MSc Interaction Technology
Final Project

Measuring Conversation Quality

Jens Salverda

Supervisor: dr. K.P. Truong,
Supervisor: dr. A.J. van Hessen

February 6, 2025

Human Media Technology (HMI)
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Abstract

Prior research has identified key factors contributing to enjoyable conversations, including rapport, synchrony and empathy among interlocutors. This study aims to operationalize these indicators within customer-agent interactions by examining both verbal and non-verbal elements of speech, such as pitch range, voice intensity and turn-taking. While also exploring the potential of physiological measures such as heart rate. The objective is to enable an assessment of conversation quality in terms of the amount of rapport the interlocutors have, which holds significant utility for contact centres. The application can empower guidance to customer service representatives, fostering alignment with customers and more harmonious interactions.

To achieve this, experiments are conducted involving dyadic conversations between a trained actor and participants (N = 6, aged 18–28). Each conversation consisted of a normal talking phase and a rapport-breaking phase to capture shifts in interaction dynamics. Multimodal data, including audio, video, and physiological measurements, were collected alongside participants' subjective perceptions of rapport via questionnaires. Key findings revealed significant differences between the normal and non-rapport phases in synchronized smiling, speech rate, and prosodic features such as the standard deviation of pitch and intensity. Machine learning models (Random Forest and Logistic Regression) achieved 84% accuracy in classifying rapport and non-rapport moments. However, the exclusion of nonverbal features, such as synchronized smiling and head movements, reduced classification accuracy to 58%, underscoring the importance of nonverbal cues in rapport measurement. This study provides insights into measuring and fostering rapport in customer-agent interactions, with implications for improving conversational quality and customer satisfaction in contact centers.

Keywords: AI mediated communication, Rapport, Synchrony

Contents

Introduction 1

1.1 What are we trying to achieve?

More and more companies are using chatbots on their website or in customer relation phone calls. This does not always work the way it is meant to work. Customers experience a lot of frustration when dealing with customer service chatbots. Even so, 72 % feel that using a chatbot for customer service is a waste of time. [1]. The goal of a customer chatbot is to provide efficient and effective customer support and assistance conversationally. Hereby customer agent chat bots reduce the cost and workload of human agents. But if chatbots fail to help customers conveniently and efficiently it is only damaging the client and thus the company.

AI-mediated communication is a field that wants to facilitate, enhance, and improve interactions between humans and computers by using artificial intelligence. When using AI in communications it is important to design, implement, and regulate these systems responsibly, AI-mediated communications wants to develop a foundational empirical understanding of their impact on a wide variety of behaviours, including impression management, trust, deception, language use, relationships, and other key factors[2]. AI mediated technology must make it possible that clients are not frustrated when using chatbots or virtual assistants when communicating with a company.

To enhance the quality of customer-agent interactions, this research focuses on rapport in phone conversations between a customer-agent and a client. Rapport is defined as "clicking" or "having chemistry together" [3]. Establishing rapport is considered crucial for effective communication in consumer-agent phone calls. If there is a mismatch in linguistic patterns or a lack of emotional resonance during phone calls, it can hinder the communication process. Pivotal roles in the formation of rapport are empathy and synchrony. Empathy is the ability to understand and share the feelings of another person [4]. Synchrony is defined as the matching of behaviours the adoption of similar behavioural rhythms, the manifestation of simultaneous movement and the interrelatedness of individual behaviours [5]. Research on how to analyze rapport during consumer-agent phone conversations is integral to the advancement of AI-mediated technology. It contributes to the creation of more user-friendly, emotionally intelligent, and trustworthy AI systems, fostering positive interactions and improving the overall effectiveness of AI in various applications.

It is well-established that non-verbal features, such as facial expressions, gestures, and body language, play a critical role in communication, accounting for a significant portion of conveyed meaning [6]. By comparing conversations with and without non-verbal cues, this study seeks to identify how rapport can be effectively fostered in phone-based communication, where traditional non-verbal channels are unavailable. Understanding these differences can guide the development of communication tools that do not have non-verbal features such as phone conversations.

For businesses this study is also important. This research is done together with Merlinq. Merlinq is a company that specialises in implementing innovative solutions. Merlinq wants to develop a tool that measures the quality of phone calls of customer agents. By improving our understanding of the features that influence rapport in phone conversations this can lead to the development of more effective communication tools and technologies. For businesses that provide customer support or sales services over the phone, the ability to measure and optimize rapport can lead to better customer experiences, increased customer satisfaction, and higher sales conversion rates. In an era where remote work and collaboration are increasingly common, understanding and promoting rapport in phone conversations can enhance teamwork, productivity, and the overall quality of remote interactions. Companies that rely on distributed teams can see improved performance and employee well-being. Rapport in phone conversations is not limited to business interactions. It can also impact personal relationships, therapy, and counselling. Improved rapport can lead to stronger personal and professional relationships, benefiting individuals and society at large.

This thesis presents a comprehensive study that delves into the dynamics of rapport-building during conversational interactions. By employing measurement techniques and machine learning models to analyze and classify rapport based on verbal and non-verbal features this research aims to provide valuable insights into the factors influencing rapport, ultimately contributing to the enhancement of customer experiences and the optimization of customer service practices.

1.2 Research Objective

The primary goal of this study is to find out how to measure and assess the level of rapport in customer-agent phone conversations. This involves quantifying various aspects of communication between participants, as well as modelling these interactions using machine learning to identify moments of non-rapport. The study aims to develop a set of metrics and methods that will allow us to quantitatively evaluate the level of rapport in these interactions.

To achieve this aim, the main research question has been formulated:

How can we quantitatively measure and assess the level of rapport in customer-agent phone conversations?.

To address the main research question, the following sub-questions are formulated:

1. What measurable features show significant differences between moments of rapport and non-rapport in conversations?
2. How well does a model perform on classifying rapport in conversations with and without non-verbal features?

The first sub-question focuses on identifying the features that distinguish moments of rapport from non-rapport in interactions. By comparing measurable aspects of communication, this analysis seeks to establish a foundation for understanding which elements are most critical in fostering or hindering rapport, enabling the quantification of these dynamics in conversations. The second sub-question explores the unique dynamics of phone conversations, where non-verbal cues—critical in face-to-face interactions—are absent. By excluding non-verbal features from the analysis, the contribution of these cues to rapport detection can be better understood.

To address these questions, a theoretical foundation is first established by defining key concepts such as rapport, synchrony, and interpersonal relationships. The research proposal has already examined the role of rapport in customer-agent interactions, and this background helps to contextualize the exploration of rapport's role in conversations while informing the methodological approach. The following sections outline how previous research has measured rapport, providing a framework for addressing the sub-questions through empirical analysis. By integrating these insights with the current study, the main research question is addressed.

1.3 Thesis outline

This thesis is split up into 10 chapters. The workflow for this thesis can be seen in figure 1.1. This chart provides a clear roadmap for the structure of this thesis. By following this workflow, the study systematically addresses its research objectives and sub-questions. The first four chapters provide the theoretical background and insight into the context and related topic of the subject. The next three chapters focus on the methodology and the process of collecting, analyzing and processing data. The last three chapters look into the findings and outcomes of this thesis.

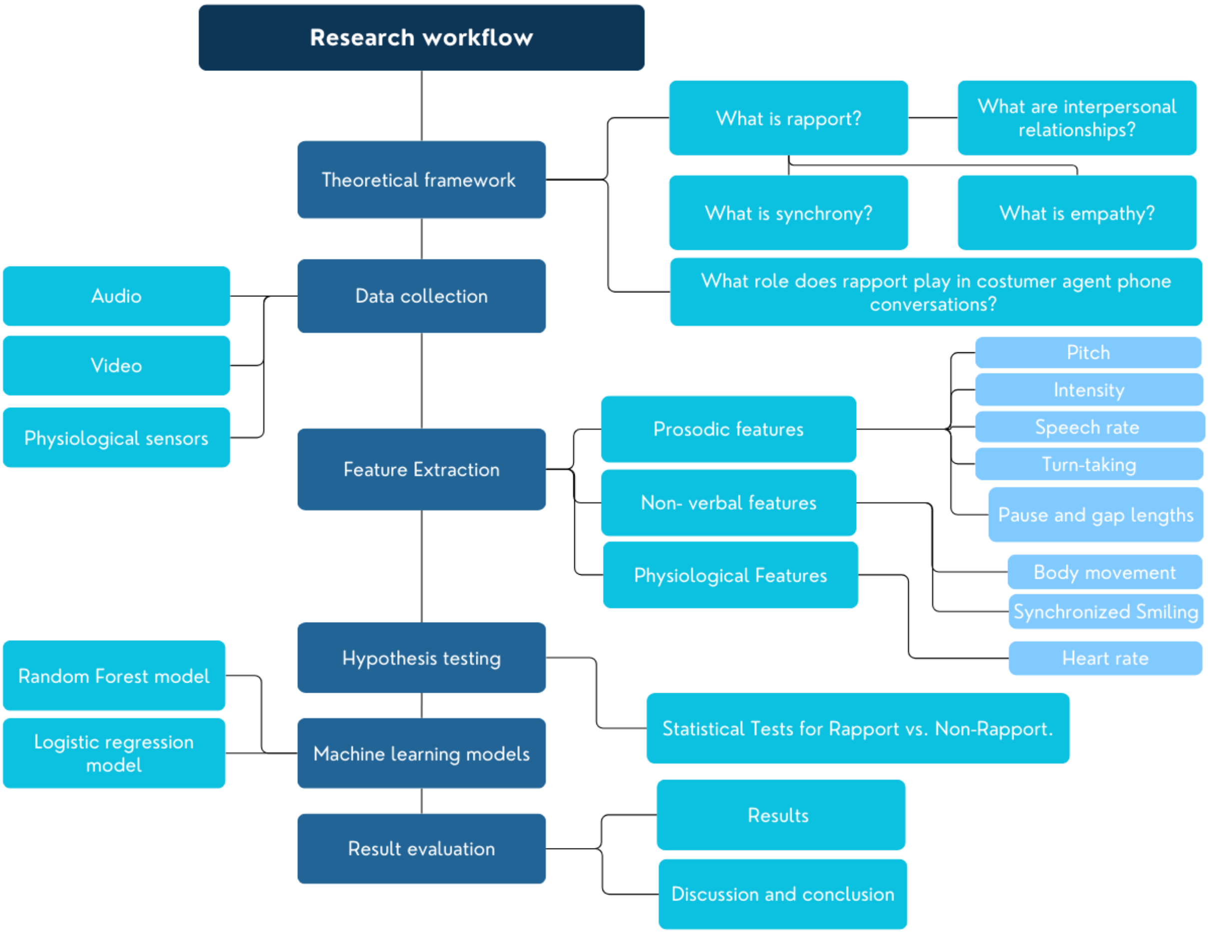

Figure 1.1: This diagram illustrates the workflow of this research, providing a step-by-step overview of the process used to address the main research question. The research begins with a review of the theoretical foundation, addressing key concepts such as rapport, synchrony, empathy, and interpersonal relationships. For data collection, the experimental setup is examined. The data is then processed to extract features relevant to rapport measurement. The outcomes of statistical tests and machine learning models are conducted, analyzed, and discussed. This includes interpreting the results, identifying limitations, and drawing conclusions about the role of verbal, non-verbal, and physiological features in rapport-building.

Principles | 2

This chapter examines the relevant principles that need to be considered, addressing the following questions: What is rapport? What are interpersonal relationships? What is synchrony? To comprehensively understand rapport, the foundational concepts of rapport, synchrony, and empathy are explored, as they play a pivotal role in facilitating smooth and meaningful interactions between customers and agents. By analyzing these interconnected principles, this chapter aims to provide a robust framework for understanding how rapport is built, maintained, and its significance in fostering meaningful interactions.

2.1 Rapport

Rapport is defined as "clicking" or "having chemistry together" [3]. It goes beyond mere acquaintance and implies a sense of mutual respect, comfort, and ease in communication. When there is rapport between people, there is a feeling of being on the same wavelength, which facilitates smooth and enjoyable interactions.

During the experience of a high degree of rapport, participants in the interaction form a cohesiveness, and become unified, through the expression of mutual attention to and involvement with one another. Their focus is directed toward the other and is other-involved. They experience the feeling as one of intense mutual interest in what the other is saying or doing [7]. In The Nature of Rapport and Its Nonverbal Correlates by LandaTickle-degnen and Robert Rosenthal, it is argued that there are three essential components during the experience of a high degree of rapport. This comes down to: *mutual attentiveness*, *positivity*. and *coordination*. In figure 2.1 it can be seen how those three components evolve over time.

Turn-taking in a conversation is a sign of having rapport. Yokozuka et al.(2020) looked into turn-taking and vocal pitch synchrony during creative problem-solving communication and investigated their possible relationship with rapport. They showed that turn-taking, rather than total utterances, is significantly positively correlated with rapport, while vocal pitch synchrony did not contribute to rapport explanations. They suggest that in the creative problem-solving discussion, the more turn-taking the conversation has, the more you feel rapport.[8] With this information, Yokozuka concludes that the amount of turn-taking is a reliable non-verbal predictor of rapport, even in cognitive goal-oriented communication.

Research has shown a connection between non-conscious mimicry, rapport, and interpersonal liking. A study by Lakin and Chartrand (2003) delves into the phenomenon of non-conscious mimicry, highlighting its significance in social interactions[9]. Non-conscious mimicry, often manifesting as subtle mirroring of another person's behaviours, is closely

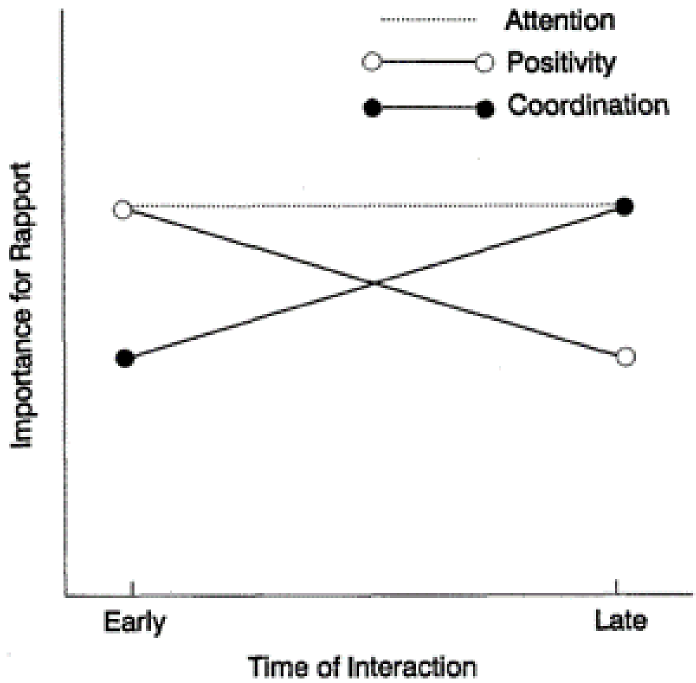

Figure 2.1: Relative importance given to each of the three components of rapport from early to late during an interaction [7] Different components of rapport gain or lose importance as the conversation unfolds, highlighting the complexity of rapport-building across time. What can be seen is that early interactions rely more heavily on positivity, while later phases benefit more from coordination, with attention remaining a constant requirement throughout.

tied to the establishment of rapport and the development of interpersonal liking. Rapport can increase the behavioural mimicry between two people[10] [11], and mimicry also increases rapport.

The scope of rapport extends beyond physical interactions. Gratch et al.(2006) explore rapport in virtual interactions, unveiling that a virtual listener's responsiveness significantly influences the speaker's willingness to communicate. The subject may even get quite frustrated when feedback is absent. Interestingly, the amount of turn-taking emerges as a key predictor of rapport, showcasing its relevance even in cognitive goal-oriented virtual communication [12]. The subjects in responsive condition talked significantly longer both in terms of overall time and word count. This shows that rapport is also felt in virtual environments.

In essence, whether in face-to-face encounters, virtual spaces, or marketing interactions, rapport emerges as a dynamic force, weaving connections through mutual attention, positivity, coordination, and the subtle dance of mimicry. Rapport has a direct impact on customer satisfaction. The quality of rapport at the outset plays a crucial role in shaping the trajectory of future business interactions [13]. A deficiency in rapport during an initial meeting can result in both immediate and enduring negative consequences. Specifically, an initial encounter lacking rapport is unlikely to evolve into a collaborative and mutually beneficial business relationship. However, Kaski et al.(2018) emphasize that a positive rapport established during the first sales meeting has the potential to mitigate the impact of weaker interaction performance in subsequent stages.

2.2 Interpersonal relationships

Interpersonal relationships refer to the connections, associations, or affiliations between individuals. These relationships can vary widely, encompassing everything from casual acquaintance ships to deep, intimate connections. They are a fundamental aspect of human experience and play a crucial role in shaping one's well-being, emotional health, and overall life satisfaction [14]. Interpersonal relationships can exist in various contexts, such as familial, social, romantic, or professional. Interpersonal relationships form the foundation upon which rapport is built. These relationships represent the broader social and emotional context, encompassing trust, mutual respect, and coordinated behaviours. They set the stage for rapport to emerge by fostering an environment where individuals feel connected and engaged.

2.3 Coordinated behaviours in conversations

Interactions between two people are affected by the quality of their relationship. Coordinated behaviours in conversations refer to the ways individuals interact with each other in a synchronized and cooperative manner. These behaviours are crucial for effective communication and building positive relationships. The higher the level of trust between them, the more satisfying the relationship between them is and the higher level of willingness of the two individuals is to exhibit extra-role behaviours. This is a behaviour that goes beyond the expected actions of a particular role or task. In conversations, coordinated behaviour typically involves individuals working together to achieve common goals, maintain social harmony, and ensure effective communication. While some behaviours are directly related to the tasks or roles individuals have within the conversation, extra-role behaviours extend beyond these specific roles and tasks. They contribute to the overall effectiveness of coordinated behaviour. Role development during a dyadic relationship makes sure the relationship progresses. The individuals start to become more familiar with each other. They may engage in more frequent and meaningful interactions and gradually move from strangers to acquaintances. As the relationship develops, trust plays a crucial role in facilitating this progression. The level of trust can be categorized as cognitive trust, affective trust, and behavioural trust, moving from the farthest to the closest [15]. This section explores the behaviors individuals engage in during relationships that influence their dynamics. To achieve this, examples of how specific behaviors impact relationships are examined.

For example, what happens when people smile at each other during a conversation? Gironzetti et al. (2016) look into how people allocate their resources when engaged in face-to-face conversations. They reveal that in humorous conversations, individuals not only reciprocate each other's smiles but also match the intensity of their smiles. This analysis identified distinct synchronous patterns of smiling and non-smiling, suggesting a multimodal connection between humorous events and smiling intensity among conversation partners[16].

In conversations, individuals often exhibit various behaviours, such as head movement. Hale et al. (2020) conducted a study on the head movements of individuals engaged in conversation, emphasizing that head motion is a precise parameter. However, the timing of this coordination remains unclear, restricting the capacity to formulate theories about the neural and cognitive mechanisms underlying social coordination [17]. Head movement synchrony is also researched during therapy sessions. Bhatia et al.(2019) explored the interpersonal coordination of head movements between patients and therapists. Their findings revealed a strong influence of patient-therapist head movement synchrony, highlighting the significance of this coordination in therapeutic interactions [18].

Another crucial role in interactions is the words people use when speaking with each other. Language is a fundamental aspect of human communication, and understanding linguistic elements contributes significantly to the analysis of coordinated behaviours in conversations. Linguistics are analyzed with, for example, text analysis. Text analysis can be used to assess the degree to which people coordinate their word use. Niederhoffer et al. (2002) uses text analysis to assess the degree to which people coordinate their word use in natural conversation. They found that individuals in dyadic interactions exhibit linguistic style matching on both the conversation level as well as on a turn-by-turn level [19].

Coordinated behaviours are the observable actions that reflect alignment and mutual engagement between individuals in a conversation. These behaviours are fundamental building blocks of rapport, as they foster mutual understanding and reinforce social bonds. By fostering seamless and enjoyable interactions, coordinated behaviours create a fertile ground for rapport.

2.4 Synchrony

Synchrony is defined as the matching of behaviours, the adoption of similar behavioural rhythms, the manifestation of simultaneous movement and the inter-relatedness of individual behaviours [5]. Research has shown synchrony to be related to positive affect in interactions and interpersonal liking and smoothness of interactions. Synchrony can increase people their self-esteem [20].

With body movement synchrony, Tsuchiya et al. (2020) state that you can predict the degree of information exchange [21]. They state that body movement synchrony occurs in a natural conversation. Their study revealed that the body movement synchrony of pairs who talked with each other was significantly higher than that of pairs who did not talk with each other and that this synchrony was positively associated with the degree of information exchange.

Synchrony also happens between people's heart rates [22]. Coutinho et al.(2020) found dyadic synchrony within couples in heart rate and heart rate variability. Another example of heart rate synchrony is in therapeutic contexts. Smits et al.(2020) looked at the correlation between heart synchrony and movement synchrony between patients and therapists and the effects on the personality problems of the patients. They found significant synchrony in most sessions between heartbeat synchrony and movement synchrony [23].

Interestingly, synchrony exhibits gender nuances. Fujiwara et al.(2019) reveal that female dyads display a higher degree of synchrony than their male counterparts, suggesting that female pairs move with similar timing, reflecting a coordinated dance of gestures and expressions [24].

Synchrony is something else than rapport. Synchrony refers to the coordination, alignment, or mirroring of behaviours, actions, or expressions between individuals engaged in social interaction. It involves a temporal matching of movements, speech patterns, or other nonverbal cues, creating a sense of harmony or connection. While, rapport is a positive and harmonious connection between individuals, encompassing shared interests, mutual respect, and a sense of camaraderie. Synchrony can be a nonverbal indicator of rapport[5]. Synchrony directly supports rapport by creating a behavioural alignment that fosters trust and smooth communication. For example, when conversational partners synchronize their gestures, movements, or vocal rhythms, it reinforces a shared rhythm and mutual focus, which are core components of rapport. Additionally, synchrony strengthens interpersonal bonds by facilitating a seamless flow in interactions, reducing potential misunderstandings, and signaling attentiveness and mutual respect.

Synchrony is also something else than convergence. Both are similar to each other, convergence involves elements arriving from distinct paths and merging, while synchrony encompasses occurrences that unfold simultaneously or operate at a uniform pace [25]. As illustrated in figure 2.2, convergence involves a gradual alignment of parameters over time, accounting for shifts and variances, while synchrony, demonstrated in the right panel of figure 2.2, captures similarity in relative values without necessitating absolute convergence. It's a simultaneous, harmonious occurrence, separate from the gradual merging seen in convergence.

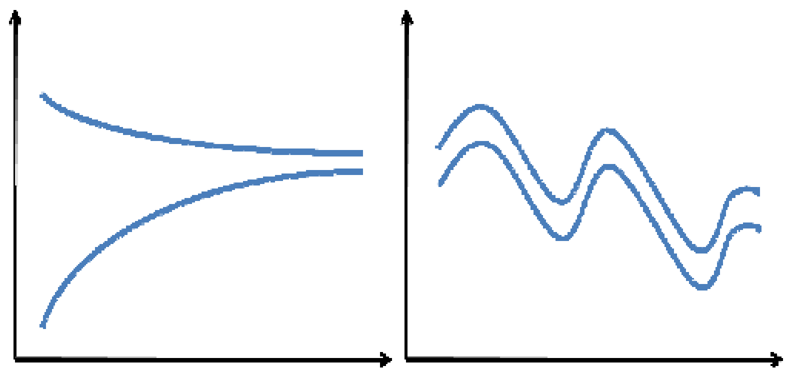

Figure 2.2: Distinct Modes of Similarity - Convergence(left) and Synchrony(right) [25]. On the left side, the concept of convergence is illustrated. Convergence refers to a gradual reduction in differences between two individuals' behaviours, attitudes, or emotional states as an interaction progresses. On the right side, synchrony is depicted as two wave-like patterns that maintain a consistent, oscillating rhythm while remaining in phase with one another.

Both synchrony and convergence are integral components of coordinated behaviour, offering nuanced perspectives on how individuals align in timing, form, and content during interactions. While synchrony focuses on the temporal aspects of coordination, convergence broadens the scope to include alignment across various modalities, providing a more comprehensive understanding of social coordination.

In this study, our main goal is to measure rapport. Synchrony and rapport are closely related. Synchrony is a behavioural coordination that underlies the development and maintenance of rapport[5]. Whether through nonverbal cues, emotional tone, or conversational rhythm, synchrony enhances the quality of interactions and strengthens the bond between individuals in social and professional settings.

2.5 Empathy

Empathy is a key component of effective interpersonal relationships. It involves the ability to understand and share the feelings of another person. When individuals in a relationship demonstrate empathy, they are better equipped to comprehend the emotions and perspectives of their counterparts. This understanding fosters a sense of connection, mutual support, and responsiveness, laying the foundation for more meaningful and positive interactions [4]. In practical terms, empathy allows individuals to validate each other's experiences, providing emotional support and creating an environment where people feel heard and understood. This, in turn, strengthens the bond between individuals and contributes to the overall health of the relationship.

Empathy is important in interpersonal relationships and coordinated behaviour in conversations. By acknowledging and validating the emotions and perspectives of others empathy fosters the relationship between people. It creates a supportive and open environment by showing understanding and compassion. In the description of empathic communication made by Hogan et al. (1975) there are three types of empathic communication found: attentive, affective, and cognitive [26]. Attentive empathy is being a "tactful and appreciative listener" and thus listening actively to customers. Affective empathy refers to the capacity to share and resonate with the emotional experiences of others. It involves feeling an emotional response that corresponds to what another person is feeling. Cognitive empathy involves understanding and intellectually grasping another person's perspective, thoughts, and emotions. It is the ability to comprehend someone else's point of view, even if you may not share the same emotional experience.

