Reciprocal Adaptation Measures for Human-Agent Interaction Evaluation

Jieyeon Woo, Catherine Pelachaud and Catherine Achard
ISIR-CNRS, Sorbonne University, 4 Place Jussieu, Paris, France

Keywords: Socially Interactive Agents, Human-Agent Interaction, Reciprocal Adaptation, Nonverbal Behaviors, Smile.

Abstract: Recent works focus on creating socially interactive agents (SIAs) that are social, engaging, and human-like. SIA development is mainly on endowing the agent with human capacities such as communication and behavior adaptation skills. Nevertheless, the task of evaluating the agent's quality remains as a challenge. Especially, the way of objectively evaluating human-agent interactions is not evident. To address this problem, we propose new measures to evaluate the agent's interaction quality. This paper focuses on interlocutors' continuous, dynamic, and reciprocal behavior adaptation during an interaction, which we refer to as reciprocal adaptation. Our reciprocal adaptation measures capture this adaptation by measuring the synchrony of behaviors including their absence of response and by assessing the behavior entrainment loop. We investigate the nonverbal adaptation, notably for smile, in dyads. Statistical analyses are conducted to improve the understanding of the adaptation phenomenon. We also studied how the presence of reciprocal adaptation may be related to different aspects of the interaction dynamics and conversational engagement. We investigate how the influence of the social dimensions of warmth and competence along with the engagement is related to reciprocal adaptation.

1 INTRODUCTION

Socially Interactive Agents (SIAs; or embodied conversational agents (ECAs)) have the goal of conducting human-like conversations while being social and engaging. Various works focus on the development of SIAs by improving the modeling of their behaviors. Nevertheless, the task of evaluating them objectively remains as a challenging problem. As SIAs are interacting with the human users, the assessment must not only be done at the agent's side but also at the interaction level considering the human interlocutor. For this, we propose new measures of reciprocal adaptation that can be used to evaluate human-agent interactions.

During an interaction, behavior adaptation between interlocutors takes place. The adaptation is done by coordinating (or synchronizing) one's behavior to that of the other and by constantly entraining and being entrained by the interacting partner.

The behavior coordination involves complex phenomena such as perceiving social signals and responding to these social signals within a given time window (Chartrand and Lakin, 2013; Burgoon et al., 1995). Conversation participants exchange by reacting to each other's social signals. The exchange is not simply alternated by taking turns between the participants (having a single reactor at the time), but the coordination involves different processes such as anticipating and producing behaviors. These behaviors are coordinated intrapersonally (between the behaviors of the same person) and interpersonally (between interlocutors). Condon and Ogston (Condon and Ogston, 1966) point out that there are intrapersonal synergies that are formed between one's behaviors and these synergies are coordinated across the interlocutors. They split the coordination into two types: intrapersonal coordination for the behavior coordination within oneself and interpersonal coordination for behavior coordination between multiple people in an interaction. To be coordinated these behaviors should match each other in action and time (Hove and Risen, 2009; Burgoon et al., 1995). We can note that Chartrand and Lakin (Chartrand and Lakin, 2013) used the term behavioral mimicry when referring to the display of a same behavior at the same time by 2 or more participants. For interpersonal coordination, an essential aspect is that the behaviors are timely aligned (Delaherche et al., 2012). This coordination of social signals may also be referred to as interpersonal synchrony. Pickering and Garrod (Pickering and Garrod, 2004) talk about alignment defined as the

Woo, J., Pelachaud, C. and Achard, C.
Reciprocal Adaptation Measures for Human-Agent Interaction Evaluation.
DOI: 10.5220/0011779300003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 1*, pages 114-125
ISBN: 978-989-758-623-1; ISSN: 2184-433X

adaptation of interlocutors' verbal behaviors. The interpersonal coordination of behaviors is an ongoing operation that turns automatically in time during a natural interaction (Schmidt and Richardson, 2008). Thus, synchrony is dynamic and is a part of reciprocal adaptation.

It is also important to note that the interpersonal coordination, that is done passively and unintentionally to match the interacting partner's behavior, has a certain delay of perception and adaptation. Chartrand and Bargh (Chartrand and Bargh, 1999), who state that the interpersonal coordination is caused by mimicry behavior, call this unconscious adaptation (or mimicry) effect the chameleon effect. This perception of interlocutors' signal is sensible to temporal alignment. For nonverbal signals the temporal alignment (or the mimicry time delay) is along a time window of 2 to 4 seconds (Leander et al., 2012).

Continuous entrainment occurs between the interlocutors (Prepin and Pelachaud, 2011). When a person shows a behavior, it entrains the mimicry behavior of their interactant. The entrainment doesn't end with a simple mimicry but it also rentrains the initial signal sender to continue performing the same behavior or to resend the same signal. We refer to this process of sequential entrainment as entrainment loop.

We are interested in understanding and measuring reciprocal adaptation, looking at the temporal synchronization and entrainment loop between participants' behaviors. We propose novel measures to understand how reciprocal adaptation emerges during an interaction. By studying these measures, we have identified different levels of synchrony and entrainment loop of dyads. We also study how synchrony and entrainment loop participate in the perception of engagement between interlocutors and in the perception of interlocutors' social attitudes. We hypothesize to see a proportional relationship between reciprocal adaptation (synchrony and entrainment loop) and engagement levels. We also hypothesize that reciprocal adaptation may have an impact on the perception of the social dimensions of warmth and competence of the interlocutors, with a positive correlation with warmth and negative relation with competence.

In our study, we focus on smile, a social signal that may convey a great variety of communicative and emotional functions (Niedenthal et al., 2010; Hess et al., 2014). Smiles are frequently observed during an interaction (Knapp et al., 2013). They can signal friendliness, positive emotions; they can be used as a polite signal to greet an acquaintance; they can be indicated as agreement, liking; etc. Smile is an important socio-emotional signal that has received a lot of interest in affective computing domains. Previous studies have highlighted the power of smiling SIAs to achieve such a goal (Wang and Ruiz, 2021; Ochs and Pelachaud, 2013).

We present new reciprocal adaptation measures that can be employed to objectively evaluating the quality of the agent in human-agent interaction. Our ultimate goal is to build socially interactive SIAs that is able to maintain user's engagement during an interaction. In the scopes of this paper, we are interested in studying reciprocal adaptation of the smile behaviors in a dyadic interaction. To do so, we propose new objective measures that studies the synchrony of behaviors including their absence of response and behavior entrainment loop to better understand how nonverbal behavior adaptation emerges during an interaction. We aim to investigate how they are displayed between the participants of an interaction and how they participate to the perception of conversational engagement and to the perception of social attitudes of the participants. We look at the relation of reciprocal adaptation with the engagement level and the social dimensions of warmth and competence.

