Estimating attraction

by quantifying bodily coordination using wearable sensors

Ö. Kapcak

Estimating attraction

by quantifying bodily coordination using wearable sensors

by

Ö. Kapcak

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Tuesday July 30, 2019 at 14:00.

Student number:	4744608	
Project duration:	November 12, 2018 – July 30, 2019	
Thesis committee:	Assoc. Prof. dr. ir. H. Hung	TU Delft, supervisor
	Prof. dr. ir. M.J.T. Reinders	TU Delft
	Dr. Odette Scharenborg	TU Delft
	Jose Vargas-Quiros	TU Delft, daily supervisor
	Stephanie Tan	TU Delft

This thesis is confidential and cannot be made public until July 30, 2019.

Abstract

In this thesis, prediction of romantic, social and sexual attraction between two people using bodily coordination features is studied. Attraction is one type of interest that can occur between interacting people and understanding the modeling of it can help understand how to model other types of interests in human-human interactions because similar methods can be used to model the non-verbal behavior that reveals interest.

Previous research in psychology and social signal processing fields showed that synchrony and convergence of both audio features and body movements of people during an interaction are indicators of interest. However, audio features require recording people's voices during interactions and it can be disturbing for people especially during more personal conversations. In addition, capturing nonverbal cues from video mostly requires recording people with a camera from front and it might make people more aware of being recorded and intervene with the naturality of the interaction. Moreover, processing the videos to extract specific nonverbal behavior can be costly. Based on these ideas, we decided to use motion channel and hypothesize that movement synchrony and convergence features can be used to automatically quantify and predict attraction. We propose a novel method of estimating romantic, social and sexual attraction between two people by quantifying their bodily coordination using wearable sensors in a speed-date setting. We developed simple synchrony and convergence features, inspired from the literature and specifically adapted to be extracted from accelerometer data. To our knowledge, this is the first time that motion convergence is used for estimating attraction.

Our features could predict one-way social attraction with a 73% Area under the ROC curve (AUC), outperforming previous work in a similar setting. We also showed that prediction performance increased when the male and female data are separated, aligning with the theories in psychology studies. We could also predict mutual romantic attraction with an AUC of 80%. We found that different types of attraction can be estimated better using different feature types, more specifically we could predict social attraction better using movement correlation features whereas for romantic and sexual interest mimicry features were better indicators. Moreover, features extracted from different types of signals recorded from accelerometers showed varying performances for different attraction types. Additionally, asymmetric features out-performed the symmetric features and our synchrony features showed better performance than convergence features. Finally, we a have seen that motion convergence can occur in people having an interaction regardless of attraction.

Preface

Human mind and human behavior have been always an interest of mine. After completing my bachelors in computer science and minor in psychology, I knew that I wanted to combine both of my interests during my masters. I think it is fascinating how we can use computers to find patterns in human behavior in order to understand the human mind better. Even though it is slightly creepy, I believe there is much more that we can learn about ourselves with the help of emerging technologies. I feel very fortunate that I have been able to conduct research in a field that I am very enthusiastic about.

Before you lies my thesis. This thesis has been written to complete my Master's Degree in Computer Science at Delft University of Technology, and is written in such a way that it should be understandable for anyone with a university degree in engineering. I hope you will enjoy your reading!

Ö. Kapcak
Delft, July 2019

Acknowledgements

First, I would like to thank Hayley for giving me the opportunity to perform my master thesis under her supervision. She always challenged my ideas, pushed me to look at things from different perspectives, think critically, and pushed me to own my research and to find my own answers. Telling me "well done for getting this far" is what kept me going for 9 months.

Next, I want to thank Jose for his guidance. He provided very useful feedback at all times, was patient with me when discussing things, showed great interest in the work and always challenged me to carry my research further.

Last, I would like to thank my friends, Neslihan, Dilay, Güneş and Beliz who are kilometers away but whom I could always count on for support. Then, I would like to thank Patrick for being my number one support in Delft. Finally, I would like to thank my parents and my sister, Ezgi, without them none of this would be possible.

Contents

1
Introduction

The advances in sensing technologies and the possibilities of sensing human behavior have brought interest in the automatic assessment of human behavior in several research communities. The field, Social Signal Processing, aims to sense and understand signals from humans during interactions using computers [36]. The importance of this field is that, assessing human behavior makes it possible to automatically analyze human-human interactions. This in turn makes possible to build tools that improve the time and possibly quality of psychological and sociological research. Additionally, automatic assessment is of interest for the creation of more naturally behaving socially-aware computing systems. Understanding the dynamics of human interactions and being able to automatically quantify them makes it possible to improve the quality of interaction between virtual agents and humans. A further application is the creation of tools that can help people assess their own behavior in their relationships, enabling them to receive feedback about behavior during social interactions which would increase the quality of their relationships with other people.

Recent promising advances in this field give insights into the relationship between little-understood phenomena like physical and emotional attraction and measurable human behavior. Attraction has been found to affect the way in which couples behave towards one another during interactions, affecting other known social phenomena like the level of synchrony in their movements [44], the degree to which they mimic one another [12, 19, 25] and the adaptation to one another's behavior [28]. This study aims to investigate how we can automatically estimate interpersonal attraction by quantifying body coordination using wearable sensors.

The goal of this study is to investigate the automatic detection of attraction in dyadic interactions using movement features that are automatically extracted from single body-worn accelerometers in an in-the-wild setting. Understanding the modeling of attraction can help understand how to model other types of interests in human-human interactions because similar methods can be used to model the non-verbal behavior that reveals interest. The reason to study non-verbal behavior is because it is argued to be less deceptive than verbal behavior since people cannot hold honest information back [11].

In the rest of this chapter, the research goals are presented and an elaboration on the main contributions is given. Subsequently, an outline of the rest of the thesis is given.

1.1. Research goals

The main research goal of this thesis, is to develop a methodology for creating a better understanding of attraction in dyadic interactions. The hypothesis is that in dyadic interactions, behavioral coordination patterns of participants can be automatically captured and used to predict different types of attraction. The focus of this work is on predicting attraction in dyadic relations, using features that were automatically extracted from accelerometer data recorded by bodily worn sensors. The first subgoal of this research is to define behavioral coordination features that are inspired from the literature, that can be extracted from single bodily worn accelerometer. The second subgoal is to examine whether there is a difference in behavioral features that are indicators of different types of attraction, such as romantic, social and sexual.

1.2. Contributions

In this thesis, it is hypothesized that interpersonal coordination of behavior in dyadic interactions can serve as an indicator for interpersonal attraction. The contributions of this research involve developing simple synchrony and convergence features, inspired from the literature and specifically adapting them to be extracted from accelerometer data. Concretely, the contributions of this study are three-fold as explained below:

Quantifying synchrony and convergence using single bodily worn sensors

Even though there is a significant amount of work that focuses on extracting nonverbal cues from audio and video during human interactions, motion channel is understudied. Moreover, the studies that worked with motion channel mostly used multiple sensors on body which is not feasible in real world scenarios. In addition, extracting body movements from video data is a costly process and using audio channel can be seen as intervening with the privacy. In this thesis, we proposed using a single bodily-worn accelerometer as data source and developed novel behavioral coordination features such as synchrony and convergence that can be extracted from that source.