In phone conversations, empathy is used to develop rapport. According to Philip et al. (2020) suicide prevention helpline counsellors use three strategies to develop rapport with clients over the telephone [27]. Empathy, emphasis on para-language cues and intentional harmonisation. The counsellors reported intentionally moderating their voice tones and other para-language cues, so clients felt comfortable to continue disclosing. They also reported intentionally mirroring clients in words, phrases, language style, tone, speed, and pace. This was done to help clients experience a sense of connection and to reinforce that the telephone counsellor was psychologically "in tune" with them. Imel et al. reported that the more synchronized the voice pitch was in psychological counselling, the higher the empathy [28].

Empathy is also used in customer-agent phone calls. Clark et al.(2013) looked at how empathy is expressed in customer service and how to explore whether empathic communication is beneficial in aftermarket customer calls. They found that empathy can help mitigate the tensions underlying the shared purposes that engender customer calls [29]. Empathy enables agents to better understand the emotional state and needs of people. This helps establish a connection between individuals. This research looks at that connection. Chapter 3.2 looks further into the role empathy plays in customer agent phone calls.

Empathy relates to rapport by providing the emotional depth and connection required for individuals to feel valued and understood. While synchrony focuses on the behavioural coordination that builds rapport, empathy ensures that interactions are emotionally fulfilling and aligned with the needs and perspectives of others. Together, these elements create a comprehensive foundation for rapport-building.

2.6 Summary

This chapter looked at the questions: What is rapport? What is synchrony? What are interpersonal relationships? It outlined how interpersonal relationships are built through coordinated behaviours and trust, and how they evolve through mutual engagement. Synchrony, defined as the matching of behaviours and rhythms, plays a critical role in enhancing communication. Empathy is the ability to understand and share the feelings or perspectives of another person. Rapport refers to a harmonious and positive connection between individuals during interactions. Rapport, characterized by trust, empathy, mutual understanding and a smooth interaction, emerges as the result of these elements reinforcing each other to create harmonious connections.

Overall, this chapter establishes that rapport is the result of coordinated behaviours, empathy and synchrony. This makes rapport a crucial element for effective interactions. This chapter lays the groundwork for the following sections, which will further explore how rapport can be measured and quantified.

Related Work 3

This chapter aims to answer the question: How can rapport be measured? To achieve this, existing research on rapport measurement is reviewed, identifying the methods used to quantify rapport. The analysis begins with a focus on prosodic features, such as pitch, intensity, and speech rate, which provide insight into vocal alignment and conversational engagement. Next, non-verbal cues are examined, along with their role in fostering connections between individuals. The discussion then shifts to methods that integrate physiological data, such as heart rate synchrony, to capture underlying emotional and cognitive states during interactions. Additionally, approaches for quantifying emotional dynamics are explored, emphasizing how subjective and objective measures are combined to understand rapport. Finally, statistical methods used in rapport and synchrony studies are reviewed. This chapter aims to establish a foundation for the methodological framework.

3.1 Prosodic measurements

Prosodic features, such as pitch, intensity, and speech rate, have been widely studied as indicators of synchrony and rapport. Research shows that alignment in these features reflects engagement and connection between speakers. Firstly, **pitch** or fundamental frequency refers to the perceived frequency of a speaker's voice. It plays a crucial role in expressing emotions and signalling engagement or interest. De Looze et al. (2014) found that higher synchrony in pitch correlates with smoother conversational flow and mutual engagement [30]. Another feature is the **intensity**, which measures the loudness of speech and intensity reflects the speaker's energy or emphasis[31]. Studies such as Levitan et al. (2011) and De Looze et al. (2014) demonstrated that variations in intensity are strongly tied to perceived speaker involvement. Next up, **speech rate** or tempo is looked at. Speech rate refers to the number of syllables or words spoken per unit of time. A faster speech rate can indicate enthusiasm, urgency, or confidence, while a slower rate may convey thoughtfulness or hesitation. Levitan et al. (2012) demonstrated that speakers entrain on speech rate during cooperative interactions, with entrainment being particularly prevalent in male-male and mixed-gender pairs [31]. This alignment has been shown to enhance dialogue smoothness and naturalness, key indicators of rapport. These prosodic features convey the emotional tone, emphasis, and rhythm of speech. Assessing these features in conversations is challenging because one participant is often silent while the other is speaking. It is therefore important to find a way to compare prosodic features between multiple interlocutors and to ensure they are time-aligned. Current approaches use three types of methods: utterance or turn-level-based [31], time-aligned moving average method [32] and hybrid utterance-sensitive window extraction [30]. The choice of method depends on the need to prioritize natural turn-taking boundaries or ensure balanced, time-aligned analysis across speakers

Other prosodic measurements that reflects the timing and rhythm of the spoken interaction are **Pauses and gap lengths**. Pause and gap lengths between turns serve as punctuation marks in conversation. Edlund et al. (2009) look at the pauses and gap lengths that happen during conversations. They state that a gap in a conversation is a silence where the other person starts talking again. A pause is a silence where the same person who initiated the silence starts talking again [25]. Their findings show that short pauses may indicate a moment of thought or signal a transition between ideas, while longer gaps may suggest contemplation or invite others to join the conversation. The management of pauses contributes to the overall pace and rhythm of the interaction, influencing the perceived attentiveness and engagement of participants. Another example of a prosodic measurement that could indicate rapport in a conversation is **turn-taking**. Turn-taking is the rhythmic exchange of speaking turns between conversation participants. It facilitates a balanced and organized conversation, preventing interruptions and allowing each participant an opportunity to contribute. Smooth turn-taking enhances the flow of communication and indicates a cooperative and respectful interaction. Yokozuka et al. (2020) show that the more turn-taking the conversation has, the more you feel rapport [8]. These studies highlight how pauses, gaps, and turn-taking influence rapport by shaping the rhythm and dynamics of interactions. Effective management of these features signals attentiveness, alignment, and shared understanding, which are critical components of rapport-building.

3.2 Non-verbal cues

Non-verbal cues play a pivotal role in interpersonal communication by conveying unspoken nuances, emotions, and relational dynamics. Prior research highlights that these cues, such as smiling, head movements, and eye gaze, significantly influence the perception and establishment of rapport in social interactions. It is well-established that non-verbal features accounting for a significant portion of conveyed meaning [6]. Understanding and analysing non-verbal cues are central to unravelling the intricacies of human interaction, especially in contexts such as consumer-agent phone calls.

For example, **body mirroring**, or mimicking the body language of another person, establishes a subconscious connection. When individuals unconsciously mimic each other's movements, it signals rapport and a sense of shared understanding. La France et al. (1979) have showed that body mirroring fosters a feeling of similarity and alignment, contributing to positive social dynamics [11].

Another example of a non-verbal cue is **smiling**. Smiling is a universal expression of positive emotions, warmth, and openness. Gironzetti et al. (2016) demonstrated that synchronized smiling not only reflects rapport but also strengthens emotional alignment between individuals [16]. In humorous conversations, participants tend to match the intensity and timing of their smiles, reinforcing shared emotional states. Facial cues provide a visual representation of shared emotions, signalling that the empathetic individual is attuned to the feelings being expressed.

To further illustrate, another feature studied is the movements of the head. **Head movements**, including nods and tilts, contribute to nonverbal communication. Nodding can signify agreement, understanding, or active listening, enhancing the sense of connection. Hale et al. (2020) found that synchronized head movements enhance the sense of connection and engagement during conversations, acting as non-verbal markers of rapport-building [17]. Lastly, **Eye gaze** is a powerful indicator of attentiveness, trust, and empathy. Tschacher et al. (2021) showed that eye gaze establishes a sense of intimacy and connection, facilitating emotional resonance between speakers. Conversely, avoiding eye contact may indicate discomfort or disengagement [33]. These findings demonstrate that non-verbal cues, particularly when synchronized, serve as visible indicators of rapport by signaling shared understanding, comfort, and mutual engagement.

3.3 Physiological cues

Together with non-verbal cues, physiological responses provide a unique and invaluable window into the visceral reactions that underlie our interactions. These responses, stemming from the autonomic nervous system, reflect the physiological changes individuals undergo during various communicative encounters. In the context of consumer-agent phone calls, exploring physiological responses becomes paramount in understanding the subtle, often subconscious, reactions that influence the dynamics of the conversation. When individuals experience a sense of connection, comfort, and positive engagement in social interaction, their physiological responses may reflect a state of relaxation and well-being.

When people are in sync with each other during interactions, their physiological states may align. Dyadic synchrony is found within couples in **heart rate** and **heart rate variability** [22]. Coutinho et al. (2020) aimed to explore the presence of couples' patterns of physiological synchrony based on cardiac activity, measured by heart rate and heart rate variability. Heart rate is a measure of the number of times the heart beats per minute. It is a fundamental physiological parameter that reflects the rate at which the heart pumps blood throughout the circulatory system. Heart rate can be influenced by various factors. Heart rate variability is a measure of the variation in time between successive heartbeats. Rather than focusing solely on the average heart rate, heart rate variability examines the fluctuations in the intervals between individual heartbeats. Coutinho et al. (2020) found that couples exhibited greater synchrony during positive interactions, suggesting that physiological alignment reflects empathy and connection. HRV, which captures fluctuations between successive heartbeats, provides a robust indicator of autonomic nervous system activity and the ability to adapt to social dynamics.

Another physiological indicator that may show a shared emotional experience is **galvanic skin response**, also known as electrodermal activity. Galvanic skin response measures the electrical conductance of the skin, which can be influenced by factors such as sweating and emotional arousal. Bouscein et al. (2013) found that while sweat secretion plays a major role in thermoregulation and sensory discrimination, changes in skin conductance are also triggered robustly by emotional stimulation [34]: the higher the arousal, the higher the skin conductance. Critchley et al. (2002) state that skin conductance is not under conscious control. Instead, it is modulated autonomously by sympathetic activity which drives aspects of human behaviour, as well as cognitive and emotional states [35]. Skin conductance, therefore, offers direct insights into autonomous emotional regulation. Thus galvanic skin response can provide some insights into a person's physiological state.

Then there are physiological responses that potentially provide insight into empathy, rapport or synchrony but are not a direct and specific indicator. For example, blood pressure, emotions and stress can influence physiological responses such as blood pressure. positive social interactions and feelings of connection may contribute to relaxation and lower blood pressure, while stress or discomfort may lead to increased blood pressure. For example, it is found that an increase in blood pressure greater than 25–40% occurs within 30 sec after the initiation of human speech [36]. So while blood pressure can be influenced by emotional states and social interactions it is not an exclusive or specific marker of empathy, synchrony or rapport.

Another example of this is the respiration rate. Various factors can influence respiration rates, such as physical activity, health conditions, and individual differences. The interpretation of physiological signals is nuanced, and context plays a crucial role.

3.4 Quantifying emotional dynamics in rapport

Because rapport is a feeling, it is difficult to quantify something that is happening as rapport. This section describes how researchers measure rapport as an emotional construct by combining subjective assessments, such as questionnaires and observer ratings, with objective measurements, behavioural or physiological cues. By comparing these approaches, a clearer understanding emerges of how emotional dynamics such as trust, comfort, and engagement are captured and analyzed.

This subjective measure could be a self-reported questionnaire about the experience. An example of a questionnaire that is used to measure the amount of rapport participants feel during an interaction is in a study done by Lucas et al. In this study they research the amount of trust people feel in a robot which makes a conversational error[37]. A 5-point scale is specifically designed to measure the feeling of having a close and harmonious relationship even after a single interaction. Example questions from this study can be seen in figure 3.1.

Another example of a study that uses a questionnaire is the study of Niederhoffer et al. (2002) They used an interaction rating questionnaire [19]. This is a questionnaire that asks the participant to rate and give feedback on how the interaction was going. The questionnaire consists of either 10 or 15 questions with the focus of assessing the degree to which participants enjoyed the conversation, and various measures of their comfort level. The questions were based on the degree to which

Each question is rated on a 5-point scale. R indicates reverse-coded items
Niki created a sense of closeness or camaraderie between us.
Niki created a sense of distance between us (R).
I think that Niki and I understood each other.
Niki communicated coldness rather than warmth (R).
Niki was warm and caring.
I wanted to maintain a sense of distance between us (R).
I felt I had a connection with Niki.
Niki was respectful to me.
I felt I had no connection with Niki (R).
I tried to create a sense of closeness or camaraderie between us.
I tried to communicate coldness rather than warmth (R).

Figure 3.1: Questionnaire that is asked after the participant had an interaction with a robot. In this questionnaire the participant can report how much rapport they felt during the interaction [37]. Reverse-coded items are included to make sure the respondents are giving consistent answers.

participants felt the interaction went smoothly, they felt comfortable during the interaction, and they truly got to know the other participant.

An additional case that uses a questionnaire to have a subjective measure of the rapport the participants feel during a conversation is the study of Bernieri et al. (1991) In their study in which they analysed the movement coordination that happens between conversation partners, they used a questionnaire with 27 items [5]. This questionnaire covered various dimensions of emotional affect and rapport with eight-point unipolar rating scales. For each item, subjects responded to the question "Exactly what were you experiencing during this brief interaction?"

Another way to get a subjective measure that can be used to compare in a study is an objective team of judges such as used by De Looze et al. (2014). In their study they look at automatic measurements of prosodic features to quantify the relationship between the people having a conversation. To get a subjective view of the rapport happening in a conversation they let judges annotate chunks of audio on aspects such as flow, engagement, and mutual involvement with a four-point Likert scale [30].

Emotional aspects of rapport are measured through subjective and objective methods. Self-reported questionnaires capture participants' internal perceptions of rapport, while observer ratings provide external perspectives on conversational dynamics. To measure rapport effectively, it is essential to use a multidimensional approach that integrates both subjective and objective methods. Subjective tools reveal how rapport feels, while objective tools uncover the mechanisms behind it.

3.5 Statistical correlation methods in rapport and synchrony studies

Statistical correlation methods are widely used to examine the relationship between features such as pauses, prosodic alignment, and rapport or synchrony during conversations. By calculating correlation coefficients, researchers can objectively evaluate how behaviors evolve over time and compare patterns of interaction across conversations. Correlation coefficients allow comparisons across conversations, enabling to exploration of patterns of interaction at a broader level, such as comparing rapport in different conversations. The following examples demonstrate how correlation methods are employed to analyze conversational dynamics and provide a foundation for understanding their role in rapport and synchrony.

The Pearson coefficient was used by Edlund et al. (2009) in their study in which they looked at pause and gap length during conversations[25]. This is done by plotting the mean length of the pauses and gap lengths over time. The Pearson correlation was used to look at the strength of the convergence and synchrony that may happen in the gaps and pauses. The analyses were run separately for gaps and pauses, and split over dialogues as well as over the pooled dialogues.

Another example is De Looze et al. (2014) which uses the Pearson correlation coefficient to measure the linear dependencies between different acoustic features of the participants during a conversation[30]. When the Pearson coefficient has a high value, and thus has strong linear dependencies, it indicates a synchronous behaviour of the prosodic parameters over the analyzed fragment. Values lower than zero indicate strong asynchronous developments of the observed parameters. For values close to zero no linear correlation is observed a state of maintenance is present.

3.6 Summary

This chapter explored the question: How to measure rapport? To address this, existing research on rapport measurement was reviewed, focusing on different domains such as prosodic features, behavioral cues, and physiological responses. Prosodic features utilize various time alignment methods to analyze speech characteristics like pitch, intensity, and speech rate. These approaches demonstrate how prosodic features reflect alignment and engagement between speakers. Behavioral cues and physiological responses play integral roles as indicators of empathy, rapport, and synchrony in interpersonal interactions. These cues provide nuanced insights into the emotional and social dynamics between individuals, revealing the subtle intricacies of human connection. While behavioral cues offer observable signs of social dynamics and visible expressions of empathy, physiological responses provide additional insights into the internal states of individuals engaged in social interactions.

Measuring rapport requires a multidimensional approach, combining prosodic, behavioural, and physiological indicators with robust analytical methods. Each domain captures a unique aspect of a conversation, but together they may provide a comprehensive framework for understanding rapport in interactions.

Hypotheses 4

This chapter looks into the hypotheses of our research. These hypotheses are based on insights from previous literature and are aligned with the study's aim to quantitatively measure rapport in customer-agent phone conversations. The analysis begins with hypotheses related to acoustic features, followed by conversational and non-verbal features. Lastly, physiological features are looked into. This study aims to determine whether these features exhibit measurable differences between normal phases and non-rapport phases. Each section discusses the domains in detail, supported by relevant citations.

4.1 Hypotheses for prosodic features

Prosodic features such as pitch, intensity, and speech rate are integral to conversational synchrony and rapport. De Looze et al. (2014) found that pitch synchrony varies dynamically during conversations, reflecting shifts in engagement and emotional alignment [30]. Our expectations on the prosodic features are based on this paper. De Looze et al. (2014) demonstrated that alignment in pitch enhances conversational flow, which is an indicator of rapport. Higher pitch synchrony reflects mutual attunement and coordination in conversation. The alternative hypothesis for the median pitch is:

- Median pitch (H 1): The median pitch correlation is significantly different in the normal period compared to the non-rapport period.

Synchrony in the standard deviation of pitch, was found to correlate positively with conversational engagement and the perception of harmony between speakers [30]. Therefore the alternative hypothesis for the standard deviation of pitch is:

- Standard deviation of pitch (H 1): The standard deviation of pitch correlation is significantly different in the normal period compared to the non-rapport period.

Synchrony in median intensity was found to follow similar trends. Higher synchrony in median intensity correlates with speakers' engagement and affinity [30]. The alternative hypothesis for the median intensity is thus:

- Median intensity (H 1): The median intensity correlation is significantly different in the normal period compared to the non-rapport period.

The standard deviation of intensity showed fewer synchrony phases compared to the other features in the paper by De Looze et al. (2014). This can indicate that the standard deviation will show less of a variation the normal and the non-rapport periods during conversations. Nevertheless, in our study, it could show something and therefore the alternative hypothesis for the standard deviation of intensity is:

- Standard deviation of intensity (H 1): The standard deviation of intensity correlation is significantly different in the normal period compared to the non-rapport period.

Speech rate synchrony is also less aligned with rapport moments in comparison with the other features. It did not show a strong correlation with conversational flow or mutual engagement in the data analyzed by De Looze et al.(2014). However, it might still play a role. The alternative hypothesis for the difference in speech rate between the normal and non-rapport period is:

- Speech rate (H 1): The mean speech rate is significantly different in the normal period compared to the non-rapport period.

4.2 Hypotheses for conversational features

To assess whether the conversational features differed significantly between the normal and non-rapport-forming periods, turn-taking and pause and gap lengths are looked into. Yokozuka et al. (2020) found that smoother turn-taking correlates with mutual attentiveness and cooperation [8]. Therefore, the following alternative hypotheses is formulated for the difference in smooth turn-taking between the normal and non-rapport period:

- Smooth turn-taking(H 1): The amount of smooth turn-taking moments per minute is significantly different in the normal period compared to the non-rapport period.

Pauses and gap lengths punctuate conversations, reflecting alignment or disconnection. Edlund et al. (2009) showed that shorter pauses and gaps are associated with more cohesive and engaging interactions [25]. The alternative hypothesis for the difference in pause lengths between the normal and non-rapport periods is:

- Pauze length (H 1): The mean pause lengths are significantly different in the normal period compared to the non-rapport period.

The alternative hypothesis for the difference in gap lengths between the normal and non-rapport period is:

- Gap length (H 1): The mean gap lengths are significantly different in the normal period compared to the non-rapport period.

4.3 Hypotheses for non-verbal features

Smiles are universally recognized as indicators of positivity and rapport. Gironzetti et al. (2016) demonstrated that synchronized smiles enhance mutual understanding and alignment [16]. The alternative hypothesis for the difference in the amount of synchronized smiling between the normal and non-rapport period:

- Synchronized smiles (H 1): The mean amount of synchronized smiles is significantly different in the normal period compared to the non-rapport period.

Kwon et al. (2023) found that synchronized head movements correlate with higher levels of interpersonal coordination [38]. The alternative hypothesis for the difference in synchronized head movements between the normal and non-rapport period is:

- Synchronized head movements (H 1): The mean amount of synchronized head movement is significantly different in the normal period compared to the non-rapport period.

4.4 Hypotheses for physiological features

For the physiological feature, the paper of Smits et al. (2020) is explored. Significant heart rate synchrony was observed across most sessions between patients and therapists [23]. Their findings suggest that physiological synchrony, including heart rate, is a meaningful indicator of interpersonal coordination. Therefore, the alternative hypothesis for the difference in the amount of heart rate synchrony between the normal and non-rapport period:

- Heart rate synchrony (H 1): The heart rate variability correlation is significantly different in the normal period compared to the non-rapport period.

4.5 Summary

This chapter outlines hypotheses for identifying measurable features that differentiate rapport-forming and non-rapport periods in conversations. Those conversations will be held in our study, the next chapter will look into the data collection and experimental setup. These hypotheses are grounded in prior research and address verbal, conversational, non-verbal, and physiological domains. Each hypothesis reflects on how a rapport is measurable with a feature, making them integral to answering the research question: How can we quantitatively measure and assess the level of rapport in customer-agent phone conversations?

Data collection | 5

This chapter outlines the study conducted to answer the research question: What observable features indicate rapport in conversations? This question is addressed through experiments involving conversations that include both a phase where rapport is present and a phase where rapport is intentionally disrupted.

5.1 Experimental setup

This section describes the physical and technical setup of the studies, ensuring a controlled environment for data collection. The setup was designed to capture audio, video, and physiological data simultaneously, ensuring synchronized and comprehensive recordings of each interaction.

For the study each participant is engaged in a 15-minute conversation with a trained actor. The conversation centered around the participant's hobby, a topic chosen to evoke personal engagement and facilitate naturalistic interaction. The actor's role was twofold: to establish rapport during the initial phase of the conversation and subsequently to disrupt rapport during the latter phase.

The 15-minute conversation was divided into two distinct phases. During the first 10 minutes, the actor focused on building rapport with the participant through attentive listening, empathetic responses, and positive reinforcement of the participant's interests. This phase aimed to create a conducive environment for rapport formation.

Following the initial rapport-building phase, the subsequent 2 minutes were dedicated to rapport disruption. The actor employed subtle conversational cues and behavioral changes designed to introduce tension and diminish the sense of connection established earlier. These disruptions ranged from abrupt topic shifts to nonverbal cues signaling disinterest or disagreement.

After the 15-minute conversation each participant filled in a questionnaire. This questionnaire was designed to capture participants' subjective experiences of the interaction. While it was not a validated instrument, it was developed based on previous literature on rapport as described in section 3.4, ensuring that the questions addressed relevant aspects of interpersonal connection and conversational dynamics. The questionnaire consisted of both Likert-scale questions and open questions, designed to evaluate the following aspects:

- Perceived connection and understanding during the interaction.
- Identification of moments of tension or disconnection.
- General reflections on rapport-building and breakdown

Both the participant and the actor wore wristbands equipped with Galvanic Skin Response (GSR) and heart rate sensors throughout the duration of the conversation. These wearable devices enabled continuous monitoring of physiological arousal levels, providing objective measures of emotional engagement and stress responses during the interaction. Next to the physiological sensors, simultaneous video and audio recordings were made of each conversation to capture verbal and nonverbal cues exchanged between the participant and the actor. After the conversation, each participant filled in a survey.

5.2 Participant group

This section outlines the selection and characteristics of the participant group, the participant group for this study comprises 10 individuals, all aged between 18 to 28 years old, with an equal gender distribution of 5 men and 5 women. All participants are students currently residing in the Netherlands. Furthermore, the inclusion of an equal number of male and female participants seeks to account for potential gender-related differences in communication styles and rapport formation strategies.

This study involved human participants, and therefore ethical approval had been granted by the ethics committee. The ethical principles of respect, confidentiality, and informed consent were upheld during the study, and all participants were fully informed about the nature of the research, the voluntary nature of their participation, and their right to withdraw at any time. No harm or discomfort was caused to participants, and the study adhered to the ethical guidelines post-approval.

5.3 Materials

Various hard- and software were used to facilitate the study. Below, a list is provided in which these materials are present, together with the corresponding purpose:

- Shure wireless audio system for individual recording of the audio of both the participant and the actor. Can be seen in figure 5.2
- Alesis iO4 audio interface for connecting the sure wireless system with the laptop.
- 2x webcam logitech quickcam pro 9000 for video recording both the participant and the actor.
- Consensys shimmer 3 bundle for measuring the GSR and heart rate of both the participant and the actor. Can be seen in figure 5.1
- Laptop with OBS software for making sure both the audio and video recording start at the same time.
- Python for processing the data.
- Praat for processing the data.

Figure 5.1: Consensys Shimmer3 bundle, which was used in this study to collect physiological data such as Galvanic Skin Response (GSR) and heart rate [39].

Figure 5.2: Shure wireless audio system, which was employed for capturing high-quality individual audio recordings of the conversations between participants and the actor in the study. This system ensures that each participant's voice is clearly recorded without interference or overlap, which is crucial for analyzing acoustic features [40].

5.4 Measurements

To ensure a smooth and synchronized data collection process, OBS (Open Broadcaster Software) is utilized for recording both video and audio during the experiments. OBS was installed on a laptop and configured to record the video from the webcams and the audio from the microphones simultaneously. This setup allowed for the capture of high-quality video recordings of the participants' facial expressions and body movements, alongside the audio of their conversation.