The paper is structured as the following: Section 2 introduces related measures for reciprocal adaptation evaluation; Section 3 explains our reciprocal adaptation measures; Section 4 presents the analyzed corpus; and Section 5 shares statistical analysis of the our reciprocal adaptation measures and their relationship with engagement and social attitudes of warmth and competence.

2 EXISTING MEASURES

During a conversation, interlocutors dynamically adapt by coordinating their speech and behaviors (Condon and Ogston, 1967; Burgoon et al., 1995; Bernieri and Rosenthal, 1991; Chartrand and Lakin, 2013). Among the various social signals that are produced during an interaction, the smile is one of the most important human interaction signals. The smile alone can express diverse information (e.g. affect state, level of engagement, and intrinsic nature) to the interacting partner in a variety of social context (Ekman, 1992; Hess et al., 2002). The presence of smile that incorporates such diverse implications can impact the perception by other partner (e.g trust, intelligence, warmth, and attractiveness) (Scharlemann et al., 2001; Lau, 1982; Reis et al., 1990). As such, we want to check the influence of smile between the interlocutors and thus hold interest in measuring the smile adaptation. To find out how to measure the adaptation of smiles, we investigate on related measures notably synchrony measures (e.g. measures for nonverbal signals and biomedical signals).

Early works on synchrony started off with manual assessment done by trained observers who were trained to perceive it directly in the data. Such evaluations were based on behavior coding methods that evaluate the interaction behaviors on a local scale by analyzing them in micro-units (Cappella, 1997; Condon and Sander, 1974). However, the training of observers is very labour-intensive which led them to switch to a judgment method that uses a Likert scale to rate behaviors on a longer time scale (Cappella, 1997; Bernieri et al., 1988). The problem with manual annotations, that rely on perception by a third party, is that it is very costly. Manual annotations are very time-consuming and there is a risk of being biased as the label decision depends heavily on the annotator. Thus, we want an objective evaluation technique that can automatically process and render a non-biased synchrony measure.

Automatic measures enable us to avoid tedious work of manual annotation by automatically capturing relevant social signals that detect the presence of synchrony. The most commonly used way to measure interpersonal synchrony is correlation (Campbell, 2008; Delaherche and Chetouani, 2010; Reidsma et al., 2010). As the behavior movements (e.g. body motion and vocal energy) are produced after the perception of the other interlocutors' motions there is a certain time delay to be considered. Several works address this by applying the time-lagged cross-correlation (Boker et al., 2002; Ashenfelter et al., 2009; Beňuš et al., 2011). A hindersome limitation of correlation is that a window length of interaction must be chosen to perform the correlation. However, the window sizes can vary for each produced motion and are not the same for interactors.

Another method of synchrony evaluation is the recurrence analysis (Shockley et al., 2003; Varni et al., 2010). The analysis assesses "recurrence points" which are points in time where similar states (or patterns of change) are visited by two different systems. The recurrent analysis depends on manipulatable states (e.g. posture state or affect state) and shows a graphical representation (a diagonal structure) of time periods when two systems visit the same state. For the recurrent analysis, the evaluation requires a fixed length of system periods and time shifts. However, the signals do not happen exactly after a certain time but within a time delay (e.g. 2 to 4 seconds) (Chartrand and Bargh, 1999; Leander et al., 2012).

The response of a smile is very dynamic. Each smile is not produced with the same length, and as stated above, the timing of the smile varies. For example, when we are asked to reproduce a smile that we have made, it is almost impossible to recreate the exact same smile with the same duration and timing. To address such dynamics, the measure must be invariant to dilations and shifts. A frequently used technique to do so is the Dynamic Time Warping (DTW) (Müller, 2007) which assesses the similarity between two temporal sequences of different speed and length. Nevertheless, the DTW matches every index of a sequence with one or more indexes from the other, which can be problematic for our case of nonverbal behaviors as both cases of a behavior occurring or not are correct answers (i.e. absence of response, for instance a person can reply with a smile or choose to not reply but both cases are plausible responses) but the DTW will consider it as an error.

New indicators characterizing synchrony phenomena were introduced by Rauzy *et al.* (Rauzy et al., 2022). They consider the two signal timescales as oscillating normal modes associated with the sum and the difference of the trajectories (x_{sum} for symmetric mode and x_{diff} for asymmetric mode). Based on the two, they propose new indicators (mode characteristic periods, coupling factor, coefficient of synchrony, and energy) to evaluate the synchrony.

As an alternative to temporal methods, spectral analysis was suggested. The evolution of relative phase for a stable time-lag between interlocutors is measured (Oullier et al., 2008; Richardson et al., 2007). It also renders information about the coordination stability with the flatness degree of the phase distribution and the overlapping frequency via the cross-spectral coherence. The synchrony can be also measured in the time-frequency domain via cross-wavelet coherence (Hale et al., 2020).

The field of biomedical signal processing also holds a big interest in such synchrony measures for applications such as detecting synchrony in EEG (Bakhshayesh et al., 2019). Various metrics are employed from point to point measures such as correlation and coherence (a linear correlation computed in the frequency domain via cross spectrum), correntropy coefficient (a correlation measure that is sensitive to nonlinear relationship and high order statistics), wav-entropy coefficient (a correntropy computed in the time-frequency domain with wavelet transforms), to measures that are solely focused on synchronization like phase synchrony (an amplitude-independent estimation of signal phase relationship) and event synchronization (a measure calculated from the number of occurrences of predefined signal events, counting events that are followed by another event in the other signal within a specified time, and their symmetric counterpart). Yet these measures are not suitable for our use as stated above for point to point measures and as for phase synchrony the subsequences of a signal might have different phase delays which could be troublesome. For event synchronization, it does not match exactly with our specific condition.