Predicting attraction using bodily coordination features

The majority of previous literature on attraction prediction targets audio cues and movement features extracted from video channel. We used the features extracted from motion channel that capture bodily coordination to model interpersonal attraction and tested them in a real life in-the-wild setting with a less intrusive approach. We show that these features can be used to predict attraction between two people. Even though motion synchrony is used for estimating attraction before, to our knowledge this is the first time that motion convergence is used for this task.

Verification of psychological theories about gender differences in attraction

We obtained experimental evidence that supports the existing theories from psychological literature about behavioral differences between men and women in a courtship setting.

1.3. Research Context

This study was performed independently at Delft University of Technology Socially Perceptive Computing Lab and can be interpreted as a subset of the larger MatchNMingle project. More information about the dataset that was used can be found in the MatchNMingle project website.

1.4. Outline

The rest of the thesis is organized as follows:

1. In Chapter 2, the related work is reviewed to provide more in-depth information on the metrics and behavioral cues used for quantifying and estimating different types of interpersonal interest.

2. In Chapter 3, the data collected for this research is discussed by giving information about the MatchNMingle dataset. The methodology of extracting the ground truth is explained and statistics over the obtained datasets are discussed.

3. In Chapter 4, the methodology of this thesis is explained. First, the initial pre-processing steps applied to raw accelerometer data are explained. Then, the motivations and the methodology for extracting the features are explained and discussed. Then, the overview of final pre-processing steps is given. Finally, the setup and the process of classification is explained.

4. In Chapter 5, the results of this study are presented. Overall prediction performance is given and compared with the state-of-the-art baseline features. Then, a more thorough examination on the relative contribution of each feature type is provided and analysis on different aspects of features, such as window length, feature category, is made. Finally, characteristics of motion convergence is investigated further.

5. In Chapter 6, concluding remarks are given and findings are discussed. The limitations of the methodology are discussed and suggestions for future research is given.

2

Automatic quantification of interpersonal interest using behavioral signals: A literature review

Gatica-Perez defines the term interest as "people's internal states related to the degree of engagement displayed, consciously or not, during social interaction" [15]. He also notes that this engagement arises because of different factors such as interest in the topic of a conversation, attraction to another person or social rapport, which means having a connection with someone else. This internal state of interest is also related to other cognitive states such as creativity and learning. Even though the focus of this thesis is the attraction in dyadic interactions, the research about other forms of interest is worth considering for related work because similar methods can be used to capture and model the non-verbal behavior that reveals interest. Interpersonal interest and its associated nonverbal behavior have been studied by psychology researchers. The automatic quantification of this behavior has also been of interest to computer scientists. Therefore, studies from both fields are reviewed here. This chapter provides a review of prior work done in modeling and prediction of interest in human interactions. First, initially the existing work over detection and recognition of romantic attraction is reviewed. After that, other research that studied different forms of interest in (mostly) dyadic interactions is reviewed. Finally, the concluding remarks are given.

2.1. Attraction

Most of the existing work that studied attraction conducted experiments in speed-date scenarios. The purpose of speed dating is to meet a large number of potential partners in a short time. One by one each male and female participant that joined the event have a short date with the opposite sex to meet with each other. After each date, they rate each other possibly on a questionnaire and continue with the next date. At the end of the event, if both people rated each other positively, a match occurs and their contact information is given to each other. The reason for using these events in attraction studies is that the responses to the questionnaires can be used as the ground truth for prediction tasks. Most of these work used features extracted from audio and video. Here, each modality is discussed separately.

In their experiments, [26, 28, 39] extracted audio features from participants' audio recordings. Madan et al. [26] extracted four types of measures from audio: activity, engagement, stress and mirroring to predict different types of attraction (romantically attracted, interested in friendship, or interested in business) between speed-date partners. They used a two-class linear classifier for each type of task and could predict the romantic attraction with an accuracy of 71% . To measure the activity, they initially trained a two stage Hidden Markov Model (HMM) to separate first the voiced segments from non-voiced segments and then within voiced segments into speaking and non-speaking segments. Conversational activity level was measured then by the z-scored percentage of speaking time plus the frequency of the voiced segments. The engagement measure aims to capture the influence that each person has on the other's turn-taking by modeling their individual turn-taking by an HMM and measure the coupling of these two dynamic systems. They measured the stress by looking at the variation in prosodic emphasis. The stress mentioned here can be either on purpose such as emphasising words, or unintentional which can be caused by discomfort or physiological stress. Finally, mirroring was measured as z-scored frequency of short utterances exchanges during the conversation. These short utterances can be short interjections such as 'uh-huh' or single-word back-and-forth exchanges such as 'OK?', 'OK!', 'Right?', 'Yeah!'.

Ranganath et al. [39] focused mainly on prosodic, dialogue, and lexical features and instead of predicting attraction, they worked on predicting flirtation intention which they measured by asking the participants how often they were flirting during their dates on a scale of 1-10. They also asked about how often the other person was flirting to measure the perception of flirting. They developed a system that can predict the intention of flirting better than humans do at a speed dating setting. The answers to the questions that they asked to the participants served as the labels for perception and intention of flirting and they ran a binary classification experiment to predict them using the features they extracted from wavefiles and transcripts of the conversations. The features they used are grouped as prosodic, dialogue and lexical features. The prosodic features were F0 and RMS amplitude features. As for dialogue features they used the number of turns, laughter, disfluencies or restarts, and overlaps which the two speakers spoke at the same time. Laughter, disfluencies and overlaps were all marked by the transcribers. As for the lexical features, they used features extracted by an autoencoder. For classification they used a SVM and could detect flirt intention of women with 71.5% and men with 69% accuracy, both scores were better than human's perception of flirtation.

Michalsky et al. [28] investigated pitch convergence and examined whether speakers' speech similarity increases during the conversation depending on perceived attractiveness and/or likability. They also conducted a speed-date experiment and found that speakers became more similar over the course of conversation. Furthermore, the degree of pitch convergence was shown to be related with the degree of perceived attractiveness and likability. They investigated the convergence both globally by comparing the first third and the last third of each conversation, and also locally on a turnwise level. They applied an acoustic analysis to the first and the last third of every conversation for global convergence. For local convergence they first segmented the speech parts of all speakers into interpausal units, then they used two interpausal units that are adjacent to a turn break with a speaker transition for the acoustic analysis. To apply acoustic analysis they extracted F0 features from the respective speech parts from both speakers and calculated the differences between these features of both speakers. In order to test the effects of convergence, they conducted a statistical analysis with linear mixed effects models. Their results show that global and local convergence do occur for most of the investigated features regardless of perceived attractiveness and likability when two people are having an interaction. In addition to that, they found that both convergence features were affected by both perceived attractiveness and likability.