By using OBS for recording, the data collection process was steamlined, ensuring that all relevant data for each study was captured consistently and efficiently, ready for further analysis using the methods described in later sections.

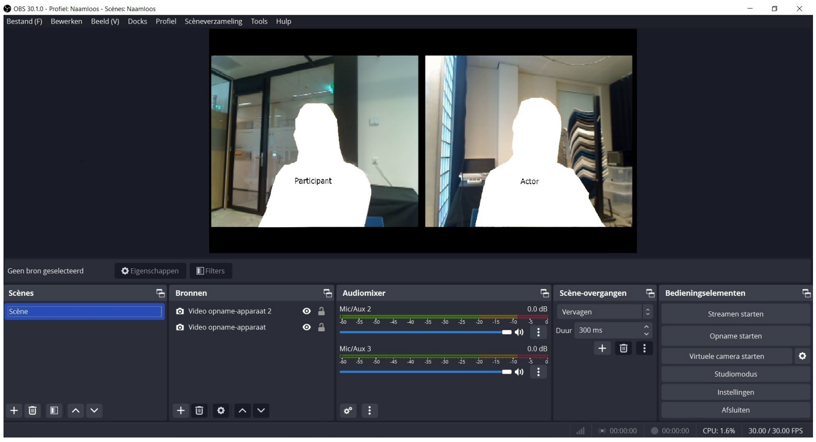

Figure 5.3: OBS (Open Broadcaster Software) interface set up for recording video and audio in an experimental setting. On the upper part of the screen, there are two video feeds displayed. These video feeds are labeled "Participant" and "Actor," indicating that the software is recording both the participant and the actor simultaneously. Below the video feeds, there's an audio mixer panel showing two audio input devices: "Mic/Aux 2" and "Mic/Aux 3". These represent the two separate microphones capturing the audio from both the participant and the actor.

In order to measure both heart rate from the participants and the actors during the studies, the Consensys Shimmer3 bundle, was utilized. The setup required attaching sensors to both participants and actors to continuously capture their physiological data throughout the conversation. On their ring-finger and pinky the electrothermal sensors were deployed and the PPG sensor was worn on the middle finger for measuring the heart rate.

Once the sensors were connected, the Consensys software was used to configure the devices. It was essential to ensure that the measurements—both heart rate and galvanic skin conductance—were accurately recorded with timestamps. This was done to synchronize these physiological measurements with the audio and video data collected through OBS. By enabling the timestamp feature, the heart rate and GSR data could later be perfectly aligned with the conversation flow, facilitating cross-modal analysis. The data was saved directly into CSV files, which provided a structured format for storing both the physiological data and the associated timestamps. This made it easier to merge and analyze the physiological responses alongside the other forms of data collected during the studies, helping to investigate the relationship between physiological states and rapport.

Although initially ten experiments were conducted, only six were included in the final analysis due to a technical issue. During data collection, the Shimmer3 wristband, used to measure physiological data such as heart rate and galvanic skin response, was not properly activated in four experiments. To maintain the integrity and reliability of the dataset, four experiments were excluded from further analysis. Despite the exclusion of these four experiments from the quantitative analysis, the questionnaire responses provided by these participants were still included. The questionnaire aimed to capture subjective perceptions of rapport and non-rapport moments, which do not rely on the missing physiological data. As these responses offer valuable qualitative insights, they were deemed relevant and incorporated into the overall analysis of participants' experiences.

Methodology 6

With the studies explained, the next step is to outline the methodology for analyzing the collected measurements. The goal is to address the following research questions:

- What measurable features show significant differences between moments of rapport and non-rapport in conversations?
- How well does a model perform on classifying rapport in conversations with and without non-verbal features?

The data is obtained in three different ways. The data is obtained through three different sources: audio recordings, video recordings, and physiological measurements. This methodology first examines how the data is used to address the first research question, followed by an exploration of how machine learning models are applied to answer the second question.

6.1 Feature analysis

A categorization of the methodological approaches was made to ensure a comprehensive analysis of both verbal and non-verbal aspects of the interactions. This categorization can be seen in the feature analysis flowchart in 6.1

The audio analysis is further divided into two key subcategories: acoustic and conversational features. In the acoustic domain, pitch, intensity, and speech rate are analyzed. Median pitch and standard deviation of pitch are calculated to assess both the central tendency and variability of the speaker's pitch during the conversation. Similarly, median intensity and standard deviation of intensity provide insights into the loudness and variability in speech delivery. The speech rate is measured using syllables per minute, helping to evaluate how quickly or slowly the speech was delivered. On the conversational side, turn-taking, pause length, and gap length are assessed. Turn-taking is examined by the method proposed by Yokozuka et al. (2020)[8], where smooth turn-taking is defined as instances when the change in speaker occurs within one second. Pause length and gap length are evaluated using the method by Edlund et al. (2009)[25], where the mean durations of pauses within a speaker's turn and the gaps between turns are measured.

The video analysis focuses on non-verbal cues, particularly head movement and smile synchrony. For head movement, Kwon's method is employed to analyze metrics such as density, mean phase difference, standard deviation of phase, and kurtosis. These metrics help in evaluating the synchrony and variability in the head movements of the participants. Similarly, smile synchrony is analyzed using Gironzetti's approach, with the same metrics—density, mean phase difference, the standard deviation of phase, and kurtosis—used to assess the synchronization of smiling

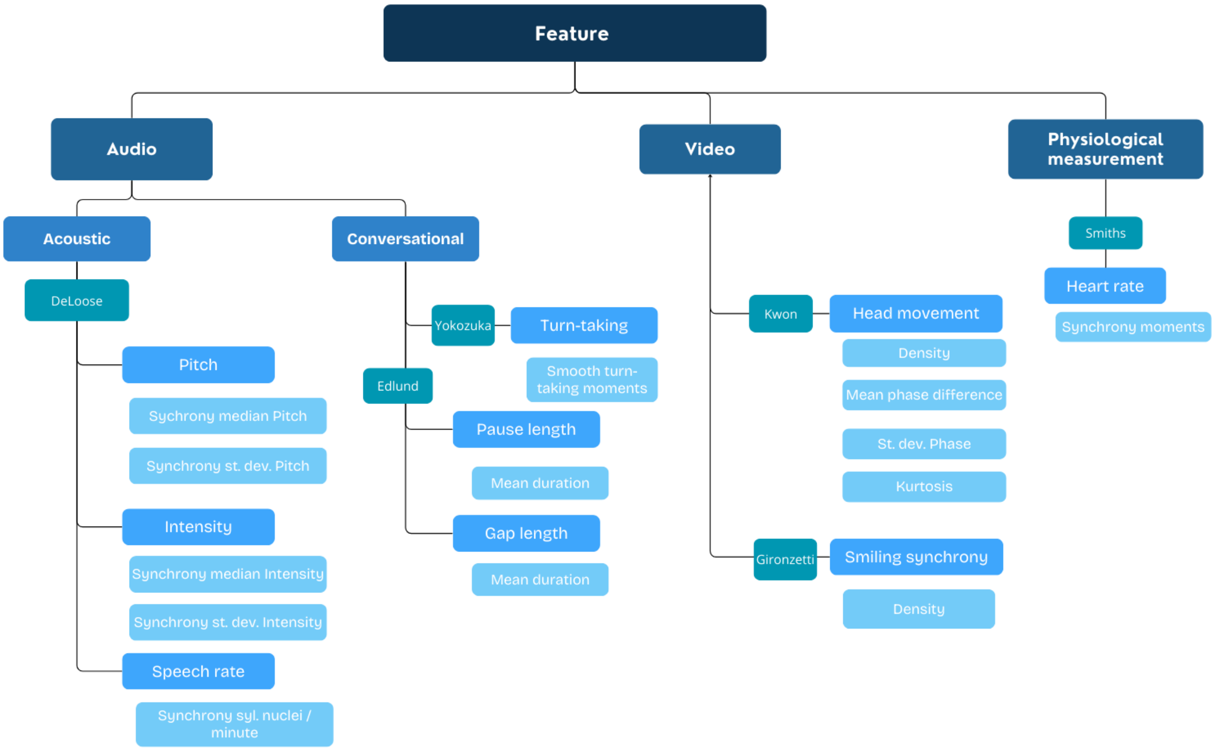

Figure 6.1: This figure presents a flowchart detailing the feature analysis approach used in this research. The features are categorized into three main domains: audio, video, and physiological measurements. The audio section of the flowchart is subdivided into acoustic and conversational features. Under the acoustic branch, pitch, intensity and speech rate are key indicators, with specific metrics such as median and standard deviation used to capture synchrony between participants. The conversational branch focuses on turn-taking and pause- and gap-lengths. The video section focuses on visual behaviours, particularly head movement and smiling synchrony. Lastly, the physiological measurement section includes heart rate data. By organizing these features into a structured analysis flowchart, the figure provides a clear representation of how different modalities (audio, video, and physiological) are integrated into the overall research framework.

behavior between the customer and agent. Finally, the physiological measurements section is focused on heart rate synchrony. Smiths' method is followed to analyze the heart rate data collected from both participants, and the correlation between the participants' heart rate is quantified.

The analysis focuses on the selected features across two distinct periods within each conversation: the normal period and the non-rapport period. Each conversation was divided into these two phases which allows to assess how the dynamics of the interaction shifted in response to changes in rapport. The normal period represents the typical flow of conversation, where rapport between the customer and agent is assumed to be intact, while the non-rapport period reflects moments of disconnect or lower engagement.

By examining the features—pitch, intensity, turn-taking, head movement synchrony, smile synchrony, and heart rate synchrony—during both periods, it is possible to determine whether significant differences exist between these two phases. This approach enables an assessment of how conversational and physiological features shift during moments of lower rapport, providing insights into which aspects of the interaction are most affected.

6.2 Data pre-processing

The data gathered from the studies are multi-modal, consisting of audio, video, and physiological measurements. Each data type undergoes specific pre-processing steps to ensure it is ready for analysis.

For the audio data, Deepgram Nova was used to transcribe the recordings into JSON files. Deepgram Nova, an automatic speech recognition (ASR) system, is known for its high accuracy and ability to handle various accents and noisy environments. The transcription process begins by feeding the raw audio recordings from customer-agent interactions into the Deepgram Nova ASR system. The system then produces a JSON file containing the transcribed text along with timestamps for each segment. This allows for precise alignment of spoken words with their corresponding times. Additionally, the JSON file includes speaker labels, enabling the differentiation between speakers. This structured format allows for easy segmentation and analysis of speech patterns, turn-taking, pause and gap lengths and conversational dynamics.

Parallel to the audio processing, OpenFace [41] is employed for the video data. OpenFace is an open-source tool developed for facial behavior analysis. OpenFace tracks 68 facial landmarks in each frame of the video, which are used to analyze expressions and movements. Furthermore, the tool is configured to estimate head orientation—such as yaw, pitch, and roll and to detect and quantify facial action units (AUs), which indicate smiling and help in assessing emotional expressions. The output generated by OpenFace is saved in a CSV file for each video, containing timestamps and detailed data on each feature extracted.

Physiological data, such as skin conductance and heart rate variability, were captured using Shimmer sensors. The Shimmer system provides detailed measurements of participants' physiological responses during the interactions. These measurements are stored in CSV files. The pre-processing of this data involves collecting raw physiological data in real-time during the studies using Shimmer sensors attached to the participants. Once the data is collected, it is cleaned to remove any artifacts or noise, ensuring accurate measurements of physiological states. The cleaned data is then exported into CSV files, with timestamps aligned to both the audio and video data. This allows for cross-modal analysis of physiological responses in relation to the participants' behavior during the interactions. After pre-processing, the audio, video, and physiological data streams are synchronized based on their timestamps. This synchronization ensures that each modality—speech, facial behavior, and physiological responses—can be analyzed in parallel, providing a holistic view of the interaction dynamics.

During the data collection phase of the studies, an issue with the Shimmer3 wristband used for measuring physiological data was encountered. Specifically, in some of the studies, the record button on the Shimmer3 device was not properly engaged, resulting in incomplete or missing data for those sessions. After reviewing the data, it was decided to exclude studies 1, 2, 3, and 7 from our analysis, as the missing physiological measurements would have compromised the accuracy and completeness of the analysis.

6.3 Audio

This section focuses on the examination of the audio data and its transcriptions. Various acoustic features, such as pitch, intensity, and speech rate, are analyzed to identify patterns and indicators of rapport or synchrony. Additionally, turn-taking behavior, as well as pause and gap lengths, are explored to assess conversational dynamics.

6.3.1 Pitch and Intensity

The pitch analysis process for the audio recordings was conducted following a method described by De Looze et al. [30]. The method looks at the temporal synchronization of the prosodic features between speakers to determine how closely aligned their speech is. First, the prosodic features of pitch and intensity are examined, followed by an analysis of speech rate.

The procedure was initiated by preparing two mono audio files derived from the original conversation recording—one from the actor and the other from the participant. A Deepgram transcription of the conversation was utilized to filter only the spoken words from these audio files, ensuring that non-speech elements were excluded from the analysis.

The filtered audio files were then loaded into Praat, a phonetic analysis software that facilitates detailed acoustic analysis. Praat was employed to extract the pitch values (in Hertz) at a fine temporal resolution, specifically every 0.01 seconds, providing 100 pitch values per second. This high-resolution data allows for a detailed examination of pitch variations over time. The next step in our analysis involved converting the pitch values from Hertz (Hz) to the octave scale. This conversion is essential for normalizing the pitch data, making it easier to compare across different speakers. The octave scale uses a logarithmic transformation, specifically log2 (Hertz), which aligns with human perception of pitch and facilitates comparison across different genders. To avoid possible pitch tracking errors, pitch floor and pitch ceiling (when creating a Pitch Object) were set to the values $p15 \cdot 0.83$ and $p65 \cdot 1.92$ respectively, where p15 and p65 denote the 15th and 65th percentile respectively.

Time-alignment
As discussed in Chapter 4.1 Prosodic measurements, when comparing prosodic features it is needed to ensure that the speakers are time-aligned. The hybrid utterance window method was used to break down the conversation into analyzable units. This method, sensitive to utterance boundaries, allowed for the creation of segments that are both temporally aligned and reflective of the conversational dynamics. This process enabled the analysis of pitch and intensity variations as a time series, offering a clear view of prosodic accommodation over the course of the conversation. Now, the median pitch and intensity values and their standard deviations were calculated for each segment, providing a statistical summary of the pitch characteristics. These values were then used to generate graphs that visually represent the prosodic dynamics between the speakers. Missing values in the data were interpolated to

maintain continuity in the analysis and ensure that the statistical calculations were based on complete datasets.

Prosodic synchrony measurement

To measure the synchrony in pitch and intensity development between the two interlocutors, the Pearson correlation coefficient was employed, which measures the linear dependency between two sets of observations, x and y, corresponding to the pitch values of the actor and the participant, respectively. The Pearson correlation coefficient, where values closer to 1 indicate strong positive synchrony, values closer to -1 indicate asynchrony and values around 0 suggest no linear relationship between the pitch developments of the two interlocutors.

$$P_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \sum_{i=1}^{N}(y_i - \bar{y})}{(N-1)s_x s_y}$$

As a Pearson correlation requires the same number of observations in both observation set x and y, a temporal window is utilized. The difference between the segments from the hybrid utterance-sensitive window method and the windows from the Pearson correlation method can be seen in figure 6.2. For the Pearson correlation method an 8-value windows size is used. This window size was determined after testing various window sizes; this size provided the most reliable and interpretable results in capturing synchrony. To maintain temporal precision and capture the dynamic nature of the conversation, a step size of four values was chosen, meaning the analysis window was advanced by four pitch values for each successive calculation. This overlap allowed for a more detailed and continuous assessment of the synchrony over time, avoiding potential gaps in the analysis that might occur with larger step sizes. Extensive experimentation with different window and step sizes confirmed that the selected configuration (window size 8 and step size 4) effectively balanced the need for temporal resolution and the stability of the correlation results. This setup was particularly effective in capturing the dynamic patterns of prosodic accommodation between the interlocutors. The significant values, with a p-value less than 0.05, are marked in red.

In figure 6.3 it can be seen how the median pitch and the standard deviation of the pitch behave over time in one study as an example. In figure 6.4 it can be seen how the median intensity and the standard deviation of the intensity behaves over time in one study also as an example. The graphs illustrate the dynamic nature of prosodic synchrony within the conversations. The median pitches of the actor and participant fluctuate over time, and the Pearson correlation indicates that the degree of synchrony between the speakers' pitches varies across different segments. Some segments show strong synchrony (high positive correlation), while others show weak or even inverse synchrony (negative correlation). This variation highlights the complex interaction patterns in human communication and suggests that synchrony is not constant but changes dynamically throughout the conversation.

In the results, the normal part of the study is compared with the moment of non-rapport. Four features are analyzed using Pearson correlation: median pitch, standard deviation of pitch, median intensity, and standard

Segment	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actor median pitch each segment	7,9	7,4	7,5	7,8	7,3	7,4	8,1	7,9	7,9	7,9	7,4	7,5	7,8	7,3	7,5	7,8	7,3	7,4	8,1	7,6
Participant median pitch each segment	6,7	6,6	6,7	6,5	6,3	6,5	6,7	6,4	6,3	6,4	6,3	6,5	6,7	6,3	6,5	6,7	6,4	6,4	6,7	6,3
	Window 1																			
					Window 2															
									Window 3											
													Window 4							

Figure 6.2: The difference between the segments from the Hybrid Utterance-Sensitive Window Method and the windows from the Windowed Pearson Correlation Method. Each segment has a value for the median pitch for the actor and the participant. The segments are aligned based on utterance boundaries and thus have varying lengths, while the fixed windows (blue boxes) for the windowed Pearson correlation coefficients, maintain a consistent size.

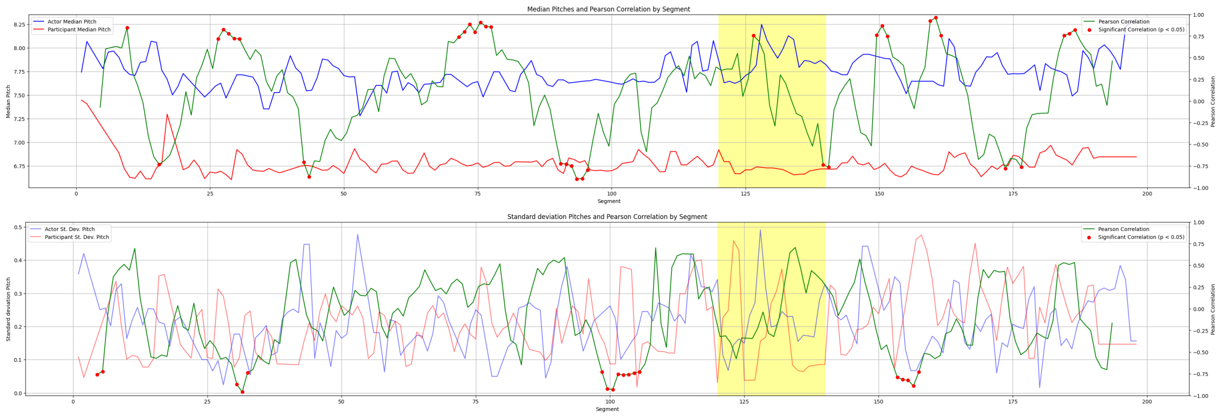

Figure 6.3: These two graphs illustrate the pitch synchrony between the actor and the participant throughout one conversation, using median pitch (top graph) and standard deviation of pitch (bottom graph). The purpose of these visualizations is to capture how pitch synchrony evolves throughout the interaction and to highlight moments where a significant correlation between the actor and participant's pitch occurs. In the top graph, the blue line represents the actor's median pitch, while the red line represents the participant's median pitch. The green line shows the Pearson correlation coefficient calculated using a windowed approach, which measures the degree of synchrony between the actor's and participant's median pitch over time. The significant correlation moments (where p-value < 0.05) are marked with red dots on the green line. The bottom graph shows how the standard deviation of pitch for both the actor and the participant fluctuates throughout the study. In both graphs, the yellow part shows the period of non-rapport.

deviation of intensity. From these Pearson correlation values, moments of high synchrony are defined as synchrony moments, characterized by a correlation higher than 0.7 and a p-value lower than 0.05. Additionally, moments of asynchrony are identified, where the correlation is lower than -0.7 with a p-value below 0.05. Moments that do not fall into either category are classified as maintenance moments. The next step is to determine whether the moments of non-rapport exhibit a different composition of synchrony, maintenance, and asynchrony moments compared to the normal period.

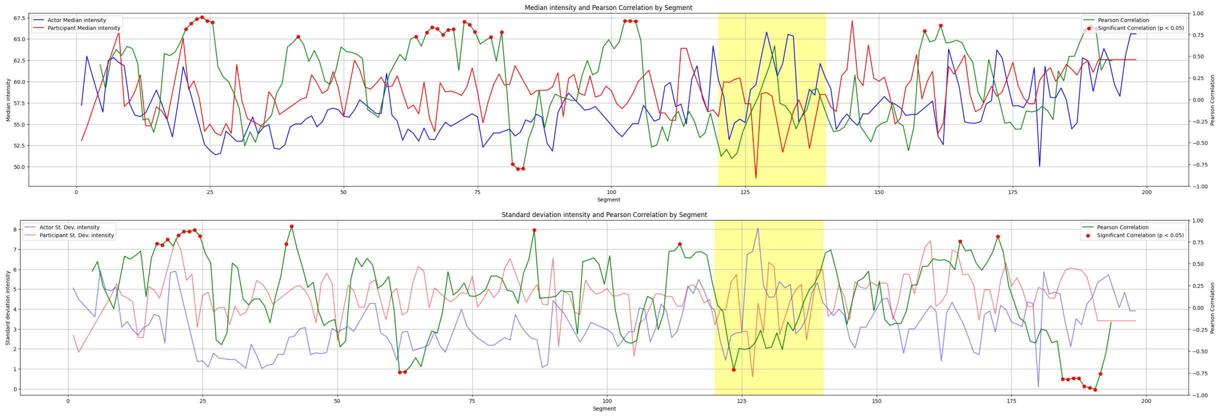

Figure 6.4: These graphs show the intensity synchrony between the actor and the participant during one conversation, using two measures: median intensity (top graph) and standard deviation of intensity (bottom graph). Again in both graphs, the blue line represents the actor, while the red line the participant. The green line represents the windowed Pearson correlation coefficient between the two intensity signals over time. Significant correlation points, where the p-value is less than 0.05, are highlighted with red dots on the green line. In both graphs, the yellow part shows the period of non-rapport. Together, these graphs provide valuable insights into how intensity synchrony unfolds over time.

6.3.2 Speech rate

Speech rate, measured as the number of spoken units (syllables or words) per unit of time, is a vital parameter for assessing conversational dynamics and fluency. For this analysis, the transcription data is used to compute syllable nuclei per minute, offering a broader view of speech tempo. To calculate the speech rate in syllable nuclei per minute, the Deepgram transcription file from each study is used. Relevant information, such as the text and timestamps for each word, is extracted. The speech rate is then calculated using the following formula:

$$Speech\ rate\ (syllables\ per\ minute) = \frac{Number\ of\ syllable\ nuclei}{Duration\ in\ seconds} x60$$

The number of syllable nuclei is identified by focusing on vowels, as syllables typically contain one vowel sound, which serves as the syllable nucleus [42]. In this approach, the transcription text is tokenized at the word level, meaning each word is treated as a token. A function is then used to iterate over each word and count the vowel occurrences, estimating the number of syllables in that word. Once the number of syllables for all words is calculated, the total duration of the speech segment is determined by subtracting the start time of the first word from the end time of the last word. With this information, the speech rate in syllable nuclei per minute is computed by dividing the total number of syllables by the duration in seconds and multiplying by 60 to convert to a per-minute rate. This method provides a statistical summary of speech rate based on syllable nuclei, with missing or incomplete data interpolated to maintain continuity in the analysis.

The next step in the analysis is calculating the synchrony in the speech rate of both interlocutors of the conversation. This is measured using a sliding window approach. The conversation is segmented into overlapping time windows of a specified size (e.g., 80 seconds), and each window is analyzed to detect correlations in speech rate between the two speakers. The process involves iterating through the conversation in steps (e.g., 40 seconds). For each window, the speech rates of both speakers are extracted and compared. If both speakers have data within the window, the Pearson correlation coefficient is calculated to assess the degree of synchrony in speech rates. If one or both speakers have insufficient data, the correlation is marked as missing. This produces a time series of correlation coefficients, which are then visualized to observe how speech rate synchrony fluctuates throughout the conversation. Figure 6.5 is an example of this visualization of one study.

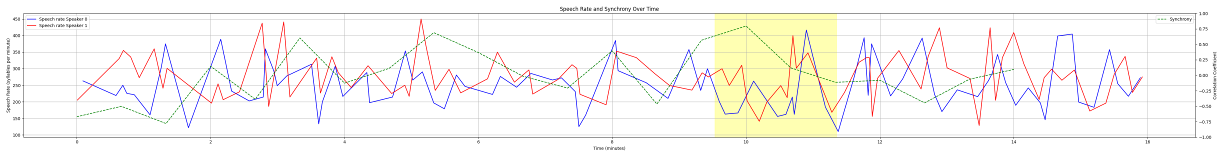

Figure 6.5: This graph visualizes the speech rate of two speakers (Speaker 0 and Speaker 1) over a 16-minute interaction during one study, as well as their speech rate synchrony over time. The speech rate for each speaker is measured in syllables per minute and plotted on the left y-axis, while the correlation coefficient (which measures synchrony between the speakers' speech rates) is plotted on the right y-axis. The yellow area in the graph shows the non-rapport moment during the conversation.