In our work, we are interested in measuring how people adapt their behavior, in particular their smile, during an interaction. During an interaction, participants may respond and adapt to each other's behavior. These interactive behaviors may serve to reinforce the relationship between the participants, their engagement in the interaction, but also to display different social attitudes. We are interested in measuring reciprocal adaptation as a function of synchrony patterns and entrainment loop. Our measure of synchrony patterns includes when participants respond or not to each other's behaviors. The absence of response is considered as an error by the point to point measures (e.g. correlation) and the DTW approach and is completely ignored by the recurrent analysis, spectral analysis, and cross-wavelet analysis. However, the absence of response may also convey important information about the interaction. In order to study the impact of absence of response, we need a new measure that is capable of detecting the addition (produced by oneself without the reaction of the other) and the suppression (produced by the other without the reaction of oneself) of signals while still being able to measure the synchrony between the interlocutors. In addition to the absence of response, we interest in observing the behavior entrainment loop. We propose to also capture this entrainment loop which is absent in the aforementioned synchrony measures.

3 OUR PROPOSED RECIPROCAL ADAPTATION MEASURES

To our knowledge, existing measures (see Section 2) are not suitable for our problem, notably regarding the absence of a response and capturing the entrainment loop. To overcome this limitation, we propose a new ways to measure the reciprocal adaptation for a dyadic pair that measure the synchrony of behaviors including their absence of response while tolerating time swift, dilation, deletion and insertion, and capture the behavior entrainment loop.

3.1 Measures of Synchrony Behavior Including their Absence of Response

We firstly address the problem by taking into account the absence of response when measuring the synchrony. Our method derives from the classical sequence dissimilarity quantification technique called edit distance or Levenshtein distance (Navarro, 2001). Its use can be mostly observed in fields such as natural language processing (Lhoussain et al., 2015) and bioinformatics (Chang and Lawler, 1994) as it compares the similarity between two strings (e.g. words) by counting the minimum number of transformation operations that are required to convert one string into the other. We grab the concepts of insertion and deletion of the edit distance while we don't use the concept of substitution.

We evaluate the synchrony with signal activation by converting continuous values to binary values and extract subsequences corresponding to active signal parts, with their starting (s) and ending (e) times. We choose to binarize the continuous values to better see the impact of absence of response. Let us consider an active subsequence (sequence of 1) A from person PA and B from person PB.

We consider that both subsequences are synchronized or paired if:

$$|e_A - e_B| + |s_A - s_B| \leq threshold \tag{1}$$

where the threshold is set to the mimicry time delay (i.e. 4 seconds which gave the best results among the thresholds of 2, 3, and 4 seconds). For our application of measuring the synchrony of smile, we took the threshold of 4 seconds (considering all responses that happen within a maximum of 4 seconds); actually the literature on nonverbal behavior mimicry states that the mimicry time delay can vary from 2 to 4 seconds (Chartrand and Bargh, 1999; Leander et al., 2012).

If several subsequences of a person check this condition with the same subsequence of the other person, a synced pair is formed with the one that has the minimum distance. The other subsequences are not paired.

Both paired subsequences and unpaired subsequences of persons A and B are considered to estimate the synchrony:

$$PA\&PB = \frac{nb.\ of\ synced\ pairs}{total\ nb.\ of\ events}$$

$$PA\&\neg PB = \frac{nb.\ of\ unpaired\ subseq.s(seqA|seqB)}{total\ nb.\ of\ events}$$

$$PB\&\neg PA = \frac{nb.\ of\ unpaired\ subseq.s(seqB|seqA)}{total\ nb.\ of\ events}$$

where the total number of events is the sum of the number of synced pairs and the number of unpaired subsequences of both persons A and B.

Each measure renders a probability that corresponds to:

- *PA&PB*: *PA* and *PB* responding to each other,
- *PA&¬PB*: *PA* is active but not *PB*,
- *PB&¬PA*: *PB* is active but not *PA*.

PA&PB means that both participants smile simultaneously or with a small delay corresponding to the reacting time; this measure represents the sync between *PA* and *PB*. For *PA&¬PB* and *PB&¬PA*, only one of the person is acting (*PA* smiles and *PB* does not smile, and vice versa), these measures indicate that *PA* and *PB* are not in sync.

3.2 Measure of Entrainment Loop

We are also interested in capturing the entrainment of smile. The smile of *PA* can entrain the smile of *PB* which then entrains *PA* to continue to smile or to smile again within a certain time delay and vice versa. We refer to this as the entrainment loop of smile. The entrainment loop consists of two types:

- Type 1: continuous smile, seen in Figure 1;
- Type 2: repeated smile with an overlap or within a certain time delay (i.e. mimicry delay of 4 seconds), seen in Figure 2 and Figure 3 respectively.

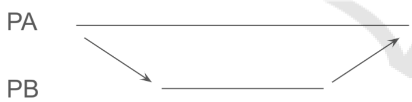

Figure 1: Entrainment loop type 1 of a continuous smile of *PA*.

Figure 2: Entrainment loop type 2 of a repeated smile of *PA* with overlap.

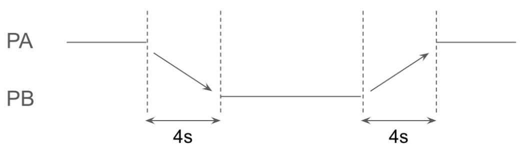

Figure 3: Entrainment loop type 2 of a repeated smile of *PA* within the mimicry delay of 4 seconds.

We capture these two types of entrainment loop and count the number of occurrence of entrainment loops for each interaction.

4 CORPUS

We chose to use the NoXi database (Cafaro et al., 2017). NoXi is a corpus of screen-mediated face-to-face interactions. It contains natural dyadic conversations talking about a common topic. Each interacting dyad consists of a pair of participants with two different roles which are called expert and novice (Cafaro et al., 2017). The expert is the one who transfers information with the goal of sharing his/her knowledge on a topic and thus who leads the conversation by talking more frequently and for longer time. The novice (the other interacting partner) receives the information and responses to the sayings of the expert on the topic.

The NoXi database consists of 3 parts depending on the recording location (France, Germany, and UK). For our work, we only use the recording from the French location which consists of 21 dyadic interactions performed by 28 participants with a total duration of 7h22. We extract the intensity of Action Unit 12 (AU12; zygomatic major) via the opensource toolkit OpenFace (Baltrusaitis et al., 2018) and preprocess it by performing median filter and linear interpolation. In the remaining of this paper, we use the term smile to refer to AU12; though we are aware that smile may be produced by different Action Units (e.g. AU11, AU13...) in combination of other Action Units (such as AU6 or AU1, AU2) (Ekman and Friesen, 1982). To get the smile activation, we binarize the continuous intensity value of smile with the threshold of 1.5/5 which is the minimal intensity (manually identified) for a smile activation.