Veenstra et al. [44] used video data from a speed-date event and extracted positional features such as position, distance, movement and synchrony. After each date, participants are asked questions regarding if they want to exchange contact information and also about how physically attracted they are to other person. As a result they obtained high accuracies for predicting them using the features extracted from video. From the recordings of top-down cameras recorded during the event, they extracted the positional information of the participants. From these information they extracted the following features: position, distance, movement and synchrony. For position, they used the difference in the angle both persons have with the table to learn about how the people are positioned with respect to each other. For distance, they used multiple features which are the difference between the average euclidean distance in the first n frames and the last n frames, how often someone moves in a particular direction, variance in position and variance in distance. They computed these movement features (except decrease in distance) both for the person of interest and also for the person he/she was dating. For synchrony features, they extracted synchrony in motion and distribution of motion reaction. In order to extract synchrony in motion, they first calculated the amount of motion per second by looking at the difference in a person's position between consecutive frames and accumulating this over a window of 1 second. Then they made a histogram with four bins, one that counts the amount of low activity of both people, two bins that count the seconds of high activity in one person and low activity in the other and last one that counts the seconds with high activity by both people. To compute the distribution of motion reaction, they investigate how they react to each other by looking at how the distance with a previous position of the other varies. In their experiments they predicted if a participant wanted to exchange contact information with the other and also the physical attraction to each other using SVM and kNN classifiers by separating the male and female data. Some of their features outperformed the baseline and they also found that different features performed better for male and female classification tasks.

Cabrera-Quiros et al. [5] has also used a speed date data set and attempted to classify attraction levels between participants by using features inspired by [44] and extracting them from accelerometer data instead of video. Even though they could only replicate some of the movement based features, they obtained good classification results by using a logistic regressor and outperformed the random baseline. They extracted the mean and variance of the magnitude of the acceleration for each date and also calculated the variance over a 1s sliding window with a shift of 0.5s. From this variance over a sliding window, they extracted the mean and the variance and obtained 4 basic features. They also applied the method of separating the male and female data but in their case it did not improve the results.

Crown et al. [8] also studied interpersonal attraction but they conducted an experiment with pairs of women in one of the three conditions of 'like', 'dislike', and 'unacquainted'. They investigated the interpersonal coordination of vocal and visual timing during the conversations of participants in these three situations. As a result, they found that different types of attraction can be differentiated by the differences in behavior coordination.

The research reviewed in this chapter is summarized in Table 2.1. In summary, speed-dates are common settings that are used in the experiments that aimed to predict interpersonal attraction because of the ease of extracting ground truth. Another common point of these studies is using binary classification or labeling for attraction. Even though the questionnaires they used are on rating scales, they are converted into binary labels. In most of the studies, audio features are used for predicting attraction with one exception of video modality and another one with motion modality. Synchrony in movement is extracted from video and used as a feature for modeling attraction but to our knowledge there is not a study that aimed to measure synchrony in motion using accelerometers to model interpersonal attraction. Moreover, convergence was only studied with audio features and there was not a study that aimed to extract motion convergence from either video or accelerometer data.

Table 2.1: Summary of research on automatic detection of attraction using nonverbal behavioral features

Ref.	**Task**	**Modality**	**Measures**
[26]	Predicting attraction	Audio	Activity level, engagement, stress, and mirroring in nonverbal audio
[39]	Predicting flirtation intention	Audio	Prosodic, dialogue, and lexical features
[44]	Predicting contact information exchange and physical attraction	Video	Position, distance, movement, synchrony
[5]	Predicting attraction	Motion	Magnitude and variance of acceleration
[28]	Examining the relation between pitch convergence and attraction	Audio	Pitch convergence
[8]	Differentiating interpersonal attraction by visual and vocal coordination	Audio, video	Temporal coordination of the visual and vocal behaviours

2.2. Other forms of interest

Automatic estimation of different forms of interest during interpersonal interactions other than attraction have also been studied by researchers. These studies can be grouped as estimating internal states of the people during interactions, predicting the outcomes the interactions and examining the link between nonverbal behavior and interaction context. Here we discussed research over each category separately.

2.2.1. Estimating internal states during interactions from nonverbal behavior

Xiao and collaborators [47, 48] focused on the head motion synchrony in dyadic interactions of spouses. They examined the link between head motion of interacting couples and their behavioral characteristics by conducting experiments using video recordings of communication sessions from real couples during a couples therapy study. They modeled the head motion using Gaussian Mixture Model (GMM) of line spectral frequencies extracted from the motion vectors of the head and quantified the similarity of head motion of couples by computing the Kullback-Leibler divergence of the GMM posteriors of their respective motion sequences. One of their findings is that the degree of synchrony between people's head motion increases as the interaction progresses. They tested this by comparing the first and second halves of the interaction. Additionally the interactions were annotated by expert judgements and specific behavioral characteristics of the couples are binarized as Acceptance, Blame, Positive, and Negative behavior. As a result of their study, they showed that the relative change of head motion similarity correlates with these behavioral characteristics and they could classify them by using the head motion models. They also found that spouses having positive affects showed increasing degree of synchrony in head motion along the interaction, whereas spouses having negative affect showed a decreasing degree of synchrony.

Other existing work addressed the problem of engagement prediction in dyadic interactions. Hsiao et al. [21] conducted a research over social engagement in dyadic conversations using microphones on smart phones to collect humans' speech behavior. They extracted 3-levels of features from audio data and used Coupled Hidden Markov Model (CHMM) and K-means algorithm to recognize patterns in audio features. By observing the patterns of turn-taking and speech emotion during a face-to-face conversation, they could classify the level of engagement of participants as "high" and "low" with a accuracy of ~79%. Huang et al. [22] used visual cues additional to audio cues to again recognize engagement levels in face-to-face dyadic interactions. They used facial information and low-level image features and also low-level auditory features to recognize engagement levels during conversations. From video data they extracted both texture features using local binary patterns (LBP) with principal component analysis (PCA) and also geometric features such as facial landmarks and head pose. From audio data they extracted low-level acoustic features like pitch level, MFCCs, and loudness, and also extracted shape and angle features from loudness curve. By using Convolutional Neural Networks (CNN) they could classify engagement in 4-levels such as Disengagement, Nature, Engagement and Strong Engagement with high accuracies. Oertel et al. [33] attempted to predict involvement levels in dyadic interactions using multimodal visual and audio cues. As for visual cues they used manually annotated mutual gaze and blinking, and for audio cues they extracted acoustic features such as pitch level and intensity. They annotated involvement levels on a scale of 0-10 but then grouped them into two or three levels to use for prediction. By using SVM with their audiovisual features, they could classify the involvement levels with an accuracy of 68%.

Recognition of emotions during dyadic interactions using audio-visual data is also studied in recent research [29, 49]. Both studies combined features extracted from body movements and speech. Muller et al. [29] focused specifically on bodily expressions. They used a dataset that was recorded in a realistic environment and contains dyadic interactions. On video recordings they used dense trajectories and body part detection for human activity recognition and from audio recordings they extracted low level audio features. Using these features separately, they attempted to classify four annotated emotions (anger, happiness, sadness, surprise) during interactions and established baseline performances. Yang et al. [49] focused specifically on modeling the mutual influence of multimodal behavior in dyadic interactions with the idea that modeling such influence is important for emotion detection in an interaction. In their dataset the emotional state of each participant was annotated as excited vs. calm and positive vs. negative. They modeled the behavior adaptation from the interlocutor to the target participant during an interaction using an interaction matrix which gathers all the behavioral information of the dyad. As a result they found that interaction patterns are dependent on the emotional states of interaction partners and their approach could significantly improve the performance of emotion recognition.