6.3.3 Turn-taking

Turn-taking is a fundamental aspect of conversational structure, reflecting the interactive nature of dialogue. By analyzing the transcription data from Deepgram, the points where speakers switch roles can be identified, marking the start and end of each turn. This analysis allows us to examine the flow and balance of the conversation, shedding light on the cooperative and dynamic aspects of the interaction.

Turn-taking in conversations was previously researched by Yokozuka et al. (2020) [8]. They state that conversations that have more rapport have more smooth turn-taking. A turn-taking moment is defined as the time interval between the end of one speaker's utterance and the start of the other speaker's response. Any intervals less than or equal to 1 second were classified as smooth turn-taking moments. The percentage of smooth turn-taking moments are then calculated in two distinct phases of the conversation: the normal period, characterized by typical interaction dynamics, and the non-rapport period, characterized by lower levels of rapport. The percentage of turn-taking moments that are smooth turn-taking moments during both periods was computed in each experimental session. This allowed us to compare conversational fluidity during normal and non-rapport periods.

6.3.4 Pause and gap lengths

This part of the analysis is based on the paper by Edlund et al. (2009) [25], which investigates the temporal aspects of conversational silences in spontaneous dialogues. Following a similar approach, the definitions of gaps and pauses are examined. A gap is a type of silence that occurs between two turns of different speakers. This silence represents a brief period where no one is speaking after one speaker finishes and before another begins. Gaps can indicate moments of thought or hesitation as the next speaker formulates their response or decides to take their turn. For example, in a multi-party conversation, if Speaker A finishes speaking and there is a silence before Speaker B starts, this silence is classified as a gap. A pause, on the other hand, occurs within the turn of a single speaker. It represents a brief silence where the same speaker continues their discourse after the silence. Pauses can be used for various rhetorical purposes, such as emphasizing a point, gathering thoughts, or marking transitions between different parts of the speech. For instance, if Speaker A is talking and briefly stops to think or emphasize before continuing, this silence is classified as a pause. Using the JSON files generated from Deepgram transcriptions, these intervals can be precisely measured. Analyzing silent periods provides valuable insights into the conversational flow and the level of synchrony between participants. The mean duration of pauses and gap lengths is examined in both the normal period and the non-rapport period of a conversation.

6.4 Video

The analysis of non-verbal cues captured in video recordings provides an understanding of the dynamics in the interactions. OpenFace is used to extract and analyze these non-verbal cues with high precision. This section focuses on two key aspects of non-verbal communication: smile synchrony and head movement.

6.4.1 Smile synchrony

Smile synchrony refers to the coordinated smiling behavior between two individuals during an interaction. It is a strong indicator of mutual engagement and positive rapport [16]. To analyze synchronized smiling behavior, the smiles of both the actor and the participant were measured during their interaction. This measurement followed the method described by Gironzetti et al. (2016), which defines a smile as a moment when Action Unit 6 (cheek raiser) and Action Unit 12 (lip corner puller) are both activated with high values. Using this criterion, instances of smiling were identified for both individuals throughout the conversation. The analysis focuses on the differences in synchronized smile behavior between the normal and non-rapport periods of the studies.

6.4.2 Head movements

Head movements are critical non-verbal cues that reflect attentiveness, agreement, and engagement during a conversation. OpenFace accurately tracks and quantifies head poses and movements, providing data on parameters such as pitch, yaw, and roll. For the head movement analysis, the method proposed by Kwon et al. (2023)[38] is followed, which analyzed head motion synchrony in conversation using accelerometers to capture participants' head movements in different dimensions. Instead of using specialized devices for measuring head movements in yaw, pitch, and roll, OpenFace is utilized, as it provides similar measurements. By tracking facial landmarks and outputting head pose data, OpenFace enables the analysis of head movements in three dimensions: yaw, pitch, and roll. This approach not only replicates the core principles of Kwon et al.'s method but also expands its applicability by enabling video-based head motion analysis.

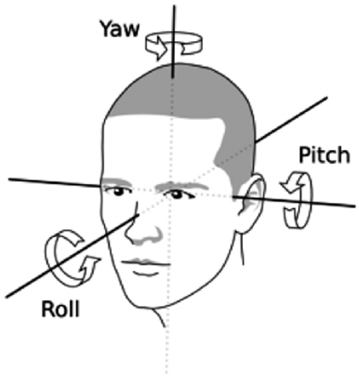

Figure 6.6: The yaw, pitch, and roll attributes of a face orientation [43].

The three dimensions of head movement — yaw, pitch, and roll — represent the different types of rotational movement that occur during human interaction. Yaw refers to the side-to-side movement of the head, where a person turns their head left or right. Pitch measures the up-and-down nodding motion, such as when a person nods in agreement. Roll, on the other hand, represents the tilting of the head from one shoulder to the other, a motion that can be seen when a person tilts their head in curiosity or confusion. These three axes allow for a detailed and nuanced understanding of how participants move their heads during conversations. These dimensions can be seen in figure 6.6.

The data provided measurements for yaw (left-right rotation), pitch (up-down rotation), and roll (tilt) for each participant in the conversation. The output included timestamped data for each of these angles, as well as a number of other facial action units. Before conducting the analysis, it was important to clean the data. This involved removing any periods where head movement data was missing or where the participants' faces were not fully visible to the camera, as these could introduce noise into the synchrony calculations. Additionally, the data was smoothed using a moving average to reduce minor fluctuations in the measurements that were not representative of meaningful head movements.

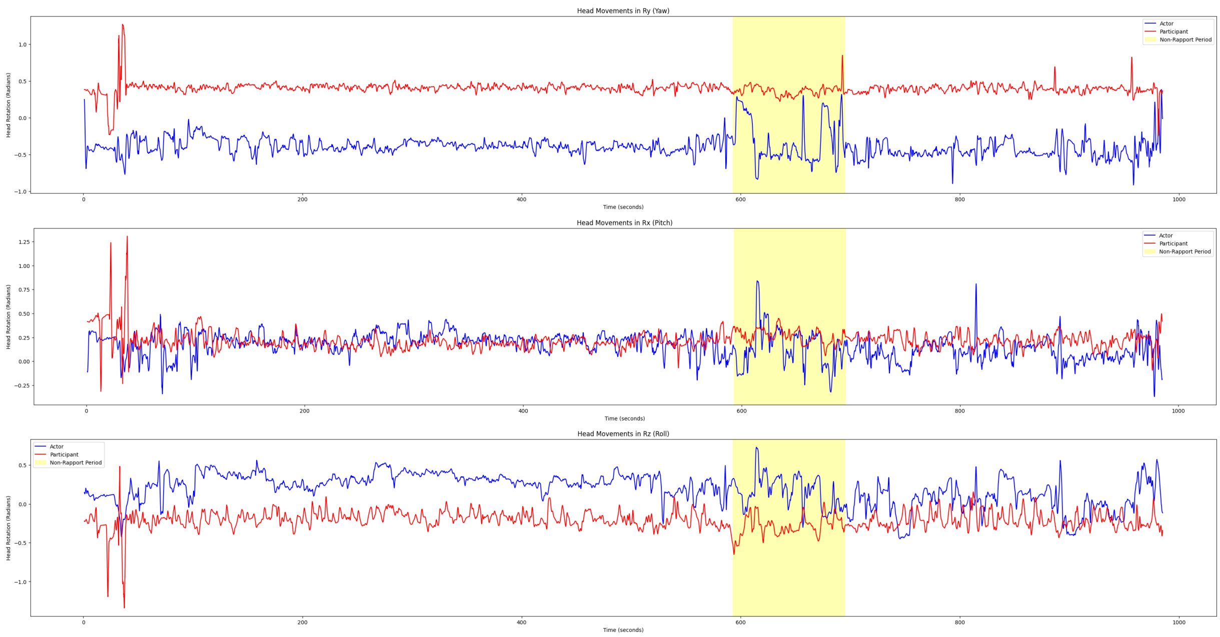

Figure 6.7: This figure illustrates the head movements of both the actor and the participant across three dimensions: yaw (horizontal rotation), pitch (vertical tilt), and roll (side tilt). The blue line represents the actor's head movements, while the red line represents the participant's head movements. The yellow-shaded area indicates the period of non-rapport in the conversation, where the rapport between the actor and participant was intentionally disrupted. In the normal period, there are instances where the actor and participant's head movements appear relatively synchronized, especially in the roll and pitch dimensions. During the non-rapport period, head movement variability increases, particularly in yaw and roll.

Using head pose data, the yaw, pitch, and roll values are extracted and analyzed for each participant during the conversation. An example of these measurements during one study can be seen in figure 6.7. For each study, synchrony metrics such as the density of synchrony events, the mean phase difference, the standard deviation, and the kurtosis for both the normal and non-rapport periods are calculated. These metrics help assess whether there were significant changes in head movement synchrony when rapport was disrupted.

- Density refers to the frequency of synchrony events, where both participants move their heads in coordination during the conversation. In my analysis, a synchrony event is defined as a moment where the head movements of both participants are either simultaneous or occur within a small time lag. The code measures this by detecting peaks in the head movement data for each participant. These peaks represent significant head movements, such as a nod or head turn. After detecting the peaks, the phase differences between corresponding peaks in the two participants' data are calculated. If the phase difference between two peaks is within a predefined threshold (1 second), it is considered a synchrony event. The density is then calculated as the total number of synchrony events that occur within a given period, normalized by the duration of that period (in minutes).
- Mean phase difference represents the average time lag between the head movements of the two participants. This metric indicates whether one person consistently leads or follows in head movements. A positive mean phase difference implies that one participant generally moves their head first, while the other follows, whereas a negative phase difference indicates the opposite. Examining the mean phase difference across the normal and non-rapport periods allows for an exploration of whether rapport influences the leadership dynamics in head movement.
- Standard deviation of phase differences measures the variability in head movement synchrony. A low standard deviation indicates that the participants are consistently in sync, while a higher standard deviation suggests more erratic or inconsistent synchronization. This metric is important because it provides insight into the stability of the interaction. If the standard deviation increases during the non-rapport period, it may suggest that the participants are struggling to maintain consistent non-verbal coordination.
- Kurtosis measures the sharpness or flatness of the distribution of phase differences. A high kurtosis value indicates that the phase differences are tightly clustered around the mean, meaning that the participants are moving their heads in a highly synchronized and predictable manner. A low or negative kurtosis suggests a flatter distribution, where synchrony is less precise, and the participants' movements are more spread out over time. Analyzing kurtosis allows me to explore whether the sharpness of the head movement synchrony changes between the normal and non-rapport periods, offering further evidence of whether rapport affects the precision of interaction.

6.5 Physiological measurements

Lastly, the physiological measurements section shows how the heart rate measurements are being analyzed.

6.5.1 Heart rate

Heart rate synchrony between participants is measured using the method established by Smits et al. (2020) [23], which utilizes PPG (photoplethysmography) measurements to track the participants' heart rates. The raw PPG data is first processed to calculate RR-intervals. These RR-intervals serve as the foundation for heart rate variability (HRV) analysis, which helps assess fluctuations in heart rate over time.

In the first stage of heart rate synchrony analysis, PPG measurements from both the actor and from the participants during the conversation were collected. PPG data captures the changes in blood volume in the body and is commonly used to estimate heart rate. From this raw PPG data, RR-intervals were calculated, which are the time intervals between consecutive heartbeats. RR-intervals are crucial for understanding the variability in heart rate, which reflects the autonomic nervous system's response to emotional and cognitive stimuli during the interaction.

Once the RR-intervals were calculated, the HRV-analysis Python program was used to clean the data. This step is necessary to remove noise and correct any irregularities caused by movement artifacts or signal loss during data collection. The HRV-analysis package offers several techniques for preprocessing RR-interval data, including artifact correction and filtering. Ensuring clean data allows for a more accurate analysis of heart rate variability and its correlation between the two participants. Figure 6.8 provides an illustration of the RR-interval.

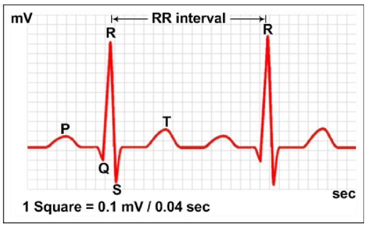

Figure 6.8: Illustration of a heartbeat with the RR-interval highlighted [44] The figure shows an electrocardiogram (ECG) waveform, which represents the electrical activity of the heart during a single cardiac cycle. The waveform is composed of several key components, each labelled with a letter (P, Q, R, S, and T) that corresponds to a specific part of the heartbeat. The RR interval, highlighted in the figure, represents the time between two consecutive R waves—the highest peaks in the ECG.

After the data was cleaned, the heart rate variability (HRV) graphs for both the customer and the agent were generated. These graphs plot the fluctuations in heart rate over time, showing the variability in the RR intervals. The HRV graphs provide a visual representation of how the participants' heart rates change throughout the conversation, and allow us to observe patterns of physiological alignment or divergence between them.

To quantify the synchrony between participants' heart rates, the windowed cross-correlation technique was applied, following the approach used for audio features described by De Looze et al. (2014)[30]. This method involves dividing the HRV time series data into smaller time windows and calculating the correlation within each window. By using this approach, it is possible to assess how participants' heart rates fluctuate in synchrony at different points in the conversation. Higher cross-correlation values indicate greater synchrony, which can be associated with periods of rapport and mutual engagement, while lower values suggest physiological disconnection or reduced engagement. A window size of 12 heart rate values and a step size of 6 were used, based on the method outlined by Smits et al.(2020).

6.6 Classifying interactions

After a detailed examination of how each feature behaves in the studies, the next step is to determine whether moments of non-rapport can be classified and which features contribute most to this classification. To achieve this, two machine learning models will be tested: Random Forest and Logistic Regression.

For this classification task, a dataset will be generated from the features extracted during both the normal and non-rapport periods of the conversations. The dataset consists of 92 samples of 30 seconds from the conversation. A consistent 30-second unit was selected for this task to ensure a standardized, balanced dataset that represents the interaction period uniformly across both normal and non-rapport phases. This length provides sufficient data to calculate synchrony features while maintaining alignment with the temporal structure of the interaction. To ensure the dataset is balanced and representative of both interaction periods, an equal number of time windows from both the normal and non-rapport phases will be sampled. Each sample in the dataset will have labels indicating whether the interaction belongs to the normal or non-rapport period. The following nine features are in each sample:

1. Median pitch synchrony: the correlation of the median pitch values of both speakers, calculated as the Pearson correlation of pitch values over the 30-second period.
2. Standard deviation of pitch synchrony: the correlation between the standard deviation of pitch for both speakers. This measurement reflects how pitch fluctuations align between speakers.
3. Median intensity synchrony: the correlation between the median loudness of both speakers, measuring alignment in their speech intensity.
4. Standard deviation of intensity synchrony: the correlation between the variability in loudness for both speakers. Measuring shared patterns in loudness fluctuations.
5. Speech rate synchrony: the correlation between the speech rates of both speakers, reflecting alignment in pacing during the conversation.
6. Smooth turn-taking moments: normalized values of the amount of smooth turn-taking moments per minute, resembling the smoothness in the conversation.
7. Smile synchrony: normalized values of the amount of synchronized smiles per minute, capturing alignment in smiling behaviour.
8. Head movement synchrony: normalized values of the amount of synchronized head movements per minute, capturing alignment in gestures.
9. Heart rate synchrony: the correlation between the heart rate variability of both speakers.

To understand the influence of non-verbal features (smile synchrony and head movement synchrony), an additional experiment will be conducted in which the models are trained on a reduced dataset that includes only audio features (pitch, intensity, and speech rate). By comparing the performance of models trained with the full set of features against those trained with only audio features, the impact of non-verbal cues on rapport detection can be determined. Additionally, this comparison assesses how well models perform when limited to features that can be extracted from phone conversations, where video data is unavailable.

The decision to use both a Random Forest model and a Logistic Regression model stems from their differing strengths in classification tasks. Random Forest, an ensemble learning method, is well-suited for handling non-linear relationships between features and can capture complex interactions across the dataset[45]. It is particularly robust to overfitting, especially with high-dimensional data, and provides feature importance metrics, which can offer insights into which features contribute most to rapport classification.

On the other hand, Logistic Regression is a simpler and more interpretable model. It assumes a linear relationship between the independent variables and the outcome, making it useful for understanding the impact of individual features on the predicted rapport classification[46]. Comparing the performance of these two models allows for an assessment of both the complexity of interaction dynamics and the relative contribution of the features. The implementation of the Random Forest and Logistic Regression models was carried out using the scikit-learn library in Python [47].

The performance of both the Random Forest and Logistic Regression models will be evaluated using standard classification metrics, including accuracy, precision, recall, and the F1 score. Additionally, cross-validation will be performed to ensure that the results are not overly dependent on any specific subset of the data. Comparing the performance of both models across these metrics will help determine which model better captures the underlying dynamics of rapport in customer-agent conversations.

Finally, the importance of individual features for both models will be analyzed. For Random Forest, feature importance will be calculated directly from the model, while for Logistic Regression, the model's coefficients will indicate the strength of each feature's contribution. This analysis will provide valuable insights into which aspects of the conversation play the most critical role in predicting moments of high and low rapport.

6.7 Summary

In this chapter, the methodology used to analyze data collected from interactions is outlined with a focus on identifying features that indicate rapport. The approach includes audio, video, and physiological measurements, which are going to be processed to extract relevant features. It is explained how the features are planned to be analyzed in order to answer the question: What measurable features show significant differences between moments of rapport and non-rapport in conversations? Additionally, it explains how machine learning models (Random Forest and Logistic Regression) are employed to classify moments of rapport and non-rapport. To evaluate the influence of non-verbal features, the models are also trained on a reduced dataset, focusing solely on audio features. Thus answer the question: How well does a model perform on classifying rapport in conversations with and without non-verbal features? The methodology outlined in this chapter highlights a systematic approach to understanding rapport. By combining feature extraction with machine learning classification, this framework aims to analyze the dynamics of rapport in conversations. Additionally, evaluating models with and without non-verbal features will focus on the role these cues play in classifying rapport, particularly in phone based communication where visual cues are absent. The findings will be explored in the next chapter, where the results are presented with their implications.

Results 7

This chapter presents the findings from the audio, video and physiological measurements analyses conducted to understand and measure the level of rapport in conversations. The primary goal of this analysis is to identify measurable differences between moments of rapport and non-rapport, as well as evaluate the effectiveness of machine learning models in distinguishing these interaction states. To ground the findings in participants' experiences, the analysis begins with an examination of the questionnaire results, which capture subjective perceptions of rapport and non-rapport moments. Following the questionnaire analysis, statistical comparisons of acoustic, conversational, and non-verbal features are conducted to highlight differences between normal and non-rapport periods. Next, the performance of machine learning models, Random Forest and Logistic Regression, is evaluated in classifying interaction moments as either normal or non-rapport. This includes a look into the feature importance to determine which attributes contribute most significantly to the models' accuracy. Through this approach, the findings in this chapter aim to validate the methodological framework presented earlier and provide critical insights into the measurable aspects of rapport.

7.1 Perceived rapport and non-rapport: questionnaire results

The primary aim of our questionnaire is to investigate whether the experiment successfully captured distinct moments of rapport and non-rapport between participants and the actor during the interactions. To evaluate this, a questionnaire was designed to capture participants' subjective experiences of the interaction. The questionnaire aimed to assess whether participants felt connected to the actor, experienced moments of mutual understanding, and noticed shifts in the interaction dynamics. This section presents the results from the questionnaire analysis, focusing on whether participants perceived a clear distinction between the normal and non-rapport phase of the conversation. The questionnaire consisted of nine questions answered with a Likert scale and five open questions. To showcase the questions that were answered with a Likert scale, boxplot were made of all the answers given. These can be seen in figure 7.1.

The boxplots reveal insights into participants' perceptions of the interaction. High scores in connectedness and understanding (top-left and top-right plots) suggest the normal phase successfully fostered moments of rapport. In contrast, the variability in perceived distance (top-middle plot) highlights differences in how participants experienced the conversations and especially the non-rapport phase. Mixed responses in curt behavior (middle-left plot) and caring (middle-center plot) further underscore the actor's ability to modulate rapport and tension dynamically.

To further confirm that people felt the normal period as a moment where rapport was created, the first two open questions asked were: Did you

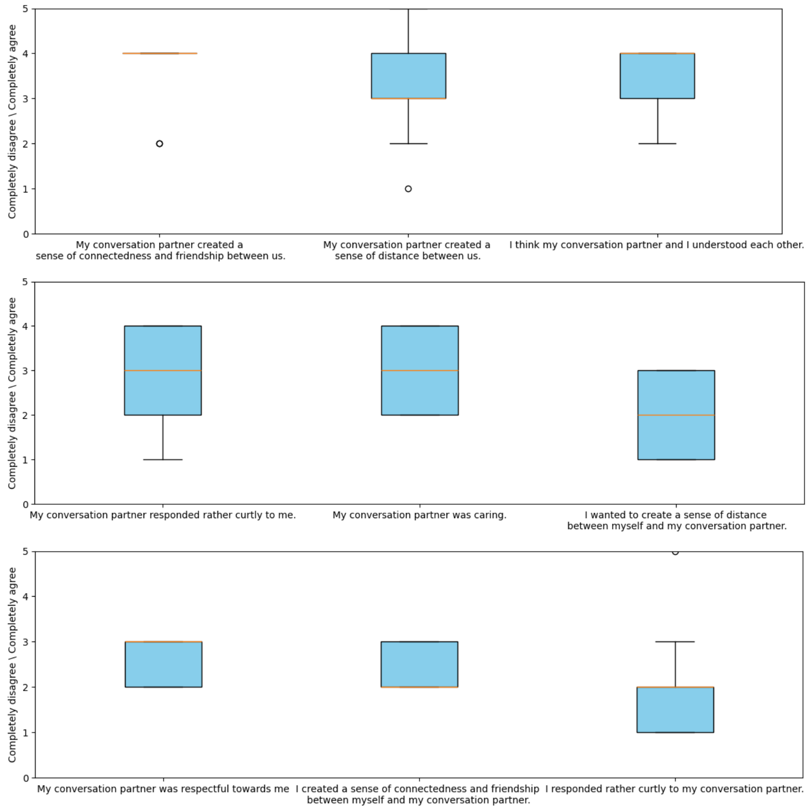

Figure 7.1: Boxplots that showcase how all the questions answered with a 5 point Likert scale(completely disagree - completely agree)

have a good connection with the actor during the conversation and "How did you feel when talking about your hobby?" From the nine questionnaires, only one stated that there was a clear distance between him/her and the actor during the entire conversation. Eight reacted that there was a good connection during parts of the conversation and one person even mentioned that he/she was surprised about how devoted he/she felt during in the moments the actor acted enthusiastic about his/her hobby. Two persons answered that they felt the conversation was “forced” or “preconceived”.

To confirm that participants experienced the non-rapport period, they were asked: “were there moments during the conversation you felt insecure and tensions? If yes, could you describe what you felt during these moments.” Every answer mentioned the period of non-rapport and the moment is described as “a period where a distance was created” or “the moment felt like she was respectless”. Four people even mentioned the moment felt like a light personal attack or that it felt like they had to defend themselves.

7.2 Statistical comparison

Different statistical methods were used to analyze the types of features examined in this study. For the acoustic features such as median pitch and standard deviation of intensity, the analysis focused on assessing the likelihood of specific events occurring during normal vs. non-rapport periods. Given the categorical nature of these events, odds ratios were used to quantify how these features increased or decreased the likelihood of rapport. For the conversational features like the amount of smooth turn-taking moments and the mean duration of pauses and gaps, the goal was to determine whether there were significant change in the average level of these behaviours between the two conditions. Since these are continuous measurements taken from matched conversations, paired t-tests were used to compare the means between the normal and non-rapport periods. This chapter first examines the acoustic and conversational features before analyzing the non-verbal cues.

7.2.1 Acoustic and conversational features

When analyzing the audio features, a distinction can be made between acoustic and conversational features. Pitch, intensity, and speech rate are categorized as acoustic features, while turn-taking, pause length, and gap length are considered conversational features. The acoustic features are analyzed using a windowed Pearson correlation. The next step is to determine whether the periods of non-rapport differ statistically from the normal period.

The four features from which the windowed Pearson correlation is calculated. The median pitch, the standard deviation of the pitch, the median intensity and the standard deviation of the intensity. From this Pearson correlation, the moments of high synchrony are defined as synchrony moments. These are moments with a correlation higher than 0.7 and a p-value lower than 0.05. Next to that, there are moments of asynchrony. These are moments with a correlation lower than -0.7 and a p-value lower than 0.05. The moments during the experiment that are neither moments of synchrony nor asynchrony are classified as maintenance moments. The next step is to see if the moments of non-rapport show a different composition of synchrony, maintenance and asynchrony moments than the normal period. This compositions can be see in table 7.1, table 7.2, table 7.3 and table 7.4.

In table 7.1 the synchrony in the median pitch can be seen across all the studies. Maintenance moments make up a significant portion of the interactions. Asynchrony moments are points where the actor's and participant's pitches diverge significantly, indicating a lack of synchrony. Interestingly, asynchrony moments remain relatively low across both periods, indicating that while synchrony may decrease during non-rapport, direct misalignment or divergence in pitch is still not very frequent. Instead, the interaction seems to shift more toward maintenance or moments of weaker synchrony rather than outright asynchrony.