5 STATISTICAL ANALYSIS & DISCUSSION

Our reciprocal adaptation measures are computed with activation state (binary activation values). At a first step, we transformed the continuous smile intensity (i.e. AU12) to smile activation with a threshold of 1.5/5, 5 being the maximal intensity in OpenFace. We found that the intensity 1.5 was the minimal intensity for a smile which was manually identified. So, a smile of intensity 1.5 corresponds to a small smile while a smile of intensity 5 to a large one.

5.1 Smile Distribution

To start off, we wanted to visualize the distribution of smiles in terms of its occurrence frequency and its duration in our database depending on the person's role (expert or novice). We annotate the novice as $P1$ and the expert as $P2$.

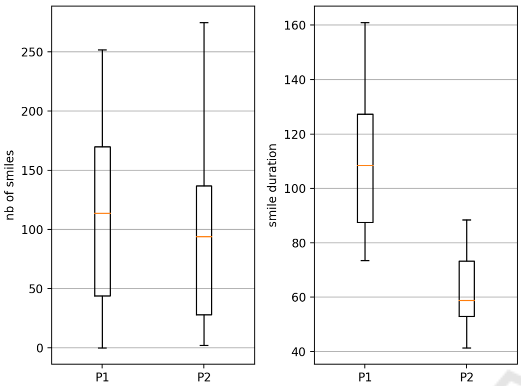

Figure 4: (left) Number of smiles produced by $P1$ and by $P2$; (right) Smile durations of $P1$ and $P2$.

With the visualization of the smile occurrence distribution in Figure 4 (left), we note that $P1$ tends to smile more often than $P2$. The context of the dyadic interaction of the NoXi corpus is mainly friendly and positive. Participants were paired between one that wanted to talk about a topic and one that wanted to learn about this topic (Cafaro et al., 2017). Within such an interaction context, having $P1$ smiling more than $P2$ can be explained by $P1$ displaying positive backchannels or showing actively his/her involvement when $P2$ is talking. Along with the number of smiles produced by the participants, we also hold interest in the smile duration distribution. Figure 4 (right) shows that $P1$ generally maintains his/her smile longer than $P2$. This can further support our analysis that $P1$'s smiles may have the purpose of showing conversational involvement.

5.2 Synchrony Behaviors Including Their Absence of Response

Going back to our initial objective of investigating the reciprocal adaptation of smile and its relation with the perception of social attitudes, we start by analyzing the smile with our measures of synchrony behaviors including their absence of response.

5.2.1 Smile Synchrony Distribution

We computed the probability densities, via our proposed measures, to visualize the distribution of 3 cases: $P1$ and $P2$ responding to each other ($P1\&P2$), $P2$ smiling to $P1$ but not reversely ($P2\&\neg P1$), and $P1$ smiling to $P2$ but not reversely ($P1\&\neg P2$).

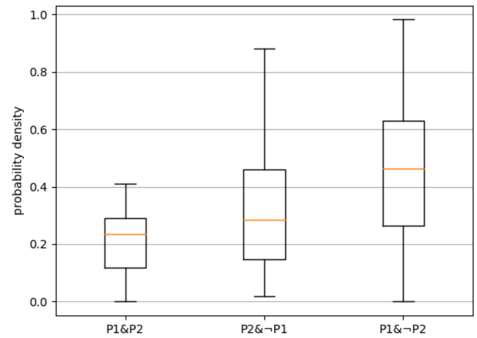

Figure 5: Probability density of smiles that are in sync ($P1\&P2$), $P2$ smiling without the response of $P1$ ($P2\&\neg P1$) and $P1$ smiling without the response of $P2$ ($P1\&\neg P2$).

We can remark, in Figure 5, that during the conversation both $P1$ and $P2$ produce smiles that are in sync responding to one another (smiling at the same time or following back within the mimicry delay of 4 seconds) and also smiles that are not responded by the other partner. As seen in Figure 4, $P1$ has a higher probability density of smiling even during the absence of the other interacting partner's response ($P1\&\neg P2$), because of his/her tendency to smile more than $P2$.

5.2.2 Synchrony Clustering

To better investigate the synchrony between the two interlocutors, we decided to first check if the smile synchrony of the 21 video dyads of the NoXi corpus can be classified into different levels. We performed a dendrogram hierarchical clustering to cluster the dyads using our obtained measures of synchrony behaviors including their absence of response ($P1\&P2$, $P2\&\neg P1$, and $P1\&\neg P2$). As seen in Figure 6, we split our data into three clusters by cutting the dendrogram with a threshold of 1.0. The cluster classes can be visualized in the 3-dimensional space of our proposed measures in Figure 7.

In Figure 8, we can note that the synchronization of level 1 ($P1\&P2 \sim 0.072$) occurs when $P2$ smiles very frequently ($P2\&\neg P1 \sim 0.924$) while $P1$ does not smile much ($P1\&\neg P2 \sim 0.004$). A level 2 synchrony ($P1\&P2 \sim 0.231$) is seen when $P1$ smiles a lot ($P1\&\neg P2 \sim 0.637$) and $P2$ smiles a bit ($P2\&\neg P1 \sim 0.146$). For level 3 synchrony ($P1\&P2 \sim 0.33$), it is observed when $P1$ and $P2$ both smile frequently ($P2\&\neg P1 \sim 0.408$ and $P1\&\neg P2 \sim 0.305$).

We can deduce from these three levels that the highest level of synchronization (level 3 where

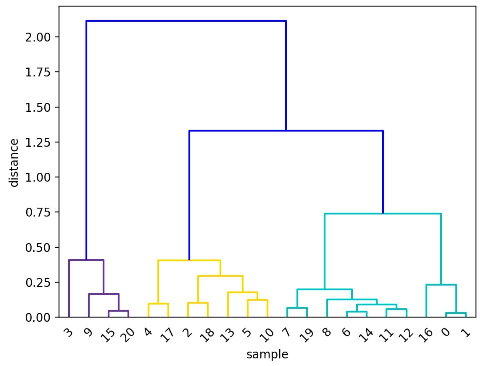

Figure 6: Dendrogram of synchrony measures where the distance is the distance between the sample points in the 3D space of our proposed measures of synchrony.