Other existing research studied the recognition of specific affective states in dyadic interactions such as enjoyment [40], prejudice [35] and agreement [2]. Sandstrom et al. [40] conducted a study using mobile phones to capture information about people's social interactions during their daily lives. They collected the data over conversations of participants that engaged during their daily lives. They used audio data recorded by participants' phones and extracted conversational features such as conversation length, rate of turn taking, proportion of speaking time and acoustic features such as volume and pitch. Using these features they could predict how much a person enjoys the conversation from these conversational properties and one interesting finding of them is that people enjoyed their conversations more when they spoke a smaller proportion of the time than usual. In an interesting study by Palazzi et al. [35] they aimed to capture and recognize prejudice towards black people from nonverbal behaviours. They extracted spatial (mutual distance, space between interlocutors, participants' movements), audio (pauses in dialogue), and biometric (related to heart rate and emotional arousal) features to measure their correlation with psychological scores of prejudice. Specifically they used Microsoft Kinect to capture the movements and Shimmer GSR to estimate heart rate and galvanic skin response. Participants of their experiments interacted with white and black people about different topics and then filled questionnaires. After that a team of social psychologists analysed these questionnaires to summarize participants' prejudice which constitutes the ground truth for their study. Using the features they extracted, they aimed to predict the prejudice scores and obtained promising results indicating their approach can be used for automatically identifying prejudice with nonverbal behavior.

Bousmalis et al. [2] targeted the recognition of spontaneous agreement and disagreement using nonverbal cues during a conversation. They automatically extracted nonverbal auditory features (fundamental frequency and energy) but they manually annotated the hand actions, head and body gestures. They proposed to use Hidden Conditional Random Fields (HCRFs) which is a dynamic discriminative model, to classify spontaneous agreement and disagreement and it outperformed SVMs and HMMs in this task, possibly because of its ability to model the hidden fine–grain dynamics of the multimodal cues. In addition, they showed that HCRFs can be automatically analysed to identify which features are the most discriminative in each class.

A different modality other than audio, video and motion that can be used to extract features while modeling human-human interactions is the physiological responses. There has been a body of work that used physiological features of people in dyadic settings to predict affective states during interactions [6, 7, 42]. Chaspari et al. [6, 7] aimed to capture the synchrony between Electrodermal Activity (EDA) streams of partners that occurs in parallel. EDA is a measure of increased skin conductivity from sweat secretion and is related to different psychological states and traits, such as how secure versus anxious people feel in their interpersonal relationships. They worked on capturing the similarity between two persons' EDA signals and proposed a synchrony measure. As a result they found that this measure shows different synchrony patterns during tasks of different emotional intensities given to the couples and also is associated with couples' attachment styles. Timmons et al. [42] made use of different modalities and a more extended set of features. They aimed to monitor problematic relationship dynamics and detect conflict in couples during their daily lives using wearable technologies. In their study, they made young-adult couples wear biosensors and measured their electrodermal and electrocardiographic activity, physical activity, and body temperature and carry smartphones that collected audio recordings and GPS coordinates for one day. Additionally, participants were asked to complete hourly self-reports on their general mood states and the quality of their interactions. Using the method from [6] they computed EDA synchrony and used other physiological measures as they are. From audio recordings they extracted linguistic and acoustic features. By using all these features they could classify the self-reported mood and quality of interaction of the couples.

Even though they are not dyadic interactions, group meeting settings are studied for detecting social cohesion in literature as well. Since cohesion can also be considered as a form of interest, research on that topic is also reviewed here. Hung et al. [23] investigated automatic estimation of cohesion in task based group meetings using automatically extracted audio, visual, and audio-visual cues. The audio cues they used can be summarized as periods between each individual's turns, times between floor exchanges, turn durations, overlapping speech, and prosody. As visual cues, they aimed to capture the amount of visual activity of each person. For audio-visual cues they looked at the visual activity during speaking and non-speaking times and also proposed two synchrony measures by combining audio and visual cues that capture self-synchrony and interpersonal synchrony. They obtained 90% accuracy with audio cues which performed the best, followed by video cues achieving 83% and audio-visual cues 82% accuracy indicating that these automatically extracted behavioral cues can be used to estimate perceived levels of cohesion in meetings. Nanninga et al. [31] also studied estimation of task and social cohesion in group meetings but they used a more specific set of features and focused on mimicry behavior. They proposed synchrony and convergence features that can be extracted from audio data to quantify the dynamic alignment of paralinguistic speech behavior. They could estimate social cohesion with 71% AUC ROC score and task cohesion with 64% AUC ROC score.

Studies that are discussed in this subsection show that audio and visual features can be used together for modeling patterns that occurs during interpersonal interactions. These features can carry more information than features extracted from a single channel. In addition, physiological responses also reveal information about people's internal states during interactions and can be used for modelling them. However, using the devices that capture these responses can intervene with the naturality of the interactions. One other interesting device that is used in these studies is Microsoft Kinect which capture the movements of people which can be an alternative to camera recordings and video processing.

2.2.2. Predicting outcomes of the interactions from nonverbal behavior

Unlike recognizing affective states, other existing works studied predicting specific outcomes of dyadic interactions. Two studies by Won et al. automatically captured body movements of pairs using Kinect cameras while they are completing given tasks to predict outcomes of these tasks such as learning [45] and creativity [46]. Using the data over the movements of joints recorded by Kinect, they created features that capture the aspects of the movements of the participants during the interactions. [45] used these features to investigate the role of nonverbal behavior in teaching and learning and classified the interactions as successful or unsuccessful learning based on the scores of the tests participants filled after. One limitation of the study is that the participants were not real teachers and students, and the experiment took place in a lab setting which possibly does not resemble a real learning setting and teaching interaction. [46] extracted the same features from Kinect data but they carried the measures one step further and computed a synchrony score by correlating movements between the two dyad members. They found a link between this nonverbal synchrony and creativity which they quantified as the number of new, valid ideas produced during the interaction. In both of the mentioned studies the metrics that capture the outcomes (learning and creativity) were imperfect and could be affected by the background of the participants.

Hirability prediction from nonverbal behavior and job interview interactions are also studied previously [30, 32, 34]. The experiments are conducted at job interview settings which includes one applicant and one or more interviewers and audio-visual features are extracted. Okada et al. [34] first extracted nonverbal behavioral features of both participants including speaking status and head nods and then identified the inter-modal and inter-person patterns by looking at frequent co-occuring events. Their proposed framework captures how one of the interactors generates nonverbal behavior when other interactors also generate nonverbal behavior. They used these features to successfully predict personality traits and hirability decisions. Behavioral features that Nguyen et al. [32] extracted include audio features such as speaking status and prosody, and visual features such as head nods, visual motion, smiling and gaze. Additionally, they used multimodal and relational features such as "nodding while speaking" and "mutual nodding". Using these features with regression models, they could predict hirability scores. In the study of Naim et al. [30], different than the previous ones they only observed the behavior of the interviewee. Moreover, they used verbal behavior additional to nonverbal behavioral cues. They extracted features from facial expressions such as smiles and head gestures, language such as word count and topic modeling and finally prosodic information such as pitch and pauses of the interviewees. Unlike the previous studies, they used these features to predict attributes such as excitement, friendliness, and engagement which are the skills that are important in interview settings.

As is seen, modeling nonverbal behavior can also be applied to the tasks that aim to predict specific outcomes of dyadic interactions. Similar to the previous studies, these research also used multimodal features that are extracted using both audio and video recordings. One important point to be considered for these studies is that they are mostly done in lab settings which are supposed to resemble the real world scenarios. However, this situation decreases their ecological validity. Therefore, the findings of the experiments should be tested in real life settings to ensure their validity.

