box have a goal? psychological reasoning in 5-month-old infants. *Psychological Science*, 16(8):601–608.

Marschinski, R. and Kantz, H. (2002). Analysing the information flow between financial time series. *The European Physical Journal B-Condensed Matter and Complex Systems*, 30(2):275–281.

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5-6):237–329.

Mediano, P. A., Rosas, F., Carhart-Harris, R. L., Seth, A. K., and Barrett, A. B. (2019). Beyond integrated information: A taxonomy of information dynamics phenomena. *arXiv preprint arXiv:1909.02297*.

Miyahara, K. (2015). Perceiving other agents: Passive experience for seeing the other body as the other's body. *Annual review of the Phenomenological Association of Japan*, (31):198–189.

Mossio, M. and Taraborelli, D. (2008). Action-dependent perceptual invariants: From ecological to sensorimotor approaches. *Consciousness and cognition*, 17(4):1324–1340.

Muir, D. and Hains, S. (1999). Young infants' perception of adult intentionality: Adult contingency and eye direction. *Early social cognition: Understanding others in the first months of life*, pages 155–187.

Murray, L. and Trevarthen, C. (1985). Emotional regulation of interactions between two-month-olds and their mothers: Social perception in infancy. *Social Perception in Infants*, pages 177–197.

Nadel, J., Carchon, I., Kervella, C., Marcelli, D., and Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Developmental science*, 2(2):164–173.

Nadel, J. and Tremblay-Leveau, H. (1999). Early perception of social contingencies and interpersonal intentionality: Dyadic and triadic paradigms. *Early social cognition: Understanding others in the first months of life*, pages 189–212.

Noë, A. (2004). *Action in perception*. MIT press.

Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5:114.

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., and Schilbach, L. (2011). A non-verbal turing test: differentiating mind from machine in gaze-based social interaction. *PloS one*, 6(11):e27591.

Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02):169–225.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(4):515–526.

Ramsøy, T. Z. and Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, 3(1):1–23.

Redcay, E. and Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, page 1.

Reddy, V. (2000). Coyness in early infancy. *Developmental Science*, 3(2):186–192.

Reddy, V. (2003). On being the object of attention: implications for self–other consciousness. *Trends in cognitive sciences*, 7(9):397–402.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., and Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human movement science*, 26(6):867–891.

Richardson, M. J., Marsh, K. L., and Schmidt, R. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1):62.

Rosas, F. E., Mediano, P. A., Jensen, H. J., Seth, A. K., Barrett, A. B., Carhart-Harris, R. L., and Bor, D. (2020). Reconciling emergences: An information-theoretic approach to identify causal emergence in multivariate data. *PLOS Computational Biology*, 16(12):e1008289.

Salge, C., Glackin, C., and Polani, D. (2014). Empowerment–an introduction. In *Guided Self-Organization: Inception*, pages 67–114. Springer.

Sato, W., Hyniewska, S., Minemoto, K., and Yoshikawa, S. (2019). Facial expressions of basic emotions in japanese laypeople. *Frontiers in psychology*, 10:259.

Schilbach, L. (2015). Eye to eye, face to face and brain to brain: Novel approaches to study the behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences*, 3:130–135.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., and Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36(4):393–414.

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., and Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences*, 107(20):9388–9393.

Schmidt, R. C., Carello, C., and Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of experimental psychology: human perception and performance*, 16(2):227.

Schmidt, R. C. and O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, 9(3):189–206.

Scholl, B. J. and Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, 4(8):299–309.

Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, 85(2):461.

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, 10(2):70–76.

Sherman, L. E., Michikyan, M., and Greenfield, P. M. (2013). The effects of text, audio, video, and in-person communication on bonding between friends. *Cyberpsychology: Journal of psychosocial research on cyberspace*, 7(2).

Shockley, K., Richardson, D. C., and Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1(2):305–319.

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

Short, J., Williams, E., and Christie, B. (1976). *The social psychology of telecommunications*. John Wiley & Sons.

Skewes, J. C., Skewes, L., Michael, J., and Konvalinka, I. (2015). Synchronised and complementary coordination mechanisms in an asymmetric joint aiming task. *Experimental brain research*, 233(2):551–565.

Sprecher, S. (2014). Initial interactions online-text, online-audio, online-video, or face-to-face: Effects of modality on liking, closeness, and other interpersonal outcomes. *Computers in Human Behavior*, 31(1):190–197.

Statista (2018). Unit shipments of virtual reality (vr) devices worldwide from 2017 to 2019 (in millions), by vendor. [Online; accessed 23-March-2020].

Statista (2020). Number of smartphone users worldwide from 2016 to 2021. [Online; accessed 23-March-2020].

Suda, M., Takei, Y., Aoyama, Y., Narita, K., Sato, T., Fukuda, M., and Mikuni, M. (2010). Frontopolar activation during face-to-face conversation: an in situ study using near-infrared spectroscopy. *Neuropsychologia*, 48(2):441–447.

Taiji, M. and Ikegami, T. (1999). Dynamics of internal models in game players. *Physica D: Nonlinear Phenomena*, 134(2):253–266.

Teh, P. S., Teoh, A. B. J., and Yue, S. (2013). A survey of keystroke dynamics biometrics. *The Scientific World Journal*, 2013.

Tognoli, E., Lagarde, J., DeGuzman, G. C., and Kelso, J. S. (2007). The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences*, 104(19):8190–8195.

Tremoulet, P. D. and Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, 29(8):943–951.

Trendafilov, D., Schmitz, G., Hwang, T.-H., Effenberg, A. O., and Polani, D. (2020). Tilting together: An information-theoretic characterization of behavioral roles in rhythmic dyadic interaction. *Frontiers in Human Neuroscience*, 14:185.

Tu, C.-H. (2002). The impacts of text-based cmc on online social presence. *Journal of Interactive Online Learning*, 1(2).

Van Overwalle, F. and Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, 48(3):564–584.

Varela, F. J. (1999). Present-time consciousness. *Journal of Consciousness Studies*, 6(2-3):111–140.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT press.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication research*, 19(1):52–90.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, 23(1):3–43.

Walther, J. B. (2015). Social information processing theory (cmc). pages 1–13.

Williams, P. L. and Beer, R. D. (2010). Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*.

Wiltshire, T. J., Lobato, E. J., McConnell, D. S., and Fiore, S. M. (2015). Prospects for direct social perception: a multi-theoretical integration to further the science of social cognition. *Frontiers in human neuroscience*, 8:1007.

Yaeger-Dror, M. (2004). Callfriend - japanese corpus. https://ca.talkbank.org/access/CallFriend/jpn.html.

Ziółko, B., Manandhar, S., Wilson, R. C., and Ziółko, M. (2006). Wavelet method of speech segmentation. In *2006 14th European Signal Processing Conference*, pages 1–5. IEEE.