Table 7.1: Comparison of synchrony, maintenance and asynchrony moments in the median pitch between normal period and non-rapport period. The data is organized by experiment (Exp 4, Exp 5, etc.), and each row shows the absolute counts and percentages of these moments during both periods. The normal period (left side of the table) and the Non-rapport period (right side of the table). Both sides show the amount of synchrony, maintenance and asynchrony moments. Synchrony moments represent the moments where the actor and participant's median pitch were significantly aligned during the period. Maintenance moments refer to periods where there is no change in the synchrony, meaning the actors and participants maintained a steady alignment in their median pitch.

Median Pitch normal period compared to non-rapport period						
	Synchrony moments normal period	Maintenance moments normal period	Asynchrony moments normal period	Synchrony moments non-rapport period	Maintenance moments non-rapport period	Asynchrony moments non-rapport period
Exp 4	22	136	12	1	18	1
Exp 4 percentage	12,94	80	7,06	5	90	5
Exp 5	17	141	4	0	27	0
Exp 5 percentage	10,49	87,04	2,47	0	100	0
Exp 6	22	161	7	8	24	0
Exp 6 percentage	11,58	84,74	3,68	25	75	0
Exp 8	21	139	8	2	31	0
Exp 8 percentage	12,50	82,74	4,76	6,06	93,94	0
Exp 9	10	152	0	5	18	1
Exp 9 percentage	6,17	93,83	0,00	20,83	75,00	4,17
Exp 10	15	131	0	6	24	3
Exp 10 percentage	10,27	89,73	0,00	18,18	72,73	9,09
Total	107	860	31	22	142	5
Total percentage	10,72	86,17	3,11	13,92	84,02	2,96

In table 7.2 the synchrony in the standard deviation of the pitch during all the studies can be seen. Maintenance moments continue to dominate the interaction, making up a large percentage across all experiments. This suggests that in both phases of the interaction, the variability in pitch between the actor and participant remains relatively stable most of the time. Even during the non-rapport period, participants and the actor tend to maintain a steady alignment in pitch variability, possibly indicating an underlying baseline synchronization even in less engaged moments. The synchrony moments decrease dramatically during the non-rapport period. While 50 synchrony moments occur during the normal period, only 1 moment of synchrony is observed in the non-rapport period.

Table 7.2: This table shows the comparison of synchrony, maintenance, and asynchrony moments in the standard deviation of the pitch between the normal period and the non-rapport period. The data is organized by experiment (Exp 4, Exp 5, etc.), and each row shows the absolute counts and percentages of these moments during both periods. The normal period is shown on the left side of the table, and the non-rapport period is shown on the right. Both sides display the number of synchrony, maintenance, and asynchrony moments in the standard deviation of pitch.

Standard deviation of the Pitch normal period compared to non-rapport period						
	Synchrony moments normal period	Maintanance moments normal period	Asynchrony moments normal period	Synchrony moments non-rapport period	Maintanance moments non-rapport period	Asynchrony moments non-rapport period
Exp 4	0	152	18	0	20	0
Percentage	0,00	89,41	10,59	0	100	0
Exp 5	4	149	9	0	24	3
Percentage	2,47	91,98	5,56	0	88,89	11,11
Exp 6	25	157	8	0	27	5
Percentage	13,16	82,63	4,21	0	84,375	15,625
Exp 8	4	157	7	0	30	3
Percentage	2,38	93,45	4,17	0,00	90,91	9,09
Exp 9	12	150	0	0	24	0
Percentage	7,41	92,59	0,00	0,00	100,00	0
Exp 10	5	127	14	1	32	0
Percentage	3,42	86,99	9,59	3,03	96,97	0
Total	50	892	56	1	157	11
Percentage	5,01	89,38	5,61	0,59	92,90	6,51

Table 7.3: This table provides a comparison of synchrony, maintenance, and asynchrony moments in median intensity for both the normal period and the non-rapport period across several experiments. The table breaks down the counts and percentages of these moments during both phases, allowing us to understand how vocal intensity alignment changes when rapport is disrupted.

Median intensity in the normal period compared to non-rapport period						
	Synchrony moments normal period	Maintanance moments normal period	Asynchrony moments normal period	Synchrony moments non-rapport period	Maintanance moments non-rapport period	Asynchrony moments non-rapport period
Exp 4	27	140	3	0	20	0
Percentage	15,88	82,35	1,76	0	100	0
Exp 5	13	147	2	8	19	0
Percentage	8,02	90,74	1,23	29,63	70,37	0
Exp 6	44	144	2	5	27	0
Percentage	23,16	75,79	1,05	15,625	84,375	0
Exp 8	30	125	13	1	32	0
Percentage	17,86	74,40	7,74	3,03	96,97	0
Exp 9	15	144	3	0	24	0
Percentage	9,26	88,89	1,85	0	100	0
Exp 10	18	120	8	0	33	0
Percentage	12,33	82,19	5,48	0	100	0
Total	147	820	31	14	155	0
Percentage	14,73	82,16	3,11	8,28	91,72	0

In table 7.3 the synchrony in the median intensity can be seen during all the studies. During the normal period, synchrony occurs more frequently, while maintenance moments remain the dominant state, and asynchrony moments are rare. Experiment 5 has a high percentage of synchrony moments which perhaps shifts the average in total. Maintenance moments make up the majority of the interaction. Asynchrony moments are absent in the non-rapport period, with 0 occurrences across all experiments. This is interesting because, while synchrony moments decrease, there is no significant increase in outright vocal intensity misalignment, suggesting that while the participants and actors may not actively synchronize, they do not actively diverge either. However it is clear that during the normal period, synchrony occurs more frequently.

Table 7.4: This table provides a comparison of the synchrony, maintenance, and asynchrony moments in the standard deviation of intensity during both the normal period and non-rapport period across several experiments.

Standard deviation of the intensity in the normal period compared to non-rapport period						
	Synchrony moments normal period	Maintanance moments normal period	Asynchrony moments normal period	Synchrony moments non-rapport period	Maintanance moments non-rapport period	Asynchrony moments non-rapport period
Exp 4	14	146	10	0	19	1
Percentage	8,24	85,88	5,88	0,00	95,00	5,00
Exp 5	8	147	7	3	24	0
Percentage	4,94	90,74	4,32	11,11	88,89	0,00
Exp 6	21	160	9	2	27	3
Percentage	11,05	84,21	4,74	6,25	84,38	9,38
Exp 8	15	141	12	1	32	0
Percentage	8,93	83,93	7,14	3,03	96,97	0,00
Exp 9	5	145	12	0	23,00	1
Percentage	3,09	89,51	7,41	0,00	95,83	4,17
Exp 10	14	127	5	0	31	2
Percentage	9,59	86,99	3,42	0,00	93,94	6,06
Total	77	866	55	6	156	7
Percentage	7,72	86,77	5,51	3,55	92,31	4,14

In table 7.4 the synchrony in the standard deviation of the intensity can be seen during all the studies. Synchrony moments decrease during the non-rapport period, dropping to only 6 moments in total (representing 3.55% of the moments). This suggests that the alignment of intensity variability between the actor and participant becomes more difficult to achieve when rapport is disrupted. In both the normal and non-rapport periods, maintenance moments dominate the interaction, suggesting that most of the time, vocal intensity variability remains steady rather than shifting frequently between synchrony and asynchrony. The next step is to compare if the proportions are significantly different in the non-rapport period compared to the normal period for all the acoustic features.

This is done by looking at the odds ratio. This odds ratio compares the odds of an event occurring in the normal period to the odds of it occurring in the non-rapport period. The odds ratio tells how many times more or less likely the event is to occur in one period compared to the other. For every event the n is the total number of events (synchrony, maintenance, and asynchrony moments) across both periods (normal and non-rapport). The results can be seen in table 7.14.

Table 7.5: The odds ratios and p-values for different acoustic features. The acoustic features are median pitch, standard deviation of pitch, median intensity, and standard deviation of intensity during normal periods and non-rapport periods.

Comparison acoustic features			
Feature	Moment	Odds ratio	p-value
Median pitch			
	Synchrony	0,82	0,433
	Maintenance	1,03	0,854
	Asynchrony	1,05	1.0
Standard deviation pitch			
	Synchrony	8,46	0,006
	Maintenance	0,96	0,764
	Asynchrony	0,86	0,598
Median intensity			
	Synchrony	1,78	0,050
	Maintenance	0,89	0,364
	Asynchrony	inf	0,0157
Standard deviation intensity			
	Synchrony	2,17	0,071
	Maintenance	0,94	0,630
	Asynchrony	1,33	0,579

It can seen in the table that the standard deviation of pitch shows a significant increase during synchrony moments, suggesting more dynamic variation in pitch when participants are aligned. Median intensity tends to increase during synchrony and shows a significant divergence during asynchrony moments, indicating clear shifts in intensity during periods of alignment and misalignment. These two features thus show a possibility of recognizing moments of asynchrony. Next to those two features, the feature median pitch and standard deviation of intensity do not show significant differences between synchrony, maintenance, or asynchrony moments, indicating that they are not as sensitive to changes in alignment during conversation.

The last acoustic feature measured was the speech rate, the speech rate for both the actor and participant was calculated during conversations, with distinct measurements taken for normal interaction periods and non-rapport periods. The speech rate, measured in syllable nuclei per minute, was averaged within each of the segments for both participants. For each segment (normal and non-rapport), the mean speech rate was calculated separately for both the actor and the participant, yielding the values presented in the table 7.6.

The overall mean speech rate during the normal period (across all experiments) is 283 words per minute. This increases slightly to 297.5 words per minute during the non-rapport period. This increase in speech rate during the non-rapport period suggests that, despite the rapport breakdown, both speakers tend to accelerate their speech. This could reflect increased tension, emotional responses, or an attempt to maintain control of the conversation during the non-rapport phase. The actor's speech rate remains relatively stable across both periods, possibly indicating that the actor maintains control of the interaction, regardless of the rapport dynamics. The participant's speech rate shows more variation, with some participants increasing their speech rate more significantly

Table 7.6: This table provides the mean speech rate (in syllable nuclei per minute) for both the actor and the participant during the normal period and non-rapport period across six experiments. The table breaks down the overall speech rate during both phases of the interaction, as well as the speech rates of the actor and participant individually. The average speech rates for each condition are also included.

	Mean speech rate normal period	Mean speech rate non-rapport period	Mean speech actor rate normal period	Mean speech rate actor non-rapport period	Mean speech rate participant normal period	Mean speech rate participant non-rapport period
Experiment 4	267	273	282	283	252	264
Experiment 5	309	321	320	320	296	322
Experiment 6	271	271	285	305	256	237
Experiment 8	279	268	292	263	266	271
Experiment 9	278	296	278	319	278	272
Experiment 10	294	296	295	302	293	289
Average	283	297,5	292	298,67	273,50	275,83

than others during the non-rapport phase. This suggests that individual participants may have different reactions to the rapport-breaking intervention. The table allows for a direct comparison between the mean speech rates during normal periods and non-rapport periods. For instance, in Experiment 4, the mean speech rate during the normal period is slightly lower than during the non-rapport period for both the actor and the participant. Across the different experiments, there is variability in how the speech rates change between normal and non-rapport periods. In some experiments, like Experiment 6, the participant's speech rate decreases significantly during the non-rapport period, while in others, such as Experiment 5, both actors' and participants' speech rates increase slightly. The data suggests that the speech rate can either increase or decrease during non-rapport periods, depending on the dynamics of the interaction. This variability could be related to how individuals adjust their speech patterns when rapport is disrupted, which may include speaking faster or slower.

To assess whether the difference in speech rates between the two periods was statistically significant, a paired t-test was chosen. This method is appropriate because the speech rates for each experiment were measured under two conditions: normal and non-rapport. The test outputs a t-statistic and a p-value (*n=6, representing 6 experiments*). The t-statistic indicates the size of the difference relative to the variation in the sample data, while the p-value determines whether the observed difference is statistically significant. The calculated t-statistic is -1.234 with a p-value of: 0.045. This result would indicate that, on average, there is a significant difference in the speech rates between the two periods.

In addition to acoustic features, conversational audio features are also analyzed. One key aspect is turn-taking, where the number of smooth turn-taking instances is measured and compared with instances that are not smooth. As told before, a turn-taking is defined as smooth when it takes less than 1 second. The percentage of turn-takings that were smooth turn-taking moments during the normal and non-rapport periods in each study can be seen in table 7.7. During the non-rapport period, the average percentage of smooth turn-taking decreases to 12.03%, indicating that fewer turn-taking moments occurred smoothly when rapport was disrupted. Across all experiments, the percentage of smooth turn-taking moments is consistently lower in the non-rapport period than in the normal period

	% of turn-taking moments shorter than 1 sec during normal period	% of turn-taking moments shorter than 1 sec during non-rapport period
Exp 4	8,97	10
Exp 5	26,17	16,22
Exp 6	15,79	16,67
Exp 8	9,28	7,69
Exp 9	16,86	10
Exp 10	11,71	0
Total	15,60	12,03

Table 7.7: This table compares the percentage of turn-taking moments that occurred smoothly (with less than 1 second between turns) during both the normal period and the non-rapport period across the studies. .

In addition to turn-taking, pause lengths were measured throughout all conversations. These can be seen in table 7.7. The mean pause duration during the normal period across all experiments is 0.57 seconds. During the non-rapport period, the average pause duration slightly decreases to 0.56 seconds. Overall, pause duration does not show a dramatic change between the normal and non-rapport periods across the experiments.

Pauses	Mean duration in normal period	Standard deviation normal period	Mean duration in non-rapport period	Standard deviation non-rapport period
EXP 4	0,63	1,25	0,80	1,02
EXP 5	0,49	0,72	0,33	0,32
EXP 6	0,68	0,97	0,45	0,49
EXP 8	0,59	0,73	0,70	0,88
EXP 9	0,52	0,95	0,66	0,60
EXP 10	0,49	0,61	0,53	0,54
Totaal	0,57	0,91	0,56	0,67

Table 7.8: Comparison of the mean duration and standard deviation of pause length in the normal period and the non-rapport period.

Lastly, the mean gap lengths in each conversation was measured during the normal phase and the non-rapport phase. These gap lengths can be seen in table 7.9. The mean gap duration during the normal period across all experiments is 0.76 seconds. During the non-rapport period, the average gap duration increases significantly to 1.07 seconds, indicating that gaps between turns tend to be longer when rapport is disrupted. The increase in gap length during the non-rapport period reflects the breakdown of conversational coordination between speakers. Longer gaps suggest hesitation, delayed responses or less fluid interactions as a result of the rapport disruption.

Table 7.9: Comparison of the mean duration and standard deviation of the gap length in the normal period and the non-rapport period.

Gaps	Mean duration in normal period	Standard deviation normal period	Mean duration in non-rapport period	Standard deviation non-rapport period
EXP 4	0,76	0,72	1,13	0,97
EXP 5	0,79	2,18	0,85	0,76
EXP 6	0,98	1,15	0,99	0,98
EXP 8	0,90	0,91	0,77	0,77
EXP 9	0,52	0,51	1,63	1,96
EXP 10	0,55	0,44	1,32	1,71
Totaal	0,76	1,09	1,07	1,18

The results of these tests are presented in Table 7.10. Conversational features, including moments of smooth turn-taking, pause length, and gap length, are compared, with *n=6* representing six experiments.

Table 7.10: conversational features during the rapport-forming and non-rapport-forming periods. The features are moments of smooth turn-taking, pause length, and gap length. The t-statistics and p-values to determine whether the differences between the two periods are statistically significant.

Comparison conversational features				
Feature	Value rapport forming	Value non-rapport forming	T-statistic	p-value
Moments of smooth turn-taking	15.36 %	12.03 %	2.07	0.094
Pauze length	0.57 seconds	0.55 seconds	0.294	0.0769
Gap length	0.76 seconds	1.06 seconds	2.490	0.014

The table highlights how conversational features shift when rapport is broken. Only the difference measured in the gap length is statistically significant. The increased gap length during the non-rapport period reflects a delay in responses, suggesting that participants are less coordinated or more hesitant when rapport is disrupted. Additionally, while smooth turn-taking shows a trend toward reduction in the non-rapport period, the difference is not statistically significant.

7.2.2 Non-verbal cues

The first non-verbal cue analyzed is the amount of smile synchrony. The results of this measurement can be seen in table 7.11. The data shows a clear reduction in synchronized smiles during the non-rapport period, with the total average dropping from 7.06 smiles per minute in the normal period to 2.28 smiles per minute in the non-rapport period. This overall reduction in synchronized smiles suggests that rapport disruption significantly affects the emotional alignment between the actor and participant, as measured by smiling behavior.

The next non-verbal cue is the measured head movements. These can be seen in table 7.12 It can be seen that synchronized head movements become more frequent during the non-rapport phase. In each experiment, the mean phase difference shifts significantly when comparing the normal period to the non-rapport period, indicating that the dominance

	Synchronized smiles on average each minute in normal period	Synchronized smiles on average each minute in non-rapport period
Exp 4	0,41	0
Exp 5	5,9	2,74
Exp 6	11,42	2,06
Exp 8	8,27	3,27
Exp 9	4,55	1,35
Exp 10	11,8	4,24
Total	7,06	2,28

Table 7.11: This table provides a comparison of the average number of synchronized smiles per minute during the normal period and the non-rapport period across six experiments.

or leadership in synchronized head movements changes during the non-rapport phase. Notably, experiment 10 has NAN values in some columns due to the absence of synchronized head movements, which explains the missing data. By comparing the density, mean phase difference, standard deviation, and kurtosis between these two conditions, the analysis seeks to determine whether the disruption of rapport significantly alters head movement synchrony between the participants.

Table 7.12: Comparison of head movements in the normal period and the non-rapport period. This table presents the density, mean phase difference, standard deviation, and kurtosis of head movements for both the normal period and non-rapport period across several experiments. These metrics provide insights into the synchrony and variability of head movements between the actor and the participant in each period.

	Density normal period	Mean phase difference normal period	Std. Dev. normal period	Kurtosis normal period	Density non-rapport period	Mean phase difference non-rapport period	Std. Dev. non-rapport period	Kurtosis non-rapport period
Exp4	8,70	507,94	299,05	-0,45	5,29	-784,33	88,50	-1,08
Exp5	0,65	-228,14	162,39	-1,85	0,81	434,50	177,50	-2,00
Exp6	1,25	103,08	666,83	-1,37	5,83	-196,82	382,99	-0,79
Exp8	0,38	-226,50	424,95	-1,26	3,67	364,11	312,44	-1,43
Exp9	4,94	-448,35	247,84	-1,10	9,02	317,05	479,90	-1,25
Exp10	0,00	NAN	NAN	NAN	0,00	NAN	NAN	NAN
Average	3,18	-58,40	360,21	-1,20	4,92	26,90	288,27	-1,31

To assess whether there was a significant difference in the density of synchronized smiling movements and head movements between the normal period and the non-rapport period, a paired t-test was conducted. The density measurements for each period were paired for each experiment, and the goal was to determine if the mean densities significantly differed between the two conditions.

To assess whether the mean difference between these two sets of paired data is statistically significant, a t-test is conducted, as shown in Table 7.13 (*n=6, representing 6 experiments*). The table includes the mean density of these features during both periods, along with the results of t-tests (t-statistics and p-values) to determine whether the differences between the two periods are statistically significant. The test show a significant result for smiling synchrony but not for head movement

Table 7.13: In this table the density of the two non-verbal features, smiling synchrony and head movement synchrony, during the rapport forming period and the non-rapport-forming period can be seen.

Comparison non-verbal cues				
Feature	Rapport forming	Non-rapport forming	t-statistic	p-value
Amount of synchronized smiling	7.06	2.28	3.60	0.016
Amount of synchronized head movements	3,18	4,92	-0.071	0.946

The findings shown in the table show that smiling synchrony is a reliable indicator of non-rapport in conversations, with a significant drop when rapport is disrupted. On the other hand, head movement synchrony appears to remain too stable, suggesting that this physical coordination is less affected by rapport breakdown or this type of physical coordination does not show a moment of non-rapport.

7.2.3 Heart rate variability

Heart rate variability was analyzed using the Pearson correlation coefficient. This resulted in moments of synchrony, moments of maintenance and moments of asynchrony. To analyze the difference between the periods of rapport and non-rapport, the moments in each period were counted. This can be seen in table 7.14.

Table 7.14: Overview of synchrony, maintenance, and asynchrony moments during normal and non-rapport periods across all the experiments. The table displays the absolute counts and corresponding percentages of each type of moment for both normal and non-rapport periods.

	Synchrony moments normal period	Maintanance moments normal period	Asynchrony moments normal period	Synchorny moments non-rapport period	Maintanance moments non-rapport period	Asynchrony moments non-rapport period
Exp 4	31	206	44	7	24	5
Exp 4 percentage	11,03	73,31	15,66	19,44	66,67	13,89
Exp 5	40	189	34	3	42	3
Exp 5 percentage	15,21	71,86	12,93	6,25	87,5	6,25
Exp 6	35	232	19	8	33	2
Exp 6 percentage	12,24	81,12	6,64	18,60	76,74	4,65
Exp 8	25	249	28	4	42	4
Exp 8 percentage	8,28	82,45	9,27	8	84	8
Exp 9	33	226	28	2	35	5
Exp 9 percentage	11,50	78,75	9,76	4,76	83,33	11,90
Exp 10	25	192	25	7	41	3
Exp 10 percentage	10,33	79,34	10,33	13,73	80,39	5,88
Total	189	1294	178	31	217	22
Total percentage	11,38	77,90	10,72	11,48	80,37	8,15

In table 7.14 it can be seen that the percentage of synchrony moments is relatively consistent across the normal and non-rapport periods, with 11.38% during the normal period and 11.48% during the non-rapport period. In contrast, the percentage of asynchrony moments shows a noticeable difference, with 10.72% during the normal period compared to 8.15% during the non-rapport period. To determine whether the observed differences in synchrony and asynchrony moments are statistically significant, the odds ratios for these moments are referenced, as presented in table 7.15. The odds ratios provide a quantitative measure of the likelihood of these moments occurring in one period compared to the other.

The results show that the odds ratio for synchrony moments is 0.98, indicating that synchrony is almost equally likely to occur in both the normal and non-rapport periods. The corresponding p-value of 0.92

Table 7.15: Comparison of odds ratios and p-values for synchrony, maintenance, and asynchrony moments between normal and non-rapport periods.

Comparison heart rate variability		
Moment	Odds ratio	p-value
Synchrony	0,98	0,92
Maintenance	0,86	0,38
Asynchrony	1,35	0,24

suggests that this difference is not statistically significant. For maintenance moments, the odds ratio is 0.86, reflecting a slight decrease in the likelihood of these moments during the normal period compared to the non-rapport period. However, with a p-value of 0.38, this difference is also not statistically significant, indicating that maintenance moments remain consistent across both conditions. Asynchrony moments, on the other hand, have an odds ratio of 1.35, suggesting that asynchrony is 35% more likely to occur during the normal period than during the non-rapport period. Despite this apparent trend, the p-value of 0.24 indicates that the difference is also not statistically significant. None of the differences in synchrony, maintenance, or asynchrony moments between normal and non-rapport periods are statistically significant as all the p-values are larger than 0.05. This suggests that heart rate variability, in terms of these specific moments, may not be a strong indicator of rapport status during interaction.

7.3 Random Forest and Logistic Regression

To explore the relationships between features and their ability to classify rapport and non-rapport moments, Random Forest and Logistic Regression models were applied to the dataset. The analysis aimed to identify which features contribute most significantly to detecting these interaction states and to evaluate the impact of excluding non-verbal features. This approach seeks to answer the question: How well does a model perform in classifying rapport in conversations with and without non-verbal features?

The Random Forest model was trained using 100 trees (estimators), with the Gini impurity criterion for determining splits. A 5-fold cross-validation was employed to ensure generalizability and avoid overfitting. This approach provided a balanced evaluation of the model's performance across different subsets of the data. The Logistic Regression model was trained with L2 regularization to prevent overfitting, using a 5-fold cross-validation approach to ensure that the model generalized well to unseen data. The regularization parameter was optimized through grid search. Both model's performance was evaluated using accuracy, precision, recall, and F1-score. In addition, ROC-AUC was calculated to assess the model's ability to distinguish between normal and non-rapport windows. The results from the random forest model can be seen in table 7.16 and the results from the logistic regression model can be seen in table 7.17.

Random Forest Classification Report				
	precision	recall	f1-score	support
	precision	recall	f1-score	support
non_rapport	0.89	0.80	0.84	10
normal	0.80	0.89	0.84	9
accuracy			0.84	19
macro avg	0.84	0.84	0.84	19
weighted avg	0.85	0.84	0.84	19

Table 7.16: Classification performance of Random Forest model. The model has an overall accuracy of 0.84.

Logistic Regression Classification Report				
	precision	recall	f1-score	support
non_rapport	1	0.7	0.82	10
normal	0.75	1	0.86	9
accuracy			0.84	19
macro avg	0.88	0.85	0.84	19
weighted avg	0.88	0.84	0.84	19

Table 7.17: Classification performance of Logistic regression model. The model has an overall accuracy of 0.84. This model has perfect precision for non-rapport moments because it missed some non-rapport moments but perfectly captured all normal moments

The Random Forest model achieved an accuracy of 84% with a recall of 80% for the non-rapport class. The Logistic Regression model also achieved an accuracy of 0.84%, with a perfect precision (100%) for the non-rapport class, meaning every instance predicted as non-rapport was correct. But a recall of 70% for identifying non-rapport periods. Indicating some false positives in predicting normal moments. Random Forest maintains a more balanced performance across both classes in terms of precision and recall. Logistic Regression could be ideal if the goal is to ensure no false positives in detecting non-rapport moments, though it may miss some instances (lower recall for non-rapport). The next step is to look into the importance of the features used by both models. For the random forest model, this is the feature importance list shown in figure 7.2. For the logistic regression model, the coefficient values are shown in figure 7.3.