2.2.3. Examining the link between nonverbal behavior and interaction context

Hart et al. [20] conducted a study to automatically measure and interpret non-verbal characteristics of doctor-patient interactions. In their experiments, an actor played as a doctor and interviewed the subjects performing two different scripts where in one he showed mimimal engagement to the subject and in other he listened actively and payed attention to the subject. They extracted motion energy from video and analyzed the cross correlation in total kinetic energy of patient and doctor to evaluate their synchrony and followership patterns. They found large differences in these measures between the two performance scenarios. In the active listening scenarios they observed more synchrony and more symmetric followership than the other scenario.

Yang et al. [50] investigated the adaptation of body language of participants to other person in a dyadic interaction given different interaction goals and context. They extracted body language features of the interlocutors from video such including body motion, posture and relative orientation and examined the degree of correlation between two people's body language in different interaction contexts(friendly vs conflictive). They performed Canonical Correlation Analysis (CCA) between target participant and the interlocutor's body language features and found that the nature and the goal of the interaction influences the coordination pattern of a dyad's behavior. Moreover, they proposed a method that combines Gaussian Mixture Model (GMM) based statistical mapping, and Fisher kernels, to predict a person's body language from the multimodal information of the other person during an interaction for specific interaction goals and obtained promising results.

Feese et al. [13] conducted a research on leadership behaviour and extracted nonverbal cues from body motion to understand if individually considerate and autocratic leaders behave differently. They used wearable motion sensors and with activity recognition methods they detected the relevant nonverbal cues such as face touch, arm positions and movements, posture, hand gestures and nodding. Further, they quantified behavioral mimicry between interacting partners using these detected cues. They computed the mimicry events as when person B displays the same nonverbal cue within a certain time after person A did. As a result, they found that individually considerate leaders moved less, mimicked the other person's cues such as nodding, face-touch and posture changes more often than the authoritarian leaders. Additionally, followers of individually considerate leaders mimicked their leader's face touch more often.

Two different studies used the same method for quantifying the nonverbal synchrony of interactants during a dyadic conversation [38, 43]. They extracted the body movements using an automated objective video analysis algorithm (Motion Energy Analysis; MEA) and to quantify the synchrony they computed the cross-correlations of participants' movement time-series with positive and negative time-lags. Ramseyer et al. [38] studied nonverbal synchrony between patients and therapists during psychotherapy sessions and found that nonverbal synchrony was positively correlated with the patient's rating of the therapeutic bond and also the therapy outcome. Tschacher et al. [43] conducted experiments on dyadic interactions where participants interact in cooperative, competitive, and 'fun task' conditions and examined the link between interactants' affectivity and their nonverbal synchrony. They found the positive affect was associated positively with synchrony whereas negative affect was associated negatively. They also found that this link between nonverbal synchrony and affect was strongest in female dyads.

The studies that are reviewed in this subsection emphasize the methods of extracting body movements for estimating the context of interpersonal interactions. Most of these studies point out the importance of body movement correlation and also behavioral mimicry in order to model the patterns of human interactions. As is seen, motion sensors can be used for activity recognition as an alternative to video recordings. However, their feasibility and practicality can differ depending on the setting and the context of the experiments.

The research reviewed in this chapter is summarized in Tables 2.2, 2.3, 2.4, 2.5 according to the modalities they used.

2.3. Concluding Remarks

In this chapter, the research over modeling of interest from multimodal nonverbal behavior is reviewed. As is seen, nonverbal cues reveal information about the content and quality of human-human interactions. Therefore, automatically extracting these cues and using them with machine learning methods can be used for prediction and detection of outcomes of the dyadic interactions.

Even though some of the studies mentioned here conducted the experiments and collected the data during in-the-wild settings, most of them are conducted in experiment rooms which are unnatural settings. In these settings, even though the conditions are resembled to real world as much as possible, it still decreases the ecological validity because people will possibly act differently than how they do in real situations. Moreover, in some cases participants are given pre-written scenarios that they needed to follow which decreases the naturality of the context even more. Another issue regarding the studies in the existing literature is the unfeasibility of the methods that they used for the detection of nonverbal cues. Some studies mentioned above used human annotators to label nonverbal actions such as hand gestures and mimicry behavior which can be very time consuming and also inaccurate and biased due to judgement of each annotator.

The most notable features used for interest modeling that appeared in the literature are synchrony and convergence of both audio features and body movements. One issue is that, even though the audio features can be very useful for prediction of affective states during dyadic interactions, people might not be very comfortable with their voice being recorded especially during a speed-date setting where sensitive information can be revealed. Specific nonverbal cues such as head gestures and gaze are also shown to be informative for interest modeling during interactions. However in most cases it requires the recording of the people with a camera from front which will make people more aware of being recorded and cause them to control their movements, thus intervene with the naturality of the interaction.

As a conclusion, in order to conduct a study over the modeling of interest and especially over the estimation of attraction, the experiments should be conducted at an in-the-wild setting and participants should be minimally aware that they are being recorded and also not be worried about the privacy of the conversation. Additionally, the feature extraction should be feasible and less prone to human bias during annotation.

Table 2.2: Research on automatic detection of interest using nonverbal behavioral features extracted from audio

Ref.	**Task**	**Measures**
[21]	Classifying engagement levels	Turn-taking and speech emotion
[31]	Estimation of task and social cohesion in group meetings	Mimicry in speech behavior
[40]	Predicting enjoyment during conversations	Conversational features such as conversation length, rate of turn taking, proportion of speaking time and acoustic features such as volume and pitch

Table 2.3: Research on automatic detection of interest using nonverbal behavioral features extracted from video

Ref.	**Task**	**Measures**
[50]	Relation between body language mimicry and interaction context	Coordination between body language features
[38]	Examining the link between therapy outcome and nonverbal synchrony	Correlation between body movements
[43]	Examining the link between affectivity and nonverbal synchrony	Correlation between body movements
[47, 48]	Examining the link between head motion of interacting couples and behavioral characteristics	Head motion synchrony
[20]	Recognizing dyadic affects in a medical setting	Cross correlation in total kinetic energy
[45]	Predicting learning in dyadic interactions	Individual's joint movements
[46]	Predicting the creativity in collaborating dyads	Correlation between individual's joint movements

Table 2.4: Research on automatic detection of interest using nonverbal behavioral features extracted from audio and video

Ref.	**Task**	**Measures**
[34]	Predicting personality traits and hirability decisions during job interviews	Co-occurrence patterns between modalities (speaking status and head nods)
[32]	Predicting hirability during job interviews	Speaking status and prosody, head nods, visual motion, smiling and gaze
[30]	Predicting attributes of social interactions in a job interview setting	Facial expressions, language, and prosodic information
[22]	Recognizing engagement levels	Facial information, low-level auditory features

[33]	Predicting involvement levels	Mutual gaze and blinking visual features, acoustic features
[29]	Emotion classification during dyadic interactions	Dense trajectories and body part detection for human activity recognition, low level audio features
[49]	Recognizing emotional states during affective interactions	Hand gesture and speech features
[23]	Estimation of cohesion in task based group meetings	Turn-based, low-level audio features, motion energy features, combination of audio and visual features (motion during speech, motion when not speaking, audio-visual synchrony)
[2]	Recognizing spontaneous agreement and disagreement during conversations	Auditory features, manually annotated head and hand gestures