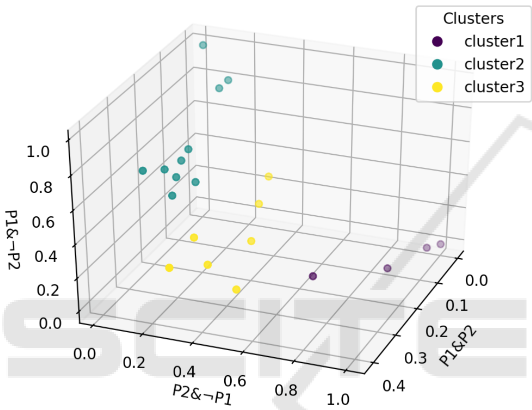

Figure 7: 3D visualization of the three synchrony classes obtained using the dendrogram.

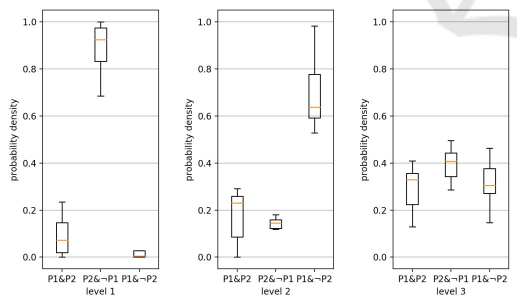

Figure 8: Probability density of smiles that are in sync ($P1\&P2$), or not ($P2\&\neg P1$ and $P1\&\neg P2$) for each class obtained with the dendrogram: (left) level 1; (middle) level 2; (right) level 3.

$P1\&P2 \sim 0.33$) is correlated with both interacting partners who tend to smile frequently, while the lower levels of synchronization, level 1 ($P1\&P2 \sim 0.072$) and level 2 ($P1\&P2 \sim 0.231$), are correlated with the situation when one of the partners, independent of his/her role, does not respond much. This shows how the presence of smile reciprocity is an important factor with respect to synchrony level; a partner that nearly does not respond to other's smile ($P1\&\neg P2 \sim 0.023$ of level 1) deteriorates the synchrony of the two even when the other interlocutor smiles a lot ($P2\&\neg P1 \sim 0.924$ of level 1). It confirms that synchronization is highly dependent on coordination between partners (Burgoon et al., 1995; Tschacher et al., 2014).

5.2.3 Relationship Between Synchrony and Engagement & Social Attitudes

We also want to see if synchrony plays a role in the perception of engagement and social attitudes of warmth and competence. As we have previously hypothesized, we expect a correlation between synchrony and engagement levels (hypothesis 1), and also for the social dimensions of warmth (hypothesis 2), and competence (hypothesis 3):

- Hypothesis 1: positive correlation between synchrony ($P1\&P2$) and engagement level,
- Hypothesis 2: positive correlation between synchrony ($P1\&P2$) and warmth level,
- Hypothesis 3: negative correlation between synchrony ($P1\&P2$) and competence level.

For the annotations, we base on previous works done on the NoXi corpus (available with the annotation tool NOVA (Heimerl et al., 2019)). For the engagement annotations, the perception change of engagement was characterized in (Dermouche and Pelachaud, 2019) with five levels (0: strongly disengaged; 1: partially disengaged; 2: neutral; 3: partially engaged; 4: strongly engaged). In (Biancardi et al., 2017), the continuous annotations of social dimensions of warmth and competence were done with scores ranging from 0 to 1 (0: very low degree of perceived warmth or competence; 1: very high degree of warmth or competence). As the work of (Biancardi et al., 2017) focuses on $P2$ (expert), we also evaluate the impact of synchrony on the three aspects of engagement, warmth, and competence of $P2$.

To test if our assumptions are correct, we will observe the engagement and the social attitudes depending on our measures of synchrony ($P1\&P2$, $P2\&\neg P1$, and $P1\&\neg P2$) and on the synchronization levels (level 1, level 2, and level 3). The analysis was done with two different methods to measure the engagement and/or social attitudes.

The first method, method 1, consists of computing the *local average value* of engagement and/or social attitudes levels only on the segments where a smile occurs, either on both participants' faces (condition $P1\&P2$) or for just on one participant's face (condition $P2\&\neg P1$ or $P1\&\neg P2$). A delay of 2 seconds is applied considering the reaction lag of the evaluator, as proposed in (Mariooryad and Busso, 2014)). We then compute the mean of all the averaged values of segments.

The second method, method 2, uses the *global average value* of the engagement level (respectively of the warmth and competence levels) over the entire video of each dyad, independent of the smile synchrony sequence. For this second method, as a single value is computed for each entire video of the corpus, we cannot use it to see the relationship that depends on our measures of synchrony ($P1\&P2$, $P2\&\neg P1$, and $P1\&\neg P2$) as they derive from a single sample (i.e. one smile occurrence).

So all in all, we evaluate the relationship between synchrony and engagement (identically for both social attitudes) using three conditions:

- Condition 1: method 1 and averaged values of segments belonging to ($P1\&P2$, $P2\&\neg P1$, and $P1\&\neg P2$),
- Condition 2: method 1 and averaged values of segments belonging to synchrony levels 1, 2, and 3,
- Condition 3: method 2 for video dyad of synchrony levels 1, 2, and 3.

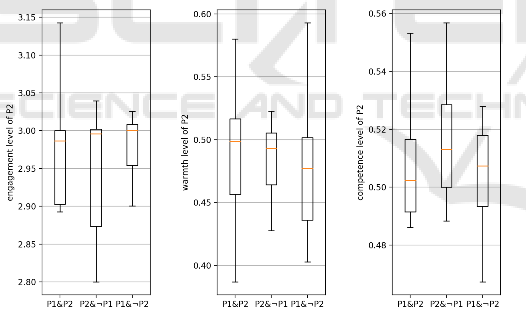

Figure 9: Distribution of engagement (left), warmth (center), and competence (right) levels measured for condition 1.

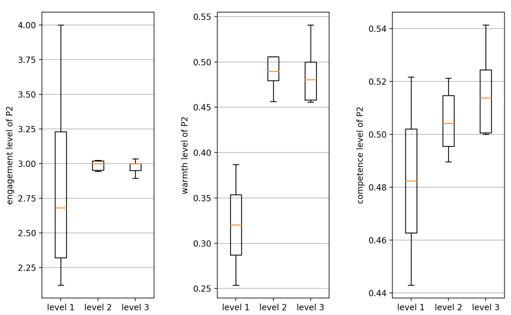

Figure 10: Distribution of engagement (left), warmth (center), and competence (right) levels measured for condition 2.