The importance values for the random forest model are a measure of how frequently and effectively a feature is used by the model's decision trees to split data for accurate classification. The most important feature in this model is the amount of synchronized smiles, with an importance score of almost 0.20. This indicates that synchronized smiling behaviour plays a significant role in determining whether the interaction is classified as rapport or non-rapport. The coefficient values representing the importance of each feature in the Logistic Regression model for predicting rapport and non-rapport periods. Unlike the Random Forest model, which ranks features based on how frequently they're used in decision trees, Logistic Regression provides coefficient values that indicate both the strength and direction (positive or negative) of the relationship between each feature and the prediction outcome. Similar to the Random Forest model, the amount of synchronized smiles is the most important feature in the Logistic Regression model, with the highest positive coefficient. Features such as the amount of synchronized head movements and smooth turn-taking moments have negative coefficients. In Logistic Regression, a negative coefficient indicates that an increase in this feature is more likely to be associated with non-rapport moments. The amount of synchronized head movements has the most significant negative coefficient, meaning higher levels of synchronized head movements reduce the likelihood of the interaction being classified as rapport and increase the likelihood of being classified as non-rapport.

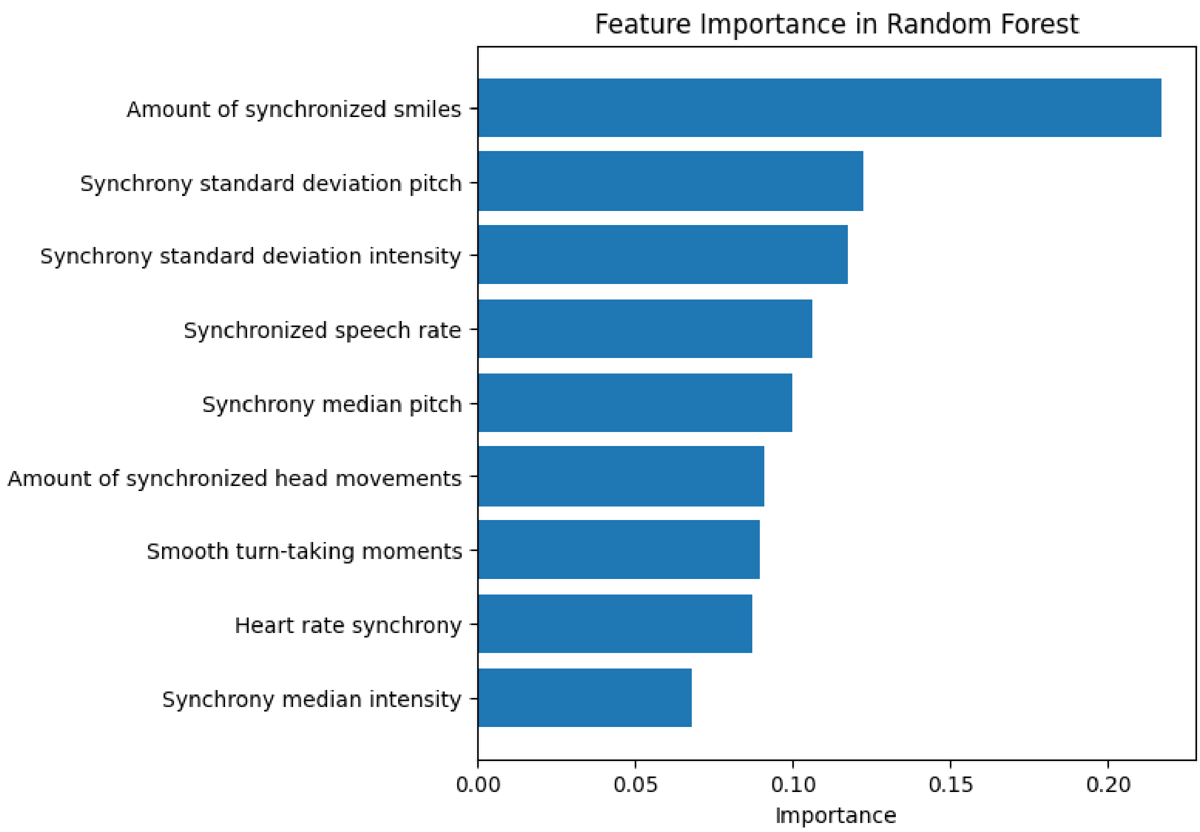

Figure 7.2: The importance of each feature in the Random Forest model, based on how much each feature contributes to distinguishing between the rapport and non-rapport moments in the conversation.

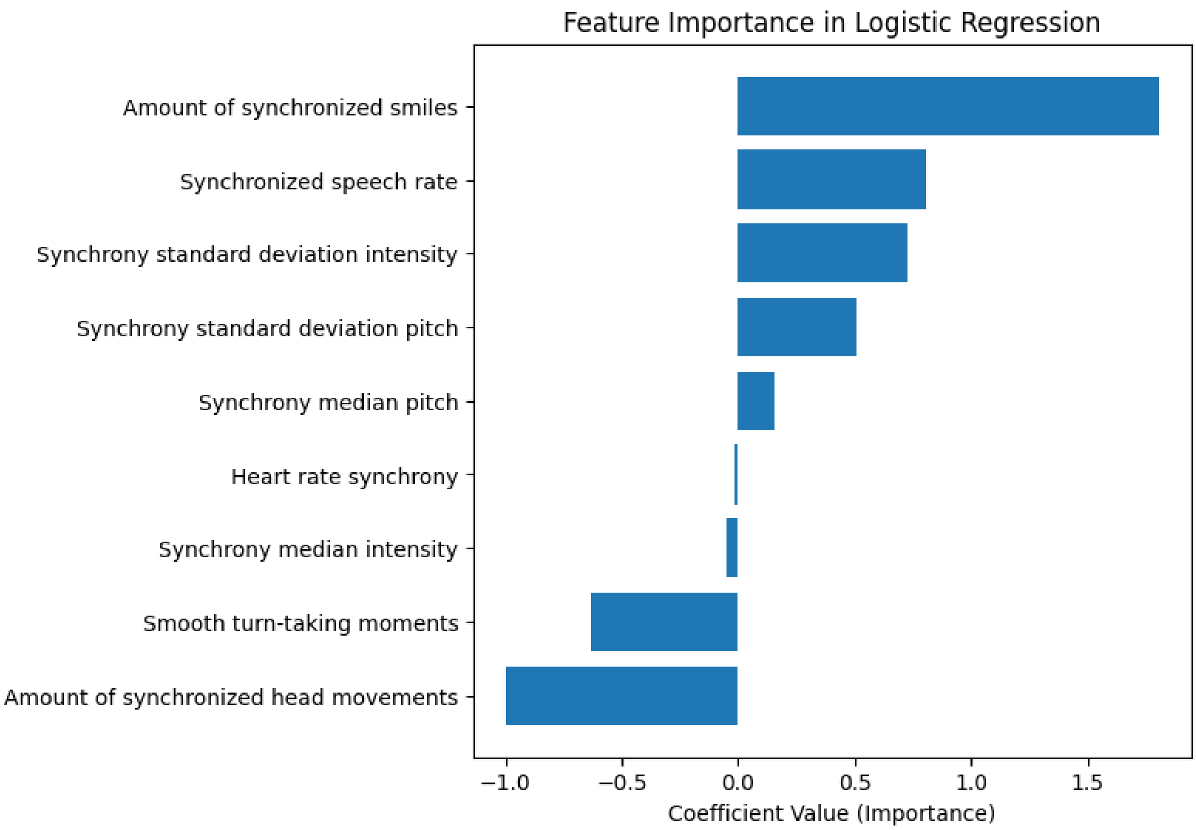

Figure 7.3: The coefficient values representing the importance of each feature in the Logistic Regression model for predicting rapport and non-rapport periods. Unlike the Random Forest model, which ranks features based on how frequently they're used in decision trees, Logistic Regression provides coefficient values that indicate both the strength and direction (positive or negative) of the relationship between each feature and the prediction outcome.

To assess the impact of non-verbal features on the models' ability to classify moments of rapport and non-rapport, both the Random Forest and Logistic Regression models were trained and evaluated on a dataset that excluded non-verbal features (synchronized smiles and head movements). The goal of this analysis was to understand how much the absence of non-verbal cues affects the models' performance and what this indicates about the differences between in-person and phone-based conversations. The results of the Random forest model without non-verbal features can be seen in table 7.20. The results of the Logistic regression model without non-verbal features can be seen in table 7.19.

Random Forest without non-verbal features Classification Report				
	precision	recall	f1-score	support
non_rapport	0.60	0.60	0.60	10
normal	0.56	0.56	0.56	9
accuracy			0.58	19
macro avg	0.58	0.58	0.58	19
weighted avg	0.58	0.58	0.58	19

Table 7.18: Classification performance of Random Forest model without non-verbal features. The model has an overall accuracy of 0.84.

The Random Forest model without non-verbal features achieved an accuracy of 58%, which is significantly lower than the 84% accuracy observed when non-verbal features were included. The precision, recall, and F1-scores for both classes were reduced, indicating that the model struggled to distinguish between rapport and non-rapport moments based solely on verbal and physiological features

Logistic Regression without non-verbal features Classification Report				
	precision	recall	f1-score	support
non_rapport	0.54	0.70	0.61	10
normal	0.50	0.33	0.40	9
accuracy			0.53	19
macro avg	0.52	0.52	0.50	19
weighted avg	0.52	0.52	0.51	19

Table 7.19: Classification performance of Logistic regression model without non-verbal features. The model has an overall accuracy of 0.84.

The Logistic Regression model without non-verbal features showed even lower performance, achieving an accuracy of 53%. While the recall for the non-rapport class remained relatively high (70%), the recall for the normal class dropped significantly to 33%, indicating that the model frequently misclassified normal moments as non-rapport. Both models experienced a notable decline in performance when non-verbal features were excluded, reinforcing the hypothesis that these cues are pivotal in detecting rapport. Without the features synchronized smiling and head movement both Random Forest and Logistic Regression struggle to achieve acceptable performance, underscoring the importance of these non-verbal cues.

7.4 Summary

In this chapter, various features were explored and their relevance in predicting and measuring rapport in conversations was looked into. With the experiment focusing on a normal and a non-rapport phase, the analysis examined which features were effective in measuring the non-rapport period, how dynamic the features were during the experiment using correlation matrices, and the performance of two different machine learning models in predicting whether a 30-second window of the conversation belonged to the normal period or the non-rapport period. Table 7.20 provides an overview of the features that were analyzed in the study, along with their effectiveness in measuring rapport during conversations, their measurability in phone conversations, and their importance as determined by the Random Forest and Logistic Regression models.

Table 7.20: Overview of the features that were analyzed in the study, along with their effectiveness in measuring rapport during conversations, their measurability in phone conversations, and their importance as determined by the Random Forest and Logistic Regression models. A measurement of rapport in the conversation is stated as effective if the statistical test was significant. Feature importance in Random Forest was determined based on whether a feature had an importance score higher than 0.5. The effectiveness in the logistic regression was determined based on the coefficient values. Features with high positive or negative coefficients were considered effective, again higher than 0.5 or lower than -0.5. Pause Lengths and Gap Lengths are also measurable in phone conversations but were not incorporated into the feature importance models due to those features being difficult to measure in moments.

Feature	Effective for measuring rapport in conversation	Possibly measurable in phone conversations	Effective for random forest	Effective for Logistic regression
Synchrony median pitch	No	Yes	No	No
Synchrony standard deviation pitch	Yes	Yes	Yes	No
Synchrony median intensity	Yes	Yes	No	No
Synchrony standard deviation intensity	No	Yes	Yes	Yes
Amount of synchronized head movements	No	No	No	Yes
Amount of synchronized smiles	Yes	No	Yes	Yes
Smooth turn-taking moments	No	Yes	No	Yes
Synchronized heart rate	No	Yes	No	No
Speech rate	Yes	Yes	Yes	Yes
Pause lengths	No	Yes	-	-
Gap lengths	Yes	Yes	-	-

No feature scores a yes in all the boxes, which is interesting for our main research question. The amount of synchronized smiles was shown to be one of the most important features in both the Random Forest and Logistic Regression models. This highlights that synchronized smiling is a key behavior associated with rapport, despite it not being measurable in phone conversations. Synchronized head movements and heart rate synchrony were less effective for rapport measurement and were not prioritized by the models. heart rate synchrony is stated as measurable during a phone conversation because in a lab setting, this could be measured.

Discussion 8

This chapter reflects on the findings of this study, addressing the main research question and the sub-questions. This chapter is divided into three parts: the first discusses the study's aims and key findings, the second highlights the limitations of the research, and the final section looks at the future directions.

8.1 Key findings and insights

The aim of our study was to answer our main research question:

How can we quantitatively measure and assess the level of rapport in customer-agent phone conversations?.

To address our main research question, the following sub-questions were formulated:

1. What measurable features show significant differences between moments of rapport and non-rapport in conversations?
2. What is the impact of removing non-verbal features on a model's ability to classify rapport in conversations?

The results of this study highlight that several measurable features, such as standard deviation of pitch, synchrony median intensity, speech rate, the gap length and synchronized smiles showed significant differences between the normal and non-rapport periods. The machine learning models showed the significant role of synchronized smiling in classifying moment. This aligns with existing research done by Gironzetti et al. (2016), which highlights smiling as a key indicator of emotional alignment and rapport [16]. Smiling often reflects shared positive affect, fostering a sense of connection and comfort. In our study, participants may have used smiling as a strategy to mitigate discomfort during non-rapport moments, particularly given the experimental nature of the setup.

In contrast, other features, like head movements, smooth turn-taking moments, heart rate synchrony, the median pitch synchrony and the synchrony in the standard deviation of the intensity did not show significant differences between rapport and non-rapport periods. This may indicate that while these non-verbal cues are relevant during conversations, they may not be as sensitive to our rapport measurements or could vary significantly depending on individual differences between participants.

When non-verbal features were removed from the machine learning models, performance dropped quite much, with Random Forest and Logistic Regression models achieving only 58% and 53% accuracy. This highlights the critical importance of non-verbal features synchronized smiles and head movements when identifying rapport in conversations. This result aligns with existing research about the role of non-verbal features in rapport building. For instance, studies by Bernieri et al. (1991)

and Tsuchiya et al. (2020) highlight that synchrony in behaviors[5], such as smiling or body movement[21], enhances information exchange and interpersonal connection. These findings reinforce the importance of non-verbal features in detecting and fostering rapport. This suggests that these cues should be prioritized in situations where visual channels are available, like face-to-face interactions or video calls. However, the notable drop in classification accuracy when non-verbal features are excluded highlights the challenges in phone-based conversations, where non-verbal cues are absent. This underscores the need to explore further strategies for specifically measuring the level of rapport in phone conversations. It could be possible that in phone conversations, the absence of non-verbal cues places greater importance on prosodic synchrony to compensate for the lack of visual feedback.

Our results align with the research on synchrony in acoustic features. The results of our measurements in pitch and intensity synchrony are quite similar to the results from the study done by De Looze et al. (2014). Which also found a difference in the synchrony of the standard deviation of the pitch. The result of their moving window analysis also reveals that some features show higher levels of synchrony than others. But their conclusion is that synchrony dynamically evolves over phases of conversations rather than increases/decreases continuously over the course of a conversation[30].

8.2 Limitations

There are several limitations to the current study that should be addressed in future research. First, the method for detecting smooth turn-taking was limited. As noted, smooth turn-taking moments could occur while one participant is still speaking, making it difficult to detect overlapping speech and its potential impact on rapport. This may have led to underestimating or overestimating the importance of these moments in the conversation.

The results of the questionnaire offered valuable insights into participants' subjective experiences during the conversations. Open-ended responses revealed that the experimental setup may have been understood by the participants, with several describing the conversation as "set up" or lacking spontaneity. Some participants were also smiling throughout the non-rapport period. This could have been because they did find it difficult to handle the moment or felt the moment was rehearsed or artificial. This perception likely diminished the intended effect, as participants may not have genuinely experienced these moments as lacking rapport.

Furthermore, individual differences in conversational style may have influenced the effects of specific features. Participants with distinct interaction styles could have displayed varying levels of rapport, which might not have been captured fully by the features analyzed. It is also possible that the experimental conditions did not create a rapport shift to affect these features significantly but it created something else. This could explain the difference in opinion about the significance of head movement synchrony during rapport stated by Kwon et al.(2023), which found head movement synchrony to be strongly indicative of interpersonal coordination[38].

Although the models, such as Random Forest and Logistic Regression, performed reasonably well in classifying periods of rapport and non-rapport, the dataset was relatively small. With only a limited number of experiments and conversations analyzed, the models may have encountered challenges in generalizing the results to broader contexts. With smaller datasets, models are more prone to overfitting, where the model performs well on the training data but fails to generalize to unseen data. This could also be seen in the huge differences between the experiments for example in the amount of synchronized smiles. One person smiles way more when talking than other people do.

8.3 Future directions

Our results have practical implications, particularly for developing tools to improve customer-agent interactions. Monitoring pitch variability and turn-taking dynamics could offer a way to detect and address moments of low rapport in real time. However, further research is needed to explore additional indicators of rapport.

Firstly, while this study analyzed measurements of audio features, future work should focus on the dynamic nature of these features over time. Investigating how pitch, intensity, and speech rate synchrony evolve across different phases of a conversation could provide deeper insights into their role in rapport-building. For instance, analyzing key interaction moments, such as agreement or conflict resolution, could enhance our understanding of rapport dynamics.

Methodological improvements are also needed to address the limitations identified in this study. For instance, creating more naturalistic non-rapport moments. Rather than relying on interactions with a very hard non-rapport moment, a calmer and softer non-rapport moment could benefit the study and elicit more authentic responses from participants. Moreover, additional tools for detecting and analyzing subtle behaviours, such as overlapping speech or habitual smiling, should be developed to better capture the nuances of dynamic conversations. Another methodological factor that could be addressed in future studies is the contextual and individual factors that may influence rapport. Expanding the dataset to include more diverse participants and scenarios would help generalize the findings.

Furthermore, exploration of rapport in conversations without non-verbal features, such as phone conversations, is necessary. The absence of non-verbal cues in phone conversations presents unique challenges for rapport detection. While this study focused on audio-only settings, future research could explore how verbal features compensate for missing non-verbal cues or examine the interplay between audio and visual synchrony in multimodal communication settings, such as video calls. Combining these modalities may lead to more accurate and robust rapport detection systems.

9 Conclusion

The main objective of this thesis was to explore how rapport in customer-agent phone conversations could be measured and assessed. By analyzing features such as pitch, intensity, head movements, synchronized smiles, speech rate, smooth turn-taking and gap length the study aimed to uncover the specific features that distinguish normal periods from moments of non-rapport.

The findings demonstrated that several features showed significant differences between normal periods and non-rapport periods, particularly

- The standard deviation of pitch. The number of synchrony moments in the standard deviation of pitch was significantly higher during non-rapport periods, with an odds ratio of 8.46 (p-value = 0.006).
- The synchrony in median intensity. The number of synchrony moments in median intensity was also significantly different, with an odds ratio of 1.78 (p-value = 0.050).
- The speech rate, the average speech rate was 5.12% slower during non-rapport periods (t-statistic is -1.234 with a p-value = 0.045).
- The number of synchronized smiles, The number of synchronized smiles dropped from 7.06 on average per minute in rapport periods to 2.28 on average per minute in non-rapport periods (p-value = 0.016).
- The mean gap length is significantly longer in the periods of non-rapport (t-statistics= 2.490, p-value = 0.014).

These features demonstrated changes that aligned with shifts in rapport, suggesting that they could serve as reliable indicators of non-rapport periods. Notably, synchronized smiles and gap length appeared to be the strongest features for distinguishing between normal and non-rapport moments. In contrast, features like head movement synchrony (t-statistics= -0.071, p-value = 0.946) and smooth turn-taking moments (difference of 3.22%, p-value = 0.094) did not show significant differences, indicating their role in rapport-building may be context-dependent.

The machine learning models demonstrated a strong performance in classifying moments of non-rapport. The Random Forest model and the Logistic Regression model achieved an accuracy of 84% in distinguishing between normal and non-rapport periods. However, when non-verbal features were excluded, accuracy dropped to 58% for Random Forest and 53% for Logistic Regression, highlighting the importance of non-verbal cues.

On the other hand, certain features, such as head movement synchrony and synchronized speech rate, did not show a strong association with rapport shifts, indicating that their role in rapport building may be more nuanced or context-dependent.

The results of this research provide insights into the measurable aspects of rapport and have practical implications for customer-agent interactions. These insights could inform the development of rapport monitoring tools, which would enable agents to adjust their communication style, fostering better interactions. However, the limitations of the dataset, as well as the exploratory nature of this research, suggest that further studies are necessary to validate these findings on a larger scale and across different contexts.

In summary, this study has provided a foundation for understanding how rapport in phone conversations can be measured through specific non-verbal and acoustic features. While certain features stood out as key indicators of rapport, others require further investigation to fully understand their role. Future research should aim to expand on these findings by exploring larger datasets, real-time applications, and additional conversational features.

Bibliography

Here are the references in citation order.

[1] Dec. 2022. URL: https://www.businesswire.com/news/home/20221206005186/en/UJET-Research-Reveals-Chatbots-Increase-Frustration-for-80-of-Consumers (cited on page 1).

[2] Jeffrey T Hancock, Mor Naaman, and Karen Levy. 'AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations'. In: *Journal of Computer-Mediated Communication* 25.1 (Jan. 2020), pp. 89–100. DOI: 10.1093/jcmc/zmz022 (cited on page 1).

[3] Stela Seo et al. 'Investigating People's Rapport Building and Hindering Behaviors When Working with a Collaborative Robot'. In: *International Journal of Social Robotics* 10 (Jan. 2018), pp. 147–161. DOI: 10.1007/s12369-017-0441-8 (cited on pages 1, 5).

[4] Helen Riess. 'The Science of Empathy'. In: *Journal of patient experience* 4 (2017), pp. 74–77 (cited on pages 1, 10).

[5] Frank Bernieri and R. Rosenthal. 'Interpersonal coordination: behavior matching and interactional synchrony'. In: *Fundamentals of Nonverbal Behavior. Studies in Emotion and Social Interaction,* (Jan. 1991) (cited on pages 1, 8, 9, 17, 64).

[6] A. Mehrabian. *Nonverbal Communication*. Aldine-Atherton, 1972 (cited on pages 1, 14).

[7] Linda Tickle-Degnen and Robert Rosenthal. 'The Nature of Rapport and Its Nonverbal Correlates'. In: *Psychological Inquiry - PSYCHOL INQ* 1 (Oct. 1990), pp. 285–293. DOI: 10.1207/s15327965pli0104_1 (cited on pages 5, 6).

[8] Takahiro Yokozuka et al. 'The Relationship Between Turn-taking, Vocal Pitch Synchrony, and Rapport in Creative Problem-Solving Communication'. In: *Speech Communication* 129 (2021), pp. 33–40. DOI: https://doi.org/10.1016/j.specom.2021.03.001 (cited on pages 5, 14, 20, 27, 34).

[9] Jessica L. Lakin and Tanya L. Chartrand. 'Using Nonconscious Behavioral Mimicry to Create Affiliation and Rapport'. In: *Psychological Science* 14.4 (2003). PMID: 12807406, pp. 334–339. DOI: 10.1111/1467-9280.14481 (cited on page 5).

[10] Janet Bavelas et al. '"I Show How You Feel". Motor Mimicry as a Communicative Act'. In: *Journal of Personality and Social Psychology* 50 (Feb. 1986), pp. 322–329. DOI: 10.1037/0022-3514.50.2.322 (cited on page 6).

[11] Marianne LaFrance. 'Nonverbal Synchrony and Rapport: Analysis by the Cross-Lag Panel Technique'. In: *Social Psychology Quarterly* 42.1 (1979), pp. 66–70. (Visited on 11/15/2023) (cited on pages 6, 14).

[12] Jonathan Gratch et al. 'Virtual Rapport'. In: Aug. 2006, pp. 14–27. DOI: 10.1007/11821830_2 (cited on page 6).

[13] Timo Kaski, Jarkko Niemi, and Ellen Pullins. 'Rapport building in authentic B2B sales interaction'. In: *Industrial Marketing Management* 69 (2018), pp. 235–252. DOI: https://doi.org/10.1016/j.indmarman.2017.08.019 (cited on page 6).

[14] Ellen Berscheid. 'Interpersonal Relationships'. In: *Annual Review of Psychology* 45.Volume 45, 1994 (1994), pp. 79–129. DOI: https://doi.org/10.1146/annurev.ps.45.020194.000455 (cited on page 7).

[15] Murray C. Clark and Roy L. Payne. 'The Nature and Structure of Workers' Trust in Management'. In: *Journal of Organizational Behavior* 18.3 (1997), pp. 205–224. (Visited on 11/28/2023) (cited on page 7).

[16] Elisa Gironzetti et al. In: *HUMOR* 29.2 (2016), pp. 301–324. DOI: doi:10.1515/humor-2016-0005 (cited on pages 7, 14, 20, 35, 63).

[17] Joanna Hale et al. 'Are You on My Wavelength? Interpersonal Coordination in Dyadic Conversations'. In: *Journal of Nonverbal Behavior* 44 (Mar. 2020). DOI: 10.1007/s10919-019-00320-3 (cited on pages 8, 15).