Table 2.5: Research on automatic detection of interest using nonverbal behavioral features extracted from other modalities

Ref.	**Task**	**Modality**	**Measures**
[6, 7]	Examining the relation between physiological synchrony and emotional states and attachment styles of partners	Physiological	Synchrony between EDA signals of two people
[42]	Detecting problematic relationship dynamics	Audio, physiological	Electrodermal activity (EDA), electrocardiogram (ECG) activity, EDA synchrony measures, language, and acoustic information
[35]	Recognizing prejudice	Audio, video, physiological	Spatial (mutual distance, space between interlocutors, participants' movements), audio (pauses in dialogue), and biometric (related to heart rate and emotional arousal)
[13]	Examining behavioral differences of different leadership styles	Motion	Mimicry of head nodding, hand gestures and posture changes

3

Data

For the experiments of this thesis, data from the MatchNMingle dataset [5] is used. The most important feature of this dataset is being recorded at an in-the-wild setting which contributes to the ecological validation of our work. In this chapter, first a description of the MatchNMingle dataset is given by explaining the experiment context and data collection procedure. Subsequently, the ground truth extraction for attraction in interactions is explained in detail. Finally, analysis of the extracted ground truth is made.

3.1. MatchNMingle Dataset

MatchNMingle is a multimodal and multisensor dataset which is recorded with the aim to be used in research about the automatic analysis of social signals and interactions for both social and data sciences [5]. It was collected in an indoor in-the-wild setting instead of a lab setting. Therefore, the social interactions between participants were as natural as possible. One of the usage intentions of the dataset is studying the relation between non-verbal behavior and attraction/attractiveness, therefore it fits the goal of this thesis perfectly.

3.1.1. Experiment context

The whole dataset was recorded during a set of activities taking place over 3 days in total in a bar. Each day the event started with a speed dating round where participants of opposite sex had a 3 minute date with each other, followed by a mingle party which lasted approximately one hour where participants could interact with each other freely. In this thesis, only the data from the first part of the event is used. Participants were recruited from a university and expected to fit the criteria of being single, heterosexual and between ages of 18 and 30. In total of 92 participants attended the event, with equal number of men and women and most of them did not know each other. During the event, participants were asked to wear devices around their necks that is shown in Figure 3.1, which record tri-axial acceleration and proximity. Because of the malfunction of hardware, some of the devices failed recording at various times. After removing these devices, in total 72 participants had sufficient data recorded by wearable devices.

3.1.2. Data collection

Acceleration data was collected using triaxial accelerometers at a frequency of 20 Hz. The proximity information is registered as binary values, with 1 representing two people who are in close proximity (~ 2-3 meters) and 0 representing two people being far away from each other. Additional to these recordings, whole event area was captured using 9 different cameras from top. Figure 3.2 shows example snapshots from recordings of these cameras. Another data source that is recorded at the event is the speed date responses. After each 3-minute date with the participant of opposite sex, participants were given 1 minute to fill a match booklet which has a questionnaire about their date partner indicating their interest on each other. Responses for these questionnaires constitute the ground truth for the tasks in this thesis.

At the online registration for the event, participants were also asked to fill different questionnaires to test individual differences. These questionnaires are 1) HEXACO personality inventory [1] 2) Self Control Scale (SCS)[37] 3) Sociosexual Orientation Inventory (SOI)[41]. In addition, in the beginning of the events hair samples from each participant are collected to gather hormonal baselines.

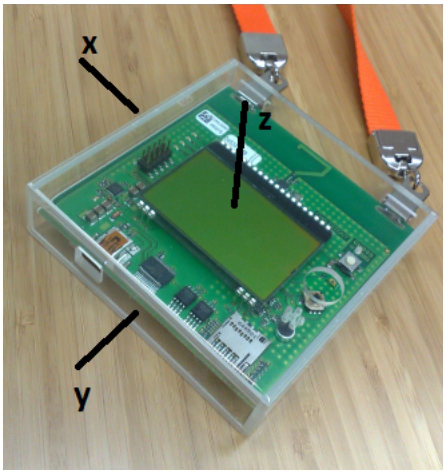


Figure 3.1: Wearable device that participants of the experiment wore during entire event that records triaxial acceleration and proximity. Three axes are shown.

As mentioned earlier, this thesis focuses on the speed date part of the events, thus the data from the accelerometer recordings and booklet responses are needed for the execution of tasks. For this reason, clean data from both of these sources were required. In a speed date, two people interact and since we focus on the dynamics of interaction between two people, the feature extraction and labeling tasks require to have valid data from both participants to be able to use this date in the experiments. Therefore, the dates for which at least one of the participants have a malfunctioning device and unreadable booklet responses are ignored and removed from the dataset. This resulted in a total number of 398 date interaction. Since each participant had their own label for each date, male and female participants of one date interaction were treated as separate samples for experimental tasks, resulting in total of 796 samples. In the following subsection, ground truth extraction process is explained in more detail.

3.2. Defining the ground truth

After each 3-minute date, participants filled the questionnaires in their match booklets to indicate their interest for the other person. The questionnaire consisted of following questions with responses on 7-point Likert scale (low = 1, high = 7):

- How much would you like to see this person again?
- How would you rate this person as a potential friend?
- How would you rate this person as a a short term sexual partner?
- How would you rate this person as a long term romantic partner?

The answers to these questions are used to quantify the attraction between participants in a date interaction. Each of these questions is used to define different tasks for the interest prediction problem as respectively *See Again*, *Friendly*, *Sexual*, or *Romantic*. The problem is treated as a binary classification problem to be consistent with the literature, meaning each date of a participant would have binary labels for each one of these tasks. For clarification, a *date* refers to the information from a single person during a speed date, whereas a *date interaction* refers to the interaction between two participants during a speed date. These two concepts are used in two challenges of this thesis. The first one is to predict if one participant is *attracted* or *not attracted* in his/her date partner. This would require labeling a person's *date* as positive or negative and thus for each speed date interaction male and female participant have their own labels. The second challenge is to predict if a *date interaction* ends up with a *match* or *no match*. Here, a *match* occurs when both participants label each other as *attracted* and for all other cases it becomes a *no match*.

To obtain labels for these situations, first the responses on Likert scale need to be binarized. One approach to do it is by taking the median of everyone's answers to each question and use it as a threshold. Dates with scores higher than this threshold are labeled as positive. However, this approach does not take the differences between people's scoring tendencies. To overcome this issue, we came up with another binarization method.

(a) Snapshot from speed date part (b) Snapshot from mingle part

Figure 3.2: Example snapshots from camera recordings during the event

Initially, each person's response scores for all of his/her dates are normalized to obtain scores between -1 and 1. Following this, dates that have positive score are labeled as *attracted* and negative score as *not attracted*. Following this, date interactions are labeled and a *match* label is given to a *date interaction* if both participants have *attracted* labels for their date and for all other cases a *no match* label is given to it.