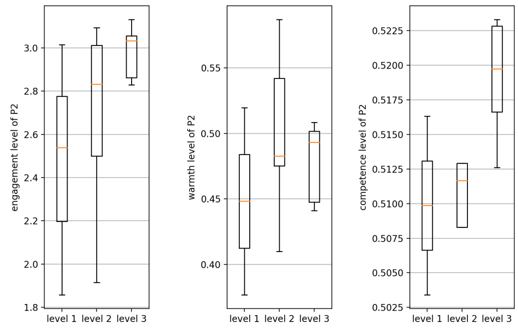

Figure 11: Distribution of engagement (left), warmth (center), and competence (right) levels measured for condition 3.

For the engagement, we can see in Figure 9 (left) that similar levels of engagement are obtained for $P2$ disregarding whether $P1$ and $P2$ are in sync ($P1\&P2 \sim 2.986$) or not ($P2\&\neg P1 \sim 2.996$ and $P1\&\neg P2 \sim 3.0$). When looking at the relationship depending on the synchrony level, in Figure 10 (left) we can observe that the level 1 (~ 2.682) indicates a lower engagement level compared to levels 2 and 3 (3.0 for both) and in Figure 11 (left) the proportional relationship between engagement and synchrony level is clearly shown (level 1 ~ 2.538, level 2 ~ 2.832, and level 3 ~ 3.033). Thus, we found a positive relationship between engagement and synchrony levels. Our analysis shows that the more engaged the participants are the more they show behavior synchronization (here smile of $P1\&P2$). It validates our first hypothesis. Condition 3 offers a clearer view. That is, providing a *global average value* for the engagement level better represents the characteristics of engagement of participants in an interaction; only looking at the short sequences of smiling moments is not sufficient to capture the whole picture of the engagement.

The warmth dimension in Figure 9 (middle) shows that when $P1$ and $P2$ are in sync (for their smile, at least) $P2$ is perceived warmer ($P1\&P2 \sim 0.499$) compared to when they are not in sync ($P2\&\neg P1 \sim 0.493$ and $P1\&\neg P2 \sim 0.477$). $P2$ is also thought to be warmer when he/she is the only one smiling ($P2\&\neg P1 \sim 0.493$) against the opposite situation ($P1\&\neg P2 \sim 0.477$; only $P1$ smiling). In Figure 10 (middle), the lower level of warmth at synchrony level 1 (~ 0.32) is distinguishable from the higher levels of warmth at synchrony levels 2 and 3 (~ 0.49 and ~ 0.481 respectively). When looking at Figure 11 (middle), we can see a rise in warmth level as the synchrony level increases (level 1 ~ 0.448, level 2 ~ 0.483, and level 3 ~ 0.493). The results for warmth tell us that being in synchrony with the other interacting participant gives a warmer impression and that the improvement of synchrony level ($P1\&P2$) conducts the growth in warmth level which validates our hypothesis 2. Moreover, the smiling tendency of the interlocutor is linked to his/her impression of warmth which is conformed with the literature that (genuine) smiles are signals of warmth (Lau, 1982; Reis et al., 1990).

In the case of the social trait of competence, we can remark in Figure 9 (right) that $P2$ is perceived as more competent when $P2$ is the only one smiling with no smiling back from $P1$ ($P2\&\neg P1 \sim 0.513$) followed up by when $P1$ is smiling alone ($P1\&\neg P2 \sim 0.507$) and then by when $P1$ and $P2$ are in sync ($P1\&P2 \sim 0.502$).

Previous researches (Bernstein et al., 2010; Biancardi et al., 2017) have highlighted that a smiling person is perceived as more affiliative and less dominant. In the context of an interaction, the interplay of participants' behaviors modulates their perception. In a study on behavior mimicry, Tiedens and Fragale (Tiedens and Fragale, 2003) have reported that when participants have different status (here in NoXi, knowledgeable on a topic vs wanted to learn on this topic), it seems to be correlated with complementarity pattern rather than mimicry. In the NoXi corpus, $P2$ acts as the "expert" that conveys information on a topic that $P1$ is interested to learn more about. Thus, $P2$ has the role of a knowledgeable person on the topic of discussion. It confers him/her a form of expertise and thus of competence. In the context of the NoXi corpus, when $P2$ displays a smile which is not responded by a smile of $P1$, $P2$ appears to be more competent than in the other smiling conditions. However, coordination of behaviors of both participants appears to modulate this inference as reported in previous studies (Tiedens and Fragale, 2003). Further studies involving other nonverbal signals (e.g. frowning, sighting) need to be conducted to see if this condition leads to complementarity.

In Figures 10 (right) and 11 (right), the increase in synchronization level leads to the rise in the perception of competence level. We could say that the higher the synchronization the more the interlocutors show involvement that gives a feeling of being more proficient around the subject of discussion and thus appearing more competent. This finding is against our hypothesis 3, of synchrony ($P1\&P2$) having an indirect relationship with competence level. Instead it follows previous literature work that saw the phenomenon of smiling people being perceived as intelligent and trustworthy (Lau, 1982; Scharlemann et al., 2001). However, it is against our hypothesis with is based on observation of Biancardi *et al.* and Cuddy *et al.* (Biancardi et al., 2017; Cuddy et al., 2011) that smiling behavior is negatively associated with competence. In our case, we remark a halo effect which occurs when the judgments of an undescribed targeted dimension (i.e. competence) goes towards the same direction as the other given dimension (i.e. warmth). Contrary to (Biancardi et al., 2017; Cuddy et al., 2011)'s study that looks only at one person, in our study we focus on the interaction and on how participants in a dyad interact with each other. This could explain the differences in our results and (Lau, 1982; Scharlemann et al., 2001) and in (Biancardi et al., 2017; Cuddy et al., 2011).

5.3 Entrainment Loop

We also want to observe the impact of entrainment loop on the aspects of engagement and social dimensions of warmth and competence.

5.3.1 Types of Entrainment Loop

We firstly check the number of occurrence of the two types of entrainment loop.

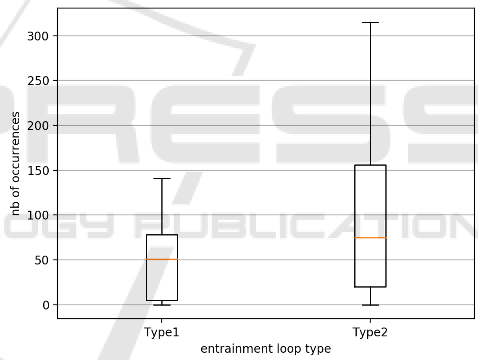

Figure 12: Number of occurrence of the two entrainment loop types.