[18] Shalini Bhatia et al. 'Automated Measurement of Head Movement Synchrony during Dyadic Depression Severity Interviews'. In: vol. 2019. May 2019. DOI: 10.1109/FG.2019.8756509 (cited on page 8).

[19] Kate G. Niederhoffer and James W. Pennebaker. 'Linguistic Style Matching in Social Interaction'. In: *Journal of Language and Social Psychology* 21.4 (2002), pp. 337–360. DOI: 10.1177/026192702237953 (cited on pages 8, 16).

[20] Joanne Lumsden et al. 'Sync or sink? Interpersonal synchrony impacts self-esteem'. In: *Frontiers in Psychology* 5 (2014) (cited on page 8).

[21] Ayaka Tsuchiya et al. 'Body Movement Synchrony Predicts Degrees of Information Exchange in a Natural Conversation'. In: *Frontiers in Psychology* 11 (2020). DOI: 10.3389/fpsyg.2020.00817 (cited on pages 8, 64).

[22] Joana Coutinho et al. 'When our hearts beat together: Cardiac synchrony as an entry point to understand dyadic co-regulation in couples'. In: *Psychophysiology* 58 (Dec. 2020). DOI: 10.1111/psyp.13739 (cited on pages 8, 15).

[23] D.J.W Smits. 'Movement synchrony and Heart Rate Synchrony in psychotherapy: A Times series and support vector regression analysis'. PhD thesis. Tilburg University. Cognitive Science and Artificial Intelligence, 2020 (cited on pages 9, 21, 38).

[24] Ken Fujiwara, Masanori Kimura, and Ikuo Daibo. 'Gender Differences in Synchrony: Females in Sync during Unstructured Dyadic Conversation'. In: *European Journal of Social Psychology* 49 (Mar. 2019). DOI: 10.1002/ejsp.2587 (cited on page 9).

[25] Jens Edlund, Mattias Heldner, and Julia Hirschberg. 'Pause and gap length in face-to-face interaction'. In: Sept. 2009, pp. 2779–2782. DOI: 10.21437/Interspeech.2009-710 (cited on pages 9, 14, 18, 20, 27, 34).

[26] Robert Hogan. 'Empathy: A Conceptual Psychometric Analysis'. In: *The Counseling Psychologist* 5.2 (1975), pp. 14–18. DOI: 10.1177/001100007500500204 (cited on page 10).

[27] Karen Phillip, Nathan Beel, and Tanya Machin. 'Understanding the Cues and Strategies Counsellors Use to Develop Rapport With Clients Through Telephone Counselling'. In: *Psychotherapy and Counselling Journal of Australia* 8 (Aug. 2020). DOI: 10.59158/001c.71253 (cited on page 10).

[28] Zac Imel et al. 'The Association of Therapist Empathy and Synchrony in Vocally Encoded Arousal'. In: *Journal of counseling psychology* 61 (Nov. 2013). DOI: 10.1037/a0034943 (cited on page 10).

[29] Colin Clark et al. 'Is Empathy Effective for Customer Service? Evidence From Call Center Interactions'. In: *Journal of Business and Technical Communication* 27 (Apr. 2013), pp. 123–153. DOI: 10.1177/1050651912468887 (cited on page 11).

[30] Céline De Looze et al. 'Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction'. In: *Speech Communication* 58 (2014), pp. 11–34. DOI: https://doi.org/10.1016/j.specom.2013.10.002 (cited on pages 13, 17–19, 30, 38, 64).

[31] Rivka Levitan and Julia Hirschberg. 'Measuring Acoustic-Prosodic Entrainment with Respect to Multiple Levels and Dimensions'. In: *Interspeech*. 2011 (cited on page 13).

[32] Kousidis S. *Times Series Analysis of Acoustic Feature Convergence in Human Dialogues*. 2008. URL: https://cir.nii.ac.jp/crid/1370290617572361860 (cited on page 13).

[33] Wolfgang Tschacher, Nikolai Tschacher, and Anja Stukenbrock. 'Eye Synchrony: A Method to Capture Mutual and Joint Attention in Social Eye Movements'. In: *Nonlinear dynamics, psychology, and life sciences* 25 (July 2021), pp. 309–333 (cited on page 15).

[34] Wolfram Boucsein. *Electrodermal activity: Second edition*. Aug. 2013, pp. 1–618 (cited on page 16).

[35] Hugo D. Critchley. 'Review: Electrodermal Responses: What Happens in the Brain'. In: *The Neuroscientist* 8.2 (2002). PMID: 11954558, pp. 132–142. DOI: 10.1177/107385840200800209 (cited on page 16).

[36] James Joseph Lynch et al. 'The Effects of Talking on the Blood Pressure of Hypertensive and Normotensive Individuals'. In: *Psychosomatic Medicine* 43 (1981), pp. 25–33 (cited on page 16).

[37] Gale Lucas et al. 'Getting to Know Each Other: The Role of Social Dialogue in Recovery from Errors in Social Robots'. In: Feb. 2018, pp. 344–351. DOI: 10.1145/3171221.3171258 (cited on pages 16, 17).

[38] Jinhwan Kwon and Hiromi Kotani. 'Head motion synchrony in unidirectional and bidirectional verbal communication'. In: *PLOS ONE* 18.5 (May 2023), pp. 1–15. DOI: 10.1371/journal.pone.0286098 (cited on pages 21, 35, 65).

[39] Francesca Gasparini, Alessandra Grossi, and Stefania Bandini. 'A Deep Learning Approach to Recognize Cognitive Load using PPG Signals'. In: *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference*. PETRA '21. Corfu, Greece: Association for Computing Machinery, 2021, pp. 489–495. DOI: 10.1145/3453892.3461625 (cited on page 25).

[40] nl-NL. URL: https://www.shure.com/nl-NL/producten/draadloze-systemen/blx_wireless/blx188-mx53-dual-earset-presenter-set?variant=BLX188E%252FMX53-K14 (cited on page 25).

[41] Tadas Baltrusaitis et al. 'OpenFace 2.0: Facial Behavior Analysis Toolkit'. In: *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*. 2018, pp. 59–66. DOI: 10.1109/FG.2018.00019 (cited on page 29).

[42] P. Ladefoged and K. Johnson. *A Course in Phonetics*. Cengage Learning, 2010 (cited on page 33).

[43] Rao Hamza Ali et al. 'Automated Identification of Astronauts on Board the International Space Station: A Case Study in Space Archaeology'. In: *Acta Astronautica* 200 (Aug. 2022). DOI: 10.1016/j.actaastro.2022.08.017 (cited on page 36).

[44] Chung-Chih Lin et al. 'Intelligent health monitoring system based on smart clothing'. In: *International Journal of Distributed Sensor Networks* 14 (Aug. 2018), p. 1550147718794318. DOI: 10.1177/1550147718794318 (cited on page 38).

[45] Leo Breiman. 'Random Forests'. In: *Machine Learning* 45.1 (2001), pp. 5–32. DOI: 10.1023/A:1010933404324 (cited on page 40).

[46] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, 2009 (cited on page 40).

[47] Fabian Pedregosa et al. 'Scikit-learn: Machine Learning in Python'. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cited on page 40).

Appendix

Appendix

Questionnaire

Questionnaire Research Measuring Conversation Quality
Tuesday 7-05-2024

What is your name? ______________________________

My interlocutor made sure there was a sense of connection and friendship between me and my interlocutor.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

My interlocutor created a sense of distance between us.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

I think my interlocutor and I understood each other.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

My interlocutor reacted to me rather gruffly.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

My interlocutor was caring.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

I was keen to create a sense of distance between me and my interlocutor.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

I felt a connection with my interlocutor.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

My interlocutor was respectful towards me.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

I felt no connection with my interlocutor.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

I made sure there was a sense of connection and friendship between me and my interlocutor.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

I responded rather gruffly to my interlocutor.

	1	2	3	4	5	
Completely disagree	o	o	o	o	o	Completely agree

Do you feel you had a good connection with your interlocutor during the conversation? Why or why not?

How did you feel when talking about your hobby? Why?

Do you have any further comments on the interview?

Were there any moments when you felt uncomfortable or tense during the conversation? If yes, can you describe those moments and what do you think caused them?

Are there things your interlocutor did or said that helped you feel more at ease during the conversation?

Praat script and audio script

Script to get the Pitch from selected file:

```
writeFileLine: "./pitch_list.txt", "time,pitch"
selectObject: 1
To Pitch: 0, 75, 600
no_of_frames = Get number of frames

for frame from 1 to no_of_frames
    time = Get time from frame number: frame
    pitch = Get value in frame: frame, "Hertz"
    appendFileLine: "pitch_list.txt", "'time','pitch'"
endfor
```

Script to get the Intensity from selected file:

```
writeFileLine: "./intensity_list.txt", "time,intensity"
selectObject: 1
To Intensity: 75, 0, "yes"
no_of_frames = Get number of frames

for frame from 1 to no_of_frames
    time = Get time from frame number: frame
    intensity = Get value in frame: frame
    appendFileLine: "intensity_list.txt", "'time','intensity'"
endfor
```

Script to make make the pitch analyzation using the pitch_list.txt file and the speaker diarisation Json file. The intensity file is the same but with the intensity_list.txt file.

```
# %%
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import statistics
from scipy.stats import pearsonr
from scipy.signal import hann

# %%
def read_file_to_dataframe(file_path):
    # Lees het bestand in
    df = pd.read_csv(file_path)

    # Vervang '--undefined--' door NaN
    df['pitch'] = df['pitch'].replace('--undefined--', np.nan)

    # Converteer de pitch kolom naar numerieke waarden (NaN-waarden blijven
behouden)
    df['pitch'] = pd.to_numeric(df['pitch'])

    return df


# %%
def plot_pitch_from_dataframe(df1, df2=None):
    plt.figure(figsize=(200, 10))

    # Plot de eerste DataFrame
    plt.plot(df1['time'], df1['pitch'], label='Actor')

    # Controleer of er een tweede DataFrame is meegegeven
    if df2 is not None:
        plt.plot(df2['time'], df2['pitch'], label='Participant')

    plt.xlabel('Time')
    plt.ylabel('Pitch')
    plt.title('Pitch over Time')
    plt.legend()
    plt.ylim(5,10)
    plt.show()


# %%
def handle_outliers(df, pitch_column, floor, ceiling):
    """Handle values outside the floor and ceiling by setting them to the
nearest boundary."""
```

```
    df.loc[df[pitch_column] < floor, pitch_column] = floor
    df.loc[df[pitch_column] > ceiling, pitch_column] = ceiling



def convert_to_octave_scale(df):
    """Convert pitch values to octave scale and handle outliers."""
    for pitch_column in ['pitch']:
        # Calculate the 15th and 65th percentiles
        p15 = np.percentile(df[pitch_column].dropna(), 15)
        p65 = np.percentile(df[pitch_column].dropna(), 65)

        # Set the pitch floor and ceiling
        pitch_floor = p15 * 0.83
        pitch_ceiling = p65 * 1.92

        # Handle outliers
        handle_outliers(df, pitch_column, pitch_floor, pitch_ceiling)

        # Convert pitch values to the octave scale
        df[pitch_column] = np.log2(df[pitch_column])

        print(f"{pitch_column} Floor: {pitch_floor}, Ceiling:
{pitch_ceiling}")

    return df




# %%


def hybrid_utterance_window(df_speaker1, df_speaker2, window_size=10,
step_size=5):
    segments = []
    start_time = 0
    segment_number = 1

    max_time = max(df_speaker1['time'].max(), df_speaker2['time'].max())

    while start_time < max_time:
        # Define the end time of the current window
        end_time = start_time + window_size

        # Extend the window to match utterance boundaries
        window_start = df_speaker1[df_speaker1['time'] >=
start_time]['time'].min()
```

```
        window_end = df_speaker1[df_speaker1['time'] <=
end_time]['time'].max()

        if pd.isna(window_start) or pd.isna(window_end):
            start_time += step_size
            continue

        # Extract pitch data within this window for both speakers
        pitch_speaker1 = df_speaker1[(df_speaker1['time'] >= window_start) &
(df_speaker1['time'] <= window_end)]
        pitch_speaker2 = df_speaker2[(df_speaker2['time'] >= window_start) &
(df_speaker2['time'] <= window_end)]

        # Merge pitch data from both speakers by time
        merged_data = pd.merge(pitch_speaker1[['time', 'pitch']],
                               pitch_speaker2[['time', 'pitch']],
                               on='time',
                               how='outer',
                               suffixes=('_actor', '_participant'))

        # Assign segment number
        merged_data['segment'] = segment_number

        segments.append(merged_data)

        # Move to the next window
        start_time += step_size
        segment_number += 1

    # Concatenate all segments into a single DataFrame
    result_df = pd.concat(segments).reset_index(drop=True)

    return result_df

# %%
# Gebruik het pad naar de bestanden die je wilt inladen
Actor_pitch = read_file_to_dataframe('Actor_pitch_exp4.txt')
Participant_pitch = read_file_to_dataframe('Participant_pitch_Exp4.txt')

Actor_converted= convert_to_octave_scale(Actor_pitch)
Participant_converted = convert_to_octave_scale(Participant_pitch)


# %%

plot_pitch_from_dataframe(Actor_converted, Participant_converted)
print(Actor_converted.median())
print(Participant_converted.median())
Actor_converted.to_csv("Actor_converted.csv")
```

```python
Participant_converted.to_csv("Participant_converted.csv")


# %%
Combined_df = hybrid_utterance_window(Actor_converted, Participant_converted)

# %%
Combined_df.to_csv("Combined_df.csv")

# %%
def interpolate_missing_values(df):
    # Interpolate missing values for pitch_actor and pitch_participant
    df['median_pitch_actor'] =
df['median_pitch_actor'].interpolate(method='linear', limit_direction='both')
    df['median_pitch_participant'] =
df['median_pitch_participant'].interpolate(method='linear',
limit_direction='both')
    df['std_pitch_participant'] =
df['std_pitch_participant'].interpolate(method='linear',
limit_direction='both')
    df['std_pitch_actor'] = df['std_pitch_actor'].interpolate(method='linear',
limit_direction='both')
    return df

def calculate_segment_statistics(df):
    # Group by segment
    grouped = df.groupby('segment')

    # Calculate median pitch and count of real values for each segment
    result = grouped.agg({
        'time': ['min', 'max'],
        'pitch_actor': ['median', lambda x: x.notna().sum(), 'std'],
        'pitch_participant': ['median', lambda x: x.notna().sum(), 'std']
    }).reset_index()

    # Rename columns for clarity
    result.columns = ['segment', 'start_time', 'end_time',
                      'median_pitch_actor', 'count_real_values_actor',
'std_pitch_actor',
                      'median_pitch_participant',
'count_real_values_participant', 'std_pitch_participant']


    interpolated_df = interpolate_missing_values(result)
    return interpolated_df

# %%
segment_statistics_df = calculate_segment_statistics(Combined_df)
```

```
# %%
def windowed_pearson_correlation(df, window_size, step_size):

    correlations = []
    p_values = []
    segments = []

    # Loop over the dataframe with a sliding window
    for start in range(0, len(df) - window_size + 1):
        end = start + window_size
        window_df = df.iloc[start:end]

        # Compute Pearson correlation and p-value for the current window
        corr, p_value = pearsonr(window_df['median_pitch_actor'],
window_df['median_pitch_participant'])

        correlations.append(corr)
        p_values.append(p_value)
        segments.append(df['segment'].iloc[start:end].mean())  # Use the mean
of the segment numbers as the segment label

    # Create a DataFrame to store the results
    result_df = pd.DataFrame({
        'segment': segments,
        'correlation': correlations,
        'p_value': p_values
    })

    return result_df

# %%
def windowed_pearson_correlation_SD(df, window_size, step_size):
    correlations = []
    p_values = []
    segments = []

    # Loop over the dataframe with a sliding window
    for start in range(0, len(df) - window_size + 1):
        end = start + window_size
        window_df = df.iloc[start:end]

        # Compute Pearson correlation and p-value for the current window
        corr, p_value = pearsonr(window_df['std_pitch_actor'],
window_df['std_pitch_participant'])

        correlations.append(corr)
```

```
        p_values.append(p_value)
        segments.append(df['segment'].iloc[start:end].mean())  # Use the mean
of the segment numbers as the segment label

    # Create a DataFrame to store the results
    result_df = pd.DataFrame({
        'segment': segments,
        'correlation': correlations,
        'p_value': p_values
    })

    return result_df




# %%
def plot_median_pitches_and_correlation(segment_df, correlation_df):
    fig, ax1 = plt.subplots(figsize=(40, 6))

    # Plot the median pitches for actor and participant on the left y-axis
    ax1.plot(segment_df['segment'], segment_df['median_pitch_actor'],
label='Actor Median Pitch', color='blue')
    ax1.plot(segment_df['segment'], segment_df['median_pitch_participant'],
label='Participant Median Pitch', color='red')
    ax1.set_xlabel('Segment')
    ax1.set_ylabel('Median Pitch')
    ax1.set_title('Median Pitches and Pearson Correlation by Segment')
    ax1.legend(loc='upper left')
    ax1.grid(True)

    ax1.axvspan(120, 140, color='yellow', alpha=0.4)

    # Create a secondary y-axis for the Pearson correlation
    ax2 = ax1.twinx()
    ax2.plot(correlation_df['segment'], correlation_df['correlation'],
label='Pearson Correlation', color='green')
    ax2.set_ylabel('Pearson Correlation')
    ax2.set_ylim(-1, 1)  # Set limits for correlation axis
    significant = correlation_df[correlation_df['p_value'] < 0.05]
    ax2.scatter(significant['segment'], significant['correlation'],
color='red', label='Significant Correlation (p < 0.05)', zorder=5)

    ax2.legend(loc='upper right')

    plt.show()


# %%
```

```
def plot_sd_pitches_and_correlation(segment_df, correlation_df):
    fig, ax1 = plt.subplots(figsize=(40, 6))

    # Plot the median pitches for actor and participant on the left y-axis
    ax1.plot(segment_df['segment'], segment_df['std_pitch_actor'],
label='Actor St. Dev. Pitch', color='blue',  alpha=0.5)
    ax1.plot(segment_df['segment'], segment_df['std_pitch_participant'],
label='Participant St. Dev. Pitch', color='red',  alpha=0.5)
    ax1.set_xlabel('Segment')
    ax1.set_ylabel('Standard deviation Pitch')
    ax1.set_title('Standard deviation Pitches and Pearson Correlation by
Segment')
    ax1.legend(loc='upper left')
    ax1.grid(True)

    ax1.axvspan(120, 140, color='yellow', alpha=0.4)

    # Create a secondary y-axis for the Pearson correlation
    ax2 = ax1.twinx()
    ax2.plot(correlation_df['segment'], correlation_df['correlation'],
label='Pearson Correlation', color='green')
    ax2.set_ylabel('Pearson Correlation')
    ax2.set_ylim(-1, 1)  # Set limits for correlation axis
    significant = correlation_df[correlation_df['p_value'] < 0.05]
    ax2.scatter(significant['segment'], significant['correlation'],
color='red', label='Significant Correlation (p < 0.05)', zorder=5)

    ax2.legend(loc='upper right')

    plt.show()


# %%


correlation_median_df = windowed_pearson_correlation(segment_statistics_df, 8,
4)
plot_median_pitches_and_correlation(segment_statistics_df,
correlation_median_df)

# %%
correlation_sd_df = windowed_pearson_correlation_SD(segment_statistics_df, 8,
4)
plot_sd_pitches_and_correlation(segment_statistics_df, correlation_sd_df)

# %%
def count_significant_correlations_in_period(correlation_df, start_segment,
end_segment, significance_level=0.05):
    # Filter for significant correlations
```

```
    significant_correlations = correlation_df[correlation_df['p_value'] <
significance_level]

    # Filter for non-significant correlations (maintenance moments)
    maintenance_moments = correlation_df[correlation_df['p_value'] >=
significance_level]

    # Filter the significant correlation DataFrame for the specified period
    inside_period =
significant_correlations[(significant_correlations['segment'] >=
start_segment) &
                                             (significant_correlations['segmen
t'] <= end_segment)]

    # Filter the significant correlation DataFrame for outside the specified
period
    outside_period =
significant_correlations[(significant_correlations['segment'] < start_segment)
|
                                              (significant_correlations['segme
nt'] > end_segment)]

    # Filter the maintenance moments for the specified period
    maintenance_inside = maintenance_moments[(maintenance_moments['segment']
>= start_segment) &
                                             (maintenance_moments['segment']
<= end_segment)]

    # Filter the maintenance moments for outside the specified period
    maintenance_outside = maintenance_moments[(maintenance_moments['segment']
< start_segment) |
                                              (maintenance_moments['segment']
> end_segment)]

    # Count positive and negative correlations inside the period
    positive_inside = (inside_period['correlation'] > 0).sum()
    negative_inside = (inside_period['correlation'] < 0).sum()

    # Count positive and negative correlations outside the period
    positive_outside = (outside_period['correlation'] > 0).sum()
    negative_outside = (outside_period['correlation'] < 0).sum()

    # Count maintenance moments (non-significant correlations) inside and
outside the period
    maintenance_inside_count = len(maintenance_inside)
    maintenance_outside_count = len(maintenance_outside)

    # Create a dictionary to store the results
    result = {
```

```
        'positive_inside': positive_inside,
        'negative_inside': negative_inside,
        'maintenance_inside': maintenance_inside_count,
        'positive_outside': positive_outside,
        'negative_outside': negative_outside,
        'maintenance_outside': maintenance_outside_count
    }

    return result


# Example usage with the correlation_df from the previous function
start_segment = 120  # Define the start of the period
end_segment = 140    # Define the end of the period

correlation_counts_sd =
count_significant_correlations_in_period(correlation_sd_df, start_segment,
end_segment)

# Display the results
print(correlation_counts_sd)

# %%
def count_significant_correlations_in_period(correlation_df, start_segment,
end_segment, significance_level=0.05):
    # Filter for significant correlations
    significant_correlations = correlation_df[correlation_df['p_value'] <
significance_level]

    # Filter for non-significant correlations (maintenance moments)
    maintenance_moments = correlation_df[correlation_df['p_value'] >=
significance_level]

    # Filter the significant correlation DataFrame for the specified period
    inside_period =
significant_correlations[(significant_correlations['segment'] >=
start_segment) &
                                             (significant_correlations['segmen
t'] <= end_segment)]

    # Filter the significant correlation DataFrame for outside the specified
period
    outside_period =
significant_correlations[(significant_correlations['segment'] < start_segment)
|
                                              (significant_correlations['segme
nt'] > end_segment)]

    # Filter the maintenance moments for the specified period
```

```
    maintenance_inside = maintenance_moments[(maintenance_moments['segment']
>= start_segment) &
                                              (maintenance_moments['segment']
<= end_segment)]

    # Filter the maintenance moments for outside the specified period
    maintenance_outside = maintenance_moments[(maintenance_moments['segment']
< start_segment) |
                                               (maintenance_moments['segment']
> end_segment)]

    # Count positive and negative correlations inside the period
    positive_inside = (inside_period['correlation'] > 0).sum()
    negative_inside = (inside_period['correlation'] < 0).sum()

    # Count positive and negative correlations outside the period
    positive_outside = (outside_period['correlation'] > 0).sum()
    negative_outside = (outside_period['correlation'] < 0).sum()

    # Count maintenance moments (non-significant correlations) inside and
outside the period
    maintenance_inside_count = len(maintenance_inside)
    maintenance_outside_count = len(maintenance_outside)

    # Create a dictionary to store the results
    result = {
        'positive_inside': positive_inside,
        'negative_inside': negative_inside,
        'maintenance_inside': maintenance_inside_count,
        'positive_outside': positive_outside,
        'negative_outside': negative_outside,
        'maintenance_outside': maintenance_outside_count
    }

    return result


# Example usage with the correlation_df from the previous function
start_segment = 120  # Define the start of the period
end_segment = 140    # Define the end of the period

correlation_counts =
count_significant_correlations_in_period(correlation_median_df, start_segment,
end_segment)

# Display the results
print(correlation_counts)

# %%
```

```python
def merge_and_save(df1, df2):
    """
    This function merges two DataFrames on a common column (e.g., 'time') and
saves the merged DataFrame as a CSV file.

    :param df1: The first DataFrame to merge.
    :param df2: The second DataFrame to merge.
    :param output_file: The path where the merged DataFrame will be saved as a
CSV file.
    :return: The merged DataFrame.
    """
    # Merge the two DataFrames on the 'time' column
    merged_df = pd.merge(df1, df2, on='segment', how='outer')

    # Save the merged DataFrame as a CSV file
    merged_df['start_time'] = merged_df['start_time'].interpolate()
    merged_df['end_time'] = merged_df['end_time'].interpolate()
    merged_df = merged_df.dropna(subset=['correlation'])
    merged_df = merged_df.dropna(axis=1, how='all')

    return merged_df

# %%
Pitchcsv = merge_and_save(segment_statistics_df, correlation_median_df)
Pitchcsv.to_csv('PitchMedian_EXP4.csv', index=False)
Pitch_SD_csv = merge_and_save(segment_statistics_df, correlation_sd_df)
Pitch_SD_csv.to_csv('Pitch_SD_EXP4.csv', index=False)
```