3.3. Data Analysis

After labeling each date and date interaction, it is worth analysing the class distribution of each task. The distribution of labels for the dates over each class for each task is shown in Figure 3.3. *SeeAgain* and *Friendly* tasks have a balanced class distribution with 49% positive labels. On the other hand, *Romantic* and *Sexual* tasks have a bias on *not attracted* class with 40% and 42% positive labels respectively. As mentioned in the previous section, additional to date labels each date interaction is labeled as a *match* or *no match*. Distribution of match labels over each class for each task is shown in Figure 3.4. With the match labels, it is observed that for all tasks the datasets are highly biased towards *no match* class with 30% positive labels for *SeeAgain* and *Friendly*, 13% for *Romantic* and 19% for *Sexual* task.

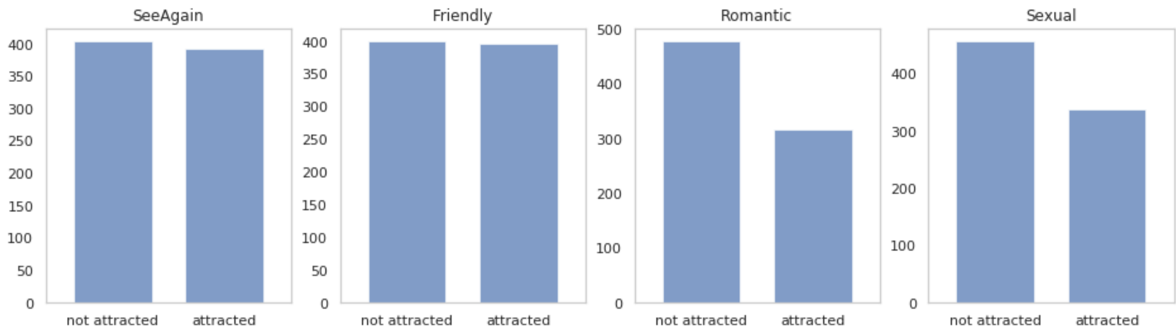


Figure 3.3: Class distribution of each task. SeeAgain and Friendly tasks have balanced class distribution, Romantic and Sexual tasks have higher number of data in *not interested* class.

In addition, a correlation analysis is conducted to understand the the correlation between attraction labels. Pearson correlation coefficients are computed using the labels of all dates per each attraction type and

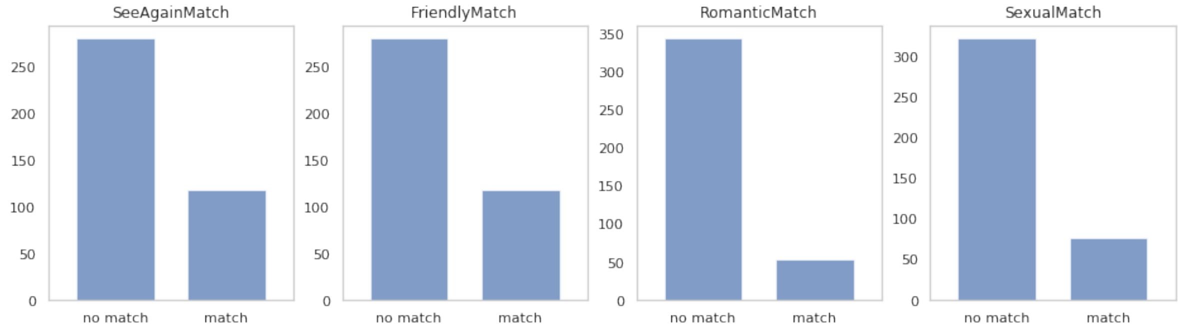


Figure 3.4: Class distribution of matches per each task. All tasks have an unbalanced distribution with higher number of *not interested* labels

presented in Tables 3.1 and 3.2 for one-way and mutual attraction respectively. One interesting observation is that even though *SeeAgain* is not considered as an attraction type, it shows the highest correlation with *Friendly* attraction labels in both table. Based on that, we can claim that people tend to give answers to *SeeAgain* question based on a friendly interest in the other person. Another observation is that in one-way attraction table, the highest correlation occurs between *Sexual* and *Romantic* labels but it is not repeated in mutual attraction table and the highest correlation is observed between *SeeAgain* and *Friendly* labels in that case. This is interesting because even though people rate their dates similarly for *Sexual* and *Romantic* attraction, matches do not necessarily occur in a similar manner for both types of attraction. In addition in one-way attraction table the lowest correlation is seen between *Friendly* and *Sexual* labels, indicating that people do not show a similar tendency while considering another person as a friend or a sexual partner. However, in mutual attraction table the lowest correlation is seen between *Friendly* and *Romantic* labels which is surprising because these attraction types actually showed a higher correlation in one-way labels.

	SeeAgain	Friendly	Sexual	Romantic
SeeAgain	1.0	0.588	0.442	0.540
Friendly	0.588	1.0	0.298	0.406
Sexual	0.442	0.298	1.0	0.599
Romantic	0.540	0.406	0.599	1.0

Table 3.1: Pearson correlation coefficients calculated between label distributions of each one-way attraction task. All correlations are significant with $p < .001$

	SeeAgainMatch	FriendlyMatch	SexualMatch	RomanticMatch
SeeAgainMatch	1.0	0.598	0.372	0.393
FriendlyMatch	0.598	1.0	0.363	0.312
SexualMatch	0.372	0.363	1.0	0.346
RomanticMatch	0.393	0.312	0.346	1.0

Table 3.2: Pearson correlation coefficients calculated between label distributions of each mutual attraction task. All correlations are significant with $p < .001$

Based on these observations we can claim that these results occur likely because of the differences of tendencies between male and female participants. Therefore correlations of male and female label distributions are calculated and presented in Tables 3.3 and 3.4 respectively. Here, we observe that even though female label correlations show a similar pattern with general label correlations, male label correlations show differences. The most interesting difference is that *SeeAgain* labels show a very high correlation with *Romantic* attraction labels indicating that unlike women, men makes their decision over wanting to see their date again based on their romantic attraction. These observations emphasize the differences between male and female in courtship settings.

	SeeAgain	Friendly	Sexual	Romantic
SeeAgain	1.0	0.592	0.506	0.635
Friendly	0.592	1.0	0.320	0.432
Sexual	0.506	0.320	1.0	0.631
Romantic	0.635	0.432	0.631	1.0

Table 3.3: Pearson correlation coefficients calculated between label distributions of male participants for each one-way attraction task. All correlations are significant with p < .001

	SeeAgain	Friendly	Sexual	Romantic
SeeAgain	1.0	0.581	0.400	0.453
Friendly	0.581	1.0	0.289	0.384
Sexual	0.400	0.289	1.0	0.565
Romantic	0.453	0.384	0.565	1.0

Table 3.4: Pearson correlation coefficients calculated between label distributions of female participants for each one-way attraction task. All correlations are significant with p < .001

Following the observations of male and female differences, it is also worth analysing the class distribution difference between genders as well. In Figure 3.5 the plots show the distribution of labels per gender. It is observed that *Friendly* and *Romantic* tasks have similar number of positive labels in female group and male group (f:52% m:48% for Friendly, f:38% m:42% for Romantic) whereas *SeeAgain* task has significantly more positive labels in female group and *Romantic* task has significantly more positive labels in male group (f:52% m:45% for SeeAgain, f:36% m:48% for Sexual). From these we can claim that female and male participants did not differ in their inclinations toward friendship and long term romantic relationship but male participants had a higher tendency towards a short term sexual relationship. These findings align with sexual strategies theory by Buss and Schmitt [3]. They hypothesized that because of the lower levels of minimum parental investment incurred by men, men will express a greater desire for short-term mates than women will. The results of their empirical tests also showed that men and women did not differ in their stated inclinations for seeking a long-term partner but more men than women reported that they were seeking short-term sexual partners at that time.