In Figure 12, we can notice that the two entrainment loop types' occurrence frequencies are not negligible. With this, we can state that both types should be considered.

5.3.2 Relationship Between Entrainment Loop and Engagement & Social Attitudes

As above, we observe the relationship of entrainment loop with the aspects of engagement and social attitudes via the aforementioned methods (using method 1: local average value or method 2: global average value of the engagement, warmth and competence levels). Before analyzing the relationships, we cluster the interactions into two groups by splitting them with the median number of occurrence of entrainment loops.

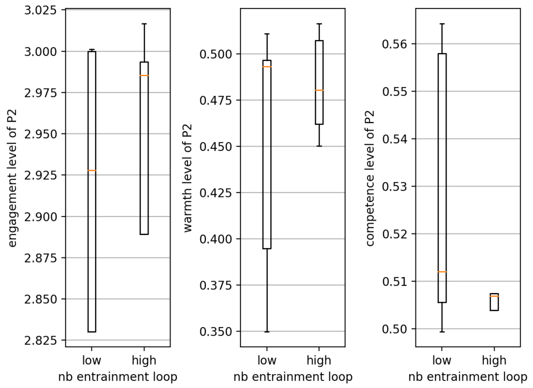

Figure 13: Distribution of engagement (left), warmth (center), and competence (right) levels measured with method 1 (local average value) for entrainment loop.

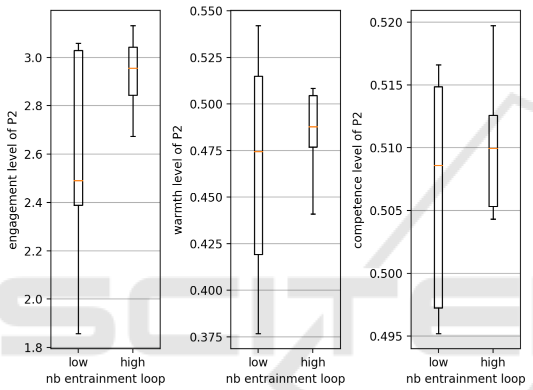

Figure 14: Distribution of engagement (left), warmth (center), and competence (right) levels measured with method 2 (global average value) for entrainment loop.

For the engagement, we can note that the engagement level increases with respect to the number of entrainment loop occurrences for both method 1 ($low \sim$ 2.490 and $high \sim$ 2.956) and method 2 ($low \sim$ 2.928 and $high \sim$ 2.985), in Figure 13 (left) and in Figure 14 (left) respectively.

For the social attitudes, when looking at them for method 1, we can remark with their median values that warmth and competence attitude levels decrease ($low \sim 0.493$ and $high \sim 0.480$, and $low \sim 0.512$ and $high \sim 0.507$ respectively) when entrainment loop occurrence transits from low to high. Nevertheless, for both cases we can see that the class for high entrainment loop occurrence is more concentrated ranging at a high warmth level ($0.394 < low < 0.497$ and $0.462 < high < 0.508$) and low competence level ($0.505 < low < 0.558$ and $0.503 < high < 0.508$). Thus, we can state that at the moment of the entrainment, the warmth level rises and the competence level decreases which is inline with the findings of (Biancardi et al., 2017; Cuddy et al., 2011).

For method 2, both warmth and competence levels increase ($low \sim 0.475$ and $high \sim 0.488$, and $low \sim 0.509$ and $high \sim 0.510$ respectively). However, no significance can be found for competence, thus validating only for warmth level to be correlated to the number of entrainment loops.

6 CONCLUSION AND DISCUSSION

As reciprocal adaptation occurs naturally as we converse, it generally passes unnoticed without giving any explicit attention towards it. Nevertheless, this aspect of reciprocal adaptation, and particularly interpersonal synchronization and entrainment loop, is an important factor for an interactive and engaging communication. With our new reciprocal adaptation evaluation measures, that assess synchrony behaviors including their response absences and measures entrainment loop, we were able to carry out several statistical analyses on smile synchrony distribution, clustering synchronization levels (level 1, level 2, and level 3) and the relationship with engagement and social dimensions (warmth and competence). Also, we observed the relation between entrainment loop occurrences and engagement and social dimensions.

We validated our hypotheses of observing a positive correlation between synchrony and entrainment loop with engagement and warmth, while we see an halo effect for competence. Thus, we can say that reciprocal adaptation, which is assessed via our measures, also has a direct relation with engagement and warmth.

Our reciprocal adaptation measures (three synchrony measures and entrainment loop measure) can be used to evaluate if the agent produced human-like behaviors with reciprocal adaptation for a human-agent interaction. This can be done by comparing the values obtained by the human-agent interaction against those obtained from human-human interaction. To detail, the human-agent interaction quality can be assessed by checking if the results of the agent, obtained via our reciprocal adaptation measures, show similar distributions with those of the real human-human interaction for both synchrony behaviors including their absence of response and behavior entrainment loop.

We are currently developing a predictive model of conversational agents with reciprocal behavior adaptation capability learned on human-human dyadic interaction. In the next future, we plan to use our reciprocal adaptation measures to objectively validate the generated agent's behaviors obtained from our predictive models to assess the quality of the interaction and the perception of the agent.

REFERENCES

Ashenfelter, K. T., Boker, S. M., Waddell, J. R., and Vitanov, N. (2009). Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4):1072.

Bakhshayesh, H., Fitzgibbon, S. P., Janani, A. S., Grummett, T. S., and Pope, K. J. (2019). Detecting synchrony in eeg: A comparative study of functional connectivity measures. *Computers in biology and medicine*, 105:1–15.

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE.

Beňuš, Š., Gravano, A., and Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, 43(12):3001–3027.

Bernieri, F. J., Reznick, J. S., and Rosenthal, R. (1988). Synchrony, pseudosynchrony, and dissynchrony: measuring the entrainment process in mother-infant interactions. *Journal of personality and social psychology*, 54(2):243.

Bernieri, F. J. and Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony.

Bernstein, M. J., Sacco, D. F., Brown, C. M., Young, S. G., and Claypool, H. M. (2010). A preference for genuine smiles following social exclusion. *Journal of experimental social psychology*, 46(1):196–199.