OpenFace script

For finding synchronized smiles:

```
# %%
import os
import pandas as pd
import subprocess
import csv
import numpy as np
import matplotlib.pyplot as plt
import cv2
import csv
import seaborn as sns
from datetime import datetime, timedelta
from scipy.stats import kurtosis

# %%
def extract_smile_timestamps_L2(csv_file):
    smile_timestamps = []
    with open(csv_file, 'r') as f:
        reader = csv.DictReader(f)
        for row in reader:
            # Check if smile is detected (you may need to adjust thresholds)
            if float(row[' AU06_c']) > 0.95 and float(row[' AU12_c']) > 0.95:
                smile_timestamps.append(float(row[' timestamp']))
    return smile_timestamps
# Path to the OpenFace CSV file
csv_file_path = "P_exp4.csv"
csv_file_path2 = "A_exp4.csv"

# Extract smile timestamps

smile_timestamps_L2 = extract_smile_timestamps_L2(csv_file_path)
smile_timestamps2_L2 = extract_smile_timestamps_L2(csv_file_path2)

# %%
def combine_timestamps(list1, list2):
    # Get the maximum length of the two lists
    max_len = max(len(list1), len(list2))

    # Extend both lists to the same length by adding None or NaN
    list1.extend([None] * (max_len - len(list1)))
    list2.extend([None] * (max_len - len(list2)))

    # Convert the two lists into a DataFrame with two columns
    df = pd.DataFrame({
        'Person 1': list1,
        'Person 2': list2
```

```
    })

    return df

# %%
combined_df = combine_timestamps(smile_timestamps_L2, smile_timestamps2_L2)

# %%
def plot_smiling_synchronization(df, yellow_start, yellow_end,
cooldown=0.0167):
    """
    Plot a graph of smiling events for two individuals and highlight a
specific time period.

    Parameters:
    - df: DataFrame with two columns ('Person 1' and 'Person 2') containing
the timestamps in seconds.
    - yellow_start: Start time of the yellow period in seconds.
    - yellow_end: End time of the yellow period in seconds.
    - cooldown: Time (in minutes) within which multiple synchronized smiles
are counted as one.

    Returns:
    - The plot and prints the synchronized smiles per minute and total
synchronized smiles inside and outside the yellow period.
    """

    # Convert timestamps from seconds to minutes
    df_minutes = df / 60.0

    # Create the plot
    plt.figure(figsize=(40, 3))

    # Plot participant and actor timestamps as vertical lines
    for timestamp in df_minutes['Person 1'].dropna():
        plt.axvline(timestamp, color='blue', linewidth=2, label='Participant'
if timestamp == df_minutes['Person 1'].dropna().iloc[0] else "")

    for timestamp in df_minutes['Person 2'].dropna():
        plt.axvline(timestamp, color='red', linewidth=2, label='Actor' if
timestamp == df_minutes['Person 2'].dropna().iloc[0] else "")

    # Mark synchronized smiles (where both timestamps are within 0.5 seconds)
    synchronized_smiles = []
    threshold = 0.0083  # 0.5 seconds in minutes
    last_sync_time = -cooldown  # Initialize with a very low value to avoid
skipping first event

    for t1 in df_minutes['Person 1'].dropna():
```

```
        for t2 in df_minutes['Person 2'].dropna():
            if abs(t1 - t2) <= threshold:
                if t1 - last_sync_time >= cooldown:  # Ensure cooldown period
has passed before counting again
                    plt.axvline(t1, color='purple', linestyle='--',
linewidth=2, label='Synchronized' if not synchronized_smiles else "")
                    synchronized_smiles.append(t1)
                    last_sync_time = t1

    # Highlight the yellow period
    yellow_start_minutes = yellow_start / 60.0
    yellow_end_minutes = yellow_end / 60.0
    plt.axvspan(yellow_start_minutes, yellow_end_minutes, color='yellow',
alpha=0.3)

    # Add title and labels
    plt.title('Synchronization of Smiling Events')
    plt.xlabel('Time (minutes)')
    plt.ylim(0, 1)
    plt.legend()

    # Calculate synchronized smile counts and rates
    inside_yellow = [s for s in synchronized_smiles if yellow_start_minutes <=
s <= yellow_end_minutes]
    outside_yellow = [s for s in synchronized_smiles if s <
yellow_start_minutes or s > yellow_end_minutes]

    total_minutes_inside = (yellow_end - yellow_start) / 60.0
    total_minutes_outside = (df_minutes.max().max() - total_minutes_inside)

    # Total counts
    total_sync_smiles_inside = len(inside_yellow)
    total_sync_smiles_outside = len(outside_yellow)

    # Rates (per minute)
    rate_inside_yellow = total_sync_smiles_inside / total_minutes_inside if
total_minutes_inside > 0 else 0
    rate_outside_yellow = total_sync_smiles_outside / total_minutes_outside if
total_minutes_outside > 0 else 0

    # Show plot
    plt.show()

    # Print synchronized smile totals and rates
    print(f"Total synchronized smiles inside the yellow area:
{total_sync_smiles_inside}")
    print(f"Synchronized smiles per minute inside the yellow area:
{rate_inside_yellow:.2f}")
```

```
    print(f"Total synchronized smiles outside the yellow area:
{total_sync_smiles_outside}")
    print(f"Synchronized smiles per minute outside the yellow area:
{rate_outside_yellow:.2f}")




# %%

start_time = 9 *60 +53
end_time = 11 * 60 +35

plot_smiling_synchronization(combined_df, yellow_start=start_time,
yellow_end=end_time)

# %%
def detect_smile_events(timestamps, threshold=1):
    """
    Function to detect smile events (start and end) from continuous
timestamps.

    Parameters:
    timestamps (list or pd.Series): List of smile timestamps for a person.
    threshold (float): The maximum gap between successive timestamps to
consider them part of the same smile.

    Returns:
    pd.DataFrame: DataFrame with 'Start' and 'End' times for each detected
smile event.
    """
    smile_events = []
    current_start = timestamps[0]

    for i in range(1, len(timestamps)):
        if timestamps[i] - timestamps[i - 1] > threshold:
            # End the current smile event
            smile_events.append((current_start, timestamps[i - 1]))
            # Start a new smile event
            current_start = timestamps[i]

    # Add the last smile event
    smile_events.append((current_start, timestamps.iloc[-1]))

    # Create a dataframe with start and end times
    return pd.DataFrame(smile_events, columns=['Start', 'End'])


def get_synchronized_smiles(df, person1_col, person2_col):
```

```
    """
    Function to calculate synchronized smiles based on timestamps from two
persons.
    A synchronized smile is when one person smiles and the other smiles within
1 second after.

    Parameters:
    df (pd.DataFrame): Input dataframe containing timestamps for two persons.
    person1_col (str): Column name for Person 1's timestamps.
    person2_col (str): Column name for Person 2's timestamps.

    Returns:
    pd.DataFrame: DataFrame containing synchronized smiles (timestamps where
one person smiled and the other smiled within 1 second).
    """

    # Step 1: Detect smile events for both persons
    person1_smile_events =
detect_smile_events(df[person1_col].dropna().sort_values().reset_index(drop=Tr
ue))
    person2_smile_events =
detect_smile_events(df[person2_col].dropna().sort_values().reset_index(drop=Tr
ue))

    # Step 2: Initialize an empty list to store synchronized smiles
    synchronized_smiles = []

    # Step 3: Check for synchronized smiles within 1 second for both persons
    for _, row1 in person1_smile_events.iterrows():
        for _, row2 in person2_smile_events.iterrows():
            # Check if Person 2 starts smiling within 1 second after Person
1's smile starts
            if row2['Start'] >= row1['Start'] and row2['Start'] <= row1['End']
+ 1:
                synchronized_smiles.append((row1['Start'], row1['End'],
row2['Start'], row2['End'], row2['Start'] - row1['Start']))

            # Check if Person 1 starts smiling within 1 second after Person
2's smile starts
            elif row1['Start'] >= row2['Start'] and row1['Start'] <=
row2['End'] + 1:
                synchronized_smiles.append((row1['Start'], row1['End'],
row2['Start'], row2['End'], row1['Start'] - row2['Start']))

    # Step 4: Create a DataFrame with the synchronized smile events
    synchronized_smiles_df = pd.DataFrame(synchronized_smiles,
columns=['Person 1 Start', 'Person 1 End', 'Person 2 Start', 'Person 2 End',
'Phase Difference'])
```

```
    return synchronized_smiles_df

# Example of usage
# Assuming you have your dataframe 'df' with columns 'Person 1' and 'Person 2'
for timestamps

# Call the function to get synchronized smiles
synchronized_df = get_synchronized_smiles(combined_df, 'Person 1', 'Person 2')

# Display the synchronized smiles dataframe
print(synchronized_df)
synchronized_df.to_csv('SynchronizedSmiles_EXP4.csv', index=False)


# %%
def detect_smile_events(timestamps, threshold=0.1):
    """
    Function to detect smile events (start and end) from continuous
timestamps.

    Parameters:
    timestamps (list or pd.Series): List of smile timestamps for a person.
    threshold (float): The maximum gap between successive timestamps to
consider them part of the same smile.

    Returns:
    pd.DataFrame: DataFrame with 'Start' and 'End' times for each detected
smile event.
    """
    smile_events = []
    current_start = timestamps[0]

    for i in range(1, len(timestamps)):
        if timestamps[i] - timestamps[i - 1] > threshold:
            # End the current smile event
            smile_events.append((current_start, timestamps[i - 1]))
            # Start a new smile event
            current_start = timestamps[i]

    # Add the last smile event
    smile_events.append((current_start, timestamps.iloc[-1]))

    # Create a dataframe with start and end times
    return pd.DataFrame(smile_events, columns=['Start', 'End'])

# %%
def get_synchronized_smiles(df, person1_col, person2_col, conv_start,
conv_end, special_start=None, special_end=None):
    """
```

```
    Function to calculate synchronized smiles based on timestamps from two
persons, within a given conversation window.
    A synchronized smile is when one person smiles and the other smiles within
1 second after.

    Parameters:
    df (pd.DataFrame): Input dataframe containing timestamps for two persons.
    person1_col (str): Column name for Person 1's timestamps.
    person2_col (str): Column name for Person 2's timestamps.
    conv_start (float): Start time of the conversation (in seconds).
    conv_end (float): End time of the conversation (in seconds).
    special_start (float): Start time of the special period to exclude
(optional).
    special_end (float): End time of the special period to exclude (optional).

    Returns:
    pd.DataFrame: DataFrame containing synchronized smiles within the time
window, excluding special periods.
    """

    # Step 1: Detect smile events for both persons
    person1_smile_events =
detect_smile_events(df[person1_col].dropna().sort_values().reset_index(drop=Tr
ue))
    person2_smile_events =
detect_smile_events(df[person2_col].dropna().sort_values().reset_index(drop=Tr
ue))

    # Step 2: Filter smiles based on conversation start and end time
    person1_smile_events = person1_smile_events[
        (person1_smile_events['Start'] >= conv_start) &
(person1_smile_events['End'] <= conv_end)
    ]

    person2_smile_events = person2_smile_events[
        (person2_smile_events['Start'] >= conv_start) &
(person2_smile_events['End'] <= conv_end)
    ]

    # Step 3: If a special time period is given, exclude smiles within that
period
    if special_start is not None and special_end is not None:
        person1_smile_events = person1_smile_events[
            ~((person1_smile_events['Start'] >= special_start) &
(person1_smile_events['End'] <= special_end))
        ]

        person2_smile_events = person2_smile_events[
```

```
            ~((person2_smile_events['Start'] >= special_start) &
(person2_smile_events['End'] <= special_end))
        ]

    # Step 4: Initialize an empty list to store synchronized smiles
    synchronized_smiles = []

    # Step 5: Check for synchronized smiles within 1 second for both persons
    for _, row1 in person1_smile_events.iterrows():
        for _, row2 in person2_smile_events.iterrows():
            # Check if Person 2 starts smiling within 1 second after Person
1's smile starts
            if row2['Start'] >= row1['Start'] and row2['Start'] <= row1['End']
+ 1:
                synchronized_smiles.append((row1['Start'], row1['End'],
row2['Start'], row2['End'], row2['Start'] - row1['Start']))

            # Check if Person 1 starts smiling within 1 second after Person
2's smile starts
            elif row1['Start'] >= row2['Start'] and row1['Start'] <=
row2['End'] + 1:
                synchronized_smiles.append((row1['Start'], row1['End'],
row2['Start'], row2['End'], row1['Start'] - row2['Start']))

    # Step 6: Create a DataFrame with the synchronized smile events
    synchronized_smiles_df = pd.DataFrame(synchronized_smiles,
columns=['Person 1 Start', 'Person 1 End', 'Person 2 Start', 'Person 2 End',
'Phase Difference'])

    return synchronized_smiles_df


def calculate_synchrony_metrics(synchronized_smiles_df, conv_start, conv_end):
    """
    Calculate synchrony metrics: density, mean phase difference, standard
deviation, and kurtosis.

    Parameters:
    synchronized_smiles_df (pd.DataFrame): DataFrame containing synchronized
smiles and phase differences.
    conv_start (float): Start time of the conversation (in seconds).
    conv_end (float): End time of the conversation (in seconds).

    Returns:
    dict: A dictionary containing the calculated metrics.
    """

    # Ensure there are synchronized smiles to analyze
    if len(synchronized_smiles_df) == 0:
```

```
        return {
            'Density (smiles per minute)': 0,
            'Mean Phase Difference': np.nan,
            'Standard Deviation of Phase Differences': np.nan,
            'Kurtosis of Phase Differences': np.nan
        }

    # Total duration in minutes
    total_duration_minutes = (conv_end - conv_start) / 60

    # Density: Smiles per minute
    density = len(synchronized_smiles_df) / total_duration_minutes

    # Mean phase difference
    mean_phase_diff = synchronized_smiles_df['Phase Difference'].mean()

    # Standard deviation of phase differences
    std_phase_diff = synchronized_smiles_df['Phase Difference'].std()

    # Kurtosis of phase differences
    kurt_phase_diff = kurtosis(synchronized_smiles_df['Phase Difference'],
fisher=True)

    return {
        'Density (smiles per minute)': density,
        'Mean Phase Difference': mean_phase_diff,
        'Standard Deviation of Phase Differences': std_phase_diff,
        'Kurtosis of Phase Differences': kurt_phase_diff
    }


# %%

conv_start = 30  # Start of the conversation (in seconds)
conv_end = 500  # End of the conversation (in seconds)
special_start = 100  # Start of the special period to exclude
special_end = 150  # End of the special period to exclude

# Get synchronized smiles
df = get_synchronized_smiles(combined_df, 'Person 1', 'Person 2', conv_start,
conv_end, special_start, special_end)

# Calculate synchrony metrics
synchrony_metrics = calculate_synchrony_metrics(df, conv_start, conv_end)

# Display the metrics
print(synchrony_metrics)
```

For finding the synchronized head movements:

```
# %%
import pandas as pd
from scipy.signal import find_peaks
import numpy as np
from scipy.stats import kurtosis
import matplotlib.pyplot as plt

# %%
def load_csv_to_dataframe(csv_file):
    # Load the data into a pandas dataframe
    data = pd.read_csv(csv_file)
    print(f"Loaded data from {csv_file}")
    print("Column Names:", data.columns)  # Optional: To verify column names
    return data

# Function to smooth the head pose data from the dataframe
def smooth_head_pose_data(df, window_size=30):
    # Extract relevant head pose data (pitch, yaw, roll) from the dataframe
    pose_Rx = df[' pose_Rx']
    pose_Ry = df[' pose_Ry']
    pose_Rz = df[' pose_Rz']

    # Smooth the data using rolling window (moving average)
    smoothed_Rx = pose_Rx.rolling(window=window_size).mean()
    smoothed_Ry = pose_Ry.rolling(window=window_size).mean()
    smoothed_Rz = pose_Rz.rolling(window=window_size).mean()

    return smoothed_Rx, smoothed_Ry, smoothed_Rz

# Function to detect head movement peaks from smoothed data
def detect_peaks(smoothed_data, threshold_value=0.2):
    peaks, _ = find_peaks(smoothed_data, height=threshold_value)

    return peaks

# Function to calculate phase difference between two participants
def calculate_phase_differences(peaks_person1, peaks_person2):
    phase_diffs = [t2 - t1 for t1, t2 in zip(peaks_person1, peaks_person2) if
abs(t2 - t1) <= 1000]  # Limit to 1s

    return phase_diffs

# Function to calculate synchrony metrics from phase differences
def calculate_synchrony_metrics(phase_diffs):
    # Density of synchrony (events per minute)
    density = len(phase_diffs)
```

```
    # Mean phase difference
    mean_phase_diff = np.mean(phase_diffs)

    # Standard deviation of phase differences
    std_phase_diff = np.std(phase_diffs)

    # Kurtosis of phase differences
    kurt_phase_diff = kurtosis(phase_diffs)

    return density, mean_phase_diff, std_phase_diff, kurt_phase_diff

# Function to visualize head movements of both participants with highlighted
non-rapport period
def plot_head_movements_with_highlight(smoothed_data_person1,
smoothed_data_person2, timestamps,
                                       start_non_rapport, end_non_rapport,
title="Head Movements"):
    plt.figure(figsize=(40, 6))

    # Plot head movements for both participants
    plt.plot(timestamps, smoothed_data_person1, label='Actor', color='blue')
    plt.plot(timestamps, smoothed_data_person2, label='Participant',
color='red')

    # Highlight the non-rapport period
    plt.axvspan(start_non_rapport, end_non_rapport, color='yellow', alpha=0.3,
label='Non-Rapport Period')

    plt.title(title)
    plt.xlabel('Time (seconds)')
    plt.ylabel('Head Rotation (Radians)')
    plt.legend()
    plt.show()

# Function to calculate synchrony metrics for a specific period (normal or
non-rapport)
def calculate_metrics_for_period(smoothed_data_person1, smoothed_data_person2,
timestamps, start, end):
    # Get data for the specified period
    period_mask = (timestamps >= start) & (timestamps <= end)

    # Detect peaks for both participants in the specified period
    peaks_Rx_person1 = detect_peaks(smoothed_data_person1[period_mask])
    peaks_Rx_person2 = detect_peaks(smoothed_data_person2[period_mask])

    # Calculate phase differences for the specified period
    phase_diffs = calculate_phase_differences(peaks_Rx_person1,
peaks_Rx_person2)
```

```
    # Calculate the duration of the period in minutes
    period_duration_seconds = timestamps[period_mask].max() - 
timestamps[period_mask].min()
    period_duration_minutes = period_duration_seconds / 60  # Convert seconds 
to minutes

    # Calculate density as the number of synchrony events per minute
    density = len(phase_diffs) / period_duration_minutes if 
period_duration_minutes > 0 else 0

    # Calculate other synchrony metrics
    mean_phase_diff = np.mean(phase_diffs)
    std_phase_diff = np.std(phase_diffs)
    kurt_phase_diff = kurtosis(phase_diffs)

    return density, mean_phase_diff, std_phase_diff, kurt_phase_diff

# Main analysis function that takes dataframes as input and analyzes both 
periods
def analyze_conversation(df_person1, df_person2, start_non_rapport, 
end_non_rapport):
    # Smooth the head pose data for both participants
    smoothed_Rx_person1, smoothed_Ry_person1, smoothed_Rz_person1 = 
smooth_head_pose_data(df_person1)
    smoothed_Rx_person2, smoothed_Ry_person2, smoothed_Rz_person2 = 
smooth_head_pose_data(df_person2)

    # Extract timestamps
    timestamps = df_person1[' timestamp']

    # Visualize the Rx (Pitch) head pose movements with non-rapport period 
highlighted
    plot_head_movements_with_highlight(smoothed_Rx_person1, 
smoothed_Rx_person2, timestamps,
                                       start_non_rapport, end_non_rapport, 
title="Head Movements in Rx (Pitch)")

    # Visualize the Ry (Yaw) head pose movements with non-rapport period 
highlighted
    plot_head_movements_with_highlight(smoothed_Ry_person1, 
smoothed_Ry_person2, timestamps,
                                       start_non_rapport, end_non_rapport, 
title="Head Movements in Ry (Yaw)")

    # Visualize the Rz (Roll) head pose movements with non-rapport period 
highlighted
    plot_head_movements_with_highlight(smoothed_Rz_person1, 
smoothed_Rz_person2, timestamps,
```

```
                                       start_non_rapport, end_non_rapport,
title="Head Movements in Rz (Roll)")


    # Calculate synchrony metrics for the normal period (before non-rapport)
    density_normal, mean_phase_diff_normal, std_phase_diff_normal,
kurt_phase_diff_normal = calculate_metrics_for_period(
        smoothed_Rx_person1, smoothed_Rx_person2, timestamps,
timestamps.min(), start_non_rapport
    )

    # Calculate synchrony metrics for the non-rapport period
    density_non_rapport, mean_phase_diff_non_rapport,
std_phase_diff_non_rapport, kurt_phase_diff_non_rapport =
calculate_metrics_for_period(
        smoothed_Rx_person1, smoothed_Rx_person2, timestamps,
start_non_rapport, end_non_rapport
    )

    # Print the synchrony metrics for both periods
    print("Normal Period:")
    print(f"Density: {density_normal}, Mean Phase Difference:
{mean_phase_diff_normal}, Std Dev: {std_phase_diff_normal}, Kurtosis:
{kurt_phase_diff_normal}")

    print("Non-Rapport Period:")
    print(f"Density: {density_non_rapport}, Mean Phase Difference:
{mean_phase_diff_non_rapport}, Std Dev: {std_phase_diff_non_rapport},
Kurtosis: {kurt_phase_diff_non_rapport}")


# %%

start_non_rapport = 9*60+ 53
end_non_rapport = 11*60 +35
df_person1 = load_csv_to_dataframe("A_exp4.csv")
df_person2 = load_csv_to_dataframe("P_exp4.csv")
analyze_conversation(df_person1, df_person2, start_non_rapport,
end_non_rapport)

# %%
def detect_synchronized_points_with_timestamps(smoothed_data_person1,
smoothed_data_person2, timestamps, phase_threshold=5):
    # Detect peaks for both participants
    peaks_person1 = detect_peaks(smoothed_data_person1)
    peaks_person2 = detect_peaks(smoothed_data_person2)

    # Calculate phase differences
    phase_diffs = []
```

```
    sync_timestamps = []
    for peak1 in peaks_person1:
        # Find the closest peak in person 2's data
        closest_peak2 = min(peaks_person2, key=lambda x: abs(x - peak1))
        phase_diff = peak1 - closest_peak2
        phase_diffs.append(phase_diff)
        sync_timestamps.append(timestamps.iloc[peak1])

    # Prepare a dataframe to store the timestamps and phase differences
    sync_df = pd.DataFrame({
        'Timestamp': sync_timestamps,
        'Phase Difference': phase_diffs
    })

    # Identify synchronized points (based on phase difference threshold)
    synchronized_points = sync_df[abs(sync_df['Phase Difference']) <=
phase_threshold]

    return synchronized_points


def plot_all_synchronized_movements(sync_points_yaw, sync_points_pitch,
sync_points_roll, title="Synchronized Head Movements"):
    plt.figure(figsize=(40, 6))

    # Plot synchronized points for each axis
    plt.scatter(sync_points_yaw['Timestamp'], sync_points_yaw['Phase
Difference'], color='red', label='Yaw')
    plt.scatter(sync_points_pitch['Timestamp'], sync_points_pitch['Phase
Difference'], color='green', label='Pitch')
    plt.scatter(sync_points_roll['Timestamp'], sync_points_roll['Phase
Difference'], color='blue', label='Roll')

    plt.title(title)
    plt.xlabel('Time (seconds)')
    plt.ylabel('Phase Difference')
    plt.legend()
    plt.show()

def analyze_all_synchronized_movements(df_person1, df_person2,
phase_threshold=5):
    # Smooth the head pose data for both participants
    smoothed_Rx_person1, smoothed_Ry_person1, smoothed_Rz_person1 =
smooth_head_pose_data(df_person1)
    smoothed_Rx_person2, smoothed_Ry_person2, smoothed_Rz_person2 =
smooth_head_pose_data(df_person2)

    # Extract timestamps
    timestamps = df_person1[' timestamp']
```

```
    # Detect synchronized points for Yaw (Rx)
    synchronized_points_yaw =
detect_synchronized_points_with_timestamps(smoothed_Rx_person1,
smoothed_Rx_person2, timestamps, phase_threshold)

    # Detect synchronized points for Pitch (Ry)
    synchronized_points_pitch =
detect_synchronized_points_with_timestamps(smoothed_Ry_person1,
smoothed_Ry_person2, timestamps, phase_threshold)

    # Detect synchronized points for Roll (Rz)
    synchronized_points_roll =
detect_synchronized_points_with_timestamps(smoothed_Rz_person1,
smoothed_Rz_person2, timestamps, phase_threshold)

    # Plot synchronized points for all axes on a single graph
    plot_all_synchronized_movements(synchronized_points_yaw,
synchronized_points_pitch, synchronized_points_roll)

    # Combine the synchronized points into a single dataframe for analysis
    sync_points_combined_df =
pd.concat([synchronized_points_yaw.assign(Axis='Yaw'),
                                          synchronized_points_pitch.assign(Axis
='Pitch'),
                                          synchronized_points_roll.assign(Axis=
'Roll')], ignore_index=True)

    print(sync_points_combined_df)  # Display the combined dataframe

    return sync_points_combined_df

synchronyzed_movement = analyze_all_synchronized_movements(df_person1,
df_person2)
synchronyzed_movement.to_csv('Head_Movement_EXP4.csv', index=False)

# %%

# df_person1 = pd.read_csv('person1.csv')
# df_person2 = pd.read_csv('person2.csv')
synchronyzed_movement = analyze_all_synchronized_movements(df_person1,
df_person2)

# %%
synchronyzed_movement.to_csv('Head_Movement_EXP4.csv', index=False)

# %% [markdown]
```