Another interesting analyze is done by looking at the *no match* cases. A *no match* can occur in three different situations at a date interaction: only the male participant is not attracted, only the female participant is not attracted and both participants are not attracted. Figure 3.6 shows the distribution of each case for each task. Here, we see that in the *SeeAgain* and *Friendly* tasks the reason for the date interaction to end without a match is mostly because the male participant was not attracted to the female participant but the difference is not significant. On the other side, in the *Sexual* task it is most likely because the female participant was not attracted. In the *Romantic* task, mostly both of the participants were not attracted to each other. These

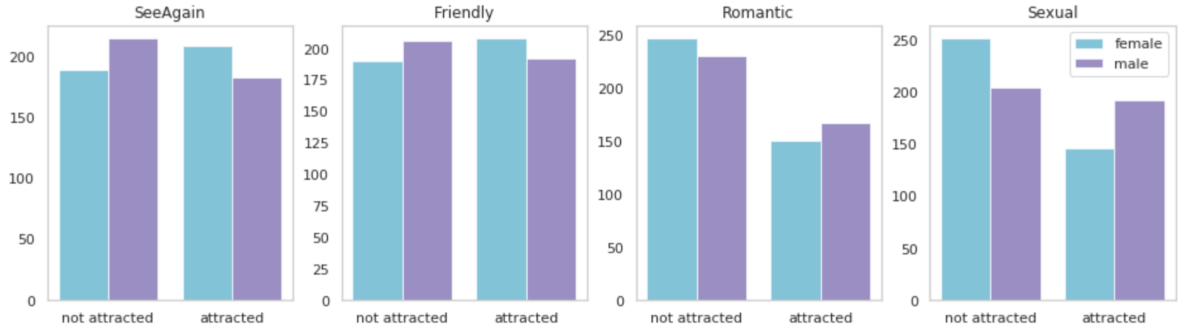


Figure 3.5: Class distribution of labels of each task per gender. *SeeAgain* task has significantly more data with *attracted* label in female group than male group and *Sexual* task significantly more data with *not attracted* label in female group than male group. *Friendly Romantic* tasks do not have a significant difference in data distribution.

findings also align with the psychological theories mentioned in the previous paragraph. Since men and women show similar inclinations for seeking a romantic partners, it can be expected that a match will not occur because of both participants' rejection. However, since more men than women are seeking for a short-term sexual partner, it is expected that female rejections are higher thus matches do not occur because of that.

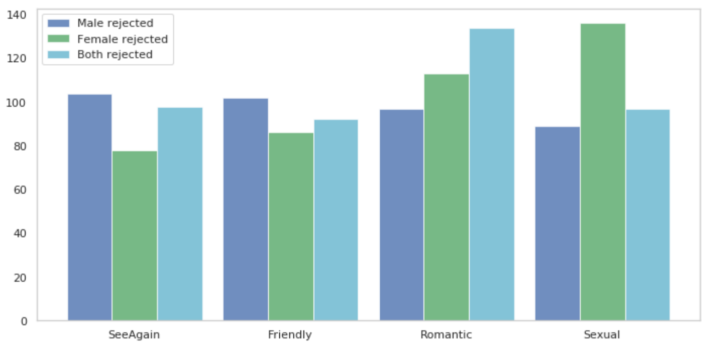


Figure 3.6: Analysis of *no match* cases. Plots show the distribution of the distribution of three different situations that cause a date interaction to be labeled as a *no match*.

4
Methodology

4.1. Introduction to Methodology

The method of this thesis aims to model the coordination of behavior between two people having a date, using behavioral features extracted from accelerometer readings, and use these features for predicting attraction between these people. This problem is defined as a binary classification task to be consistent with the literature. The first goal is to classify a single person as attracted or not attracted to the person he/she had a date with and the second goal is to classify a date interaction as a match, which both participants of the date are attracted to each other, or no match.

An initial preprocessing step is needed to extract low-level features, before extracting behavioral features, from the raw triaxial acceleration signal recorded by the accelerometers that participants were wearing. With this step, simple statistical and spectral features are extracted from the signal using a sliding window approach. This is explained more thoroughly in Section 4.2.

Following that, using these low-level features more complex features are extracted to model the behavior of individuals. These features are grouped into two categories: *Synchrony* and *Convergence*. The aim is to capture the coordination and similarity of behavior of two people in a date using both of their signals. The methodology for this part is explained in Section 4.3. Some of these features carry the same information for both people in an interaction which are considered *Symmetric* features whereas some of them can have a different meaning for each person which are considered *Asymmetric*. These concepts are also explained in more detail in Section 4.3.3. After extracting all features, before training models with them it is necessary to apply some preprocessing to have more reliable results. Applied methods for this purpose are explained in Section 4.3.4.

Finally, the methodology for classification tasks and the evaluation of them are explained in Section 4.4.

4.2. Preprocessing

This section explains the initial steps taken to transform the raw data into a format that can be used easily for extracting behavioral features from the acceleration signals. The accelerometer data recorded by the device consists of 3-dimensional readings with the X axis capturing the left-right movements; the Y axis up-down movements and Z axis forward-backward movements, with a sampling frequency of 20 Hz. Initially each axis of each person's recordings is normalized by computing the z-score within itself to remove interpersonal differences in movement intensity, using that person's entire accelerometer recording. Then, these normalized raw recordings are treated in multiple ways: the raw values of each axis, absolute values of each axis and the magnitude of the acceleration which is computed as $\sqrt{(x^2 + y^2 + z^2)}$. These different interpretations have been used in the previous research that used the data from bodily worn accelerometer to model human behavior [16, 17]. Especially in a context where two people are sitting opposite to each other, it is important to capture the direction of movement therefore raw values are useful. However, there can also be movements where the direction does not matter and that is the reason to use the absolute values. Additionally, magnitude will give an idea about the size of general movements but it ignores the specific movements where direction of action matters. Therefore it is important to use all these different interpretations of one signal recording. In total 7 different types of signal are derived from one accelerometer reading.

These signals can be treated as they are and behavioral features can be extracted from them. However in the literature, the studies that focused on extracting body movements from the accelerometers data have used statistical (mean and variance) and spectral (power spectral densities) features extracted from raw data using a sliding window approach [4, 16, 17, 24, 27]. This step is shown to be useful for detecting movements using the acceleration signal. Therefore it is also decided to use this approach in this thesis. In those studies, their tasks were different than ours and mostly focused on action detection, thus using short window lengths was more appropriate for their tasks. However since we do not focus on actions in short time periods but would like to learn more about the behavioral coordination between two people during the entire interaction, longer time periods can also reveal useful information. Based on this idea, it is decided to try varying time lengths for sliding windows and compare the performances of the features that are extracted using different window lengths. An illustration of the preprocessing steps is shown in Figure 4.1.

Each of the 7 signals derived from the raw recording mentioned earlier, is divided into n-second windows using a sliding-window approach, with *n/2* second shifts between each window. Since the optimal window size that captures necessary information is not known, the possible values of n are chosen as 1, 3, 5 and 10 seconds. Further, the effect of different window sizes can be evaluated.