Biancardi, B., Cafaro, A., and Pelachaud, C. (2017). Analyzing first impressions of warmth and competence from observable nonverbal cues in expert-novice interactions. In *Proceedings of the 19th acm international conference on multimodal interaction*, pages 341–349.

Boker, S. M., Rotondo, J. L., Xu, M., and King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological methods*, 7(3):338.

Burgoon, J. K., Stern, L. A., and Dillman, L. (1995). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge University Press.

Cafaro, A., Wagner, J., Baur, T., Dermouche, S., Torres Torres, M., Pelachaud, C., André, E., and Valstar, M. (2017). The noxi database: multimodal recordings of mediated novice-expert interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 350–359.

Campbell, N. (2008). Multimodal processing of discourse information; the effect of synchrony. In *2008 Second International Symposium on Universal Communication*, pages 12–15. IEEE.

Cappella, J. N. (1997). Behavioral and judged coordination in adult informal social interactions: Vocal and kinesic indicators. *Journal of personality and social psychology*, 72(1):119.

Chang, W. I. and Lawler, E. L. (1994). Sublinear approximate string matching and biological applications. *Algorithmica*, 12(4):327–344.

Chartrand, T. L. and Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893.

Chartrand, T. L. and Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual review of psychology*, 64:285–308.

Condon, W. S. and Ogston, W. D. (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of nervous and mental disease*.

Condon, W. S. and Ogston, W. D. (1967). A segmentation of behavior. *Journal of psychiatric research*, 5(3):221–235.

Condon, W. S. and Sander, L. W. (1974). Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. *Science*, 183(4120):99–101.

Cuddy, A. J., Glick, P., and Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in organizational behavior*, 31:73–98.

Delaherche, E. and Chetouani, M. (2010). Multimodal coordination: exploring relevant features and measures. In *Proceedings of the 2nd international workshop on Social signal processing*, pages 47–52.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., and Cohen, D. (2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, 3(3):349–365.

Dermouche, S. and Pelachaud, C. (2019). Engagement modeling in dyadic interaction. In *2019 international conference on multimodal interaction*, pages 440–445.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.

Ekman, P. and Friesen, W. (1982). Felt, false, miserable smiles. *Journal of Nonverbal Behavior*, 6(4):238–251.

Hale, J., Ward, J. A., Buccheri, F., Oliver, D., and Hamilton, A. F. d. C. (2020). Are you on my wavelength? interpersonal coordination in dyadic conversations. *Journal of nonverbal behavior*, 44(1):63–83.

Heimerl, A., Baur, T., Lingenfelser, F., Wagner, J., and André, E. (2019). Nova - a tool for explainable cooperative machine learning. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), 3-6 September 2019, Cambridge, UK*, pages 109 – 115.

Hess, U., Beaupré, M. G., Cheung, N., et al. (2002). Who to whom and why–cultural differences and similarities in the function of smiles. *An empirical reflection on the smile*, 4:187.

Hess, U., Houde, S., and Fischer, A. (2014). Do we mimic what we see or what we know. pages 94–107.

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

Knapp, M. L., Hall, J. A., and Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Lau, S. (1982). The effect of smiling on person perception. *The Journal of social psychology*, 117(1):63–67.

Leander, N. P., Chartrand, T. L., and Bargh, J. A. (2012). You give me the chills: Embodied reactions to inappropriate amounts of behavioral mimicry. *Psychological science*, 23(7):772–779.

Lhoussain, A. S., Hicham, G., and Abdellah, Y. (2015). Adaptating the levenshtein distance to contextual spelling correction. *International Journal of Computer Science and Applications*, 12(1):127–133.

Mariooryad, S. and Busso, C. (2014). Correcting time-continuous emotional labels by modeling the reaction lag of evaluators. *IEEE Transactions on Affective Computing*, 6(2):97–108.

Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, pages 69–84.

Navarro, G. (2001). A guided tour to approximate string matching. *ACM computing surveys (CSUR)*, 33(1):31–88.

Niedenthal, P. M., Mermillod, M., Maringer, M., and Hess, U. (2010). The simulation of smiles (sims) model: Embodied simulation and the meaning of facial expression.

Ochs, M. and Pelachaud, C. (2013). Socially aware virtual characters: The social signal of smiles. *IEEE Signal Processing Magazine*, 30(2):128–132.

Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., and Scott Kelso, J. (2008). Social coordination dynamics: Measuring human bonding. *Social neuroscience*, 3(2):178–192.

Pickering, M. J. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2):169–190.

Prepin, K. and Pelachaud, C. (2011). Basics of intersubjectivity dynamics: Model of synchrony emergence when dialogue partners understand each other. In *International Conference on Agents and Artificial Intelligence*, pages 302–318. Springer.

Rauzy, S., Amoyal, M., and Priego-Valverde, B. (2022). A measure of the smiling synchrony in the conversational face-to-face interaction corpus paco-cheese. In *Workshop SmiLa, Language Resources and Evaluation Conference, LREC 2022*.

Reidsma, D., Nijholt, A., Tschacher, W., and Ramseyer, F. (2010). Measuring multimodal synchrony for human-computer interaction. In *2010 international conference on cyberworlds*, pages 67–71. IEEE.

Reis, H. T., Wilson, I. M., Monestere, C., Bernstein, S., Clark, K., Seidl, E., Franco, M., Gioioso, E., Freeman, L., and Radoane, K. (1990). What is smiling is beautiful and good. *European Journal of Social Psychology*, 20(3):259–267.

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.

Scharlemann, J. P., Eckel, C. C., Kacelnik, A., and Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, 22(5):617–640.

Schmidt, R. C. and Richardson, M. J. (2008). Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics*, pages 281–308. Springer.

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.

Tiedens, L. Z. and Fragale, A. R. (2003). Power moves: complementarity in dominant and submissive nonverbal behavior. *Journal of personality and social psychology*, 84(3):558.

Tschacher, W., Rees, G. M., and Ramseyer, F. (2014). Nonverbal synchrony and affect in dyadic interactions. *Frontiers in psychology*, 5:1323.

Varni, G., Volpe, G., and Camurri, A. (2010). A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. *IEEE Transactions on Multimedia*, 12(6):576–590.

Wang, I. and Ruiz, J. (2021). Examining the use of nonverbal communication in virtual agents. *International Journal of Human–Computer Interaction*, 37(17):1648–1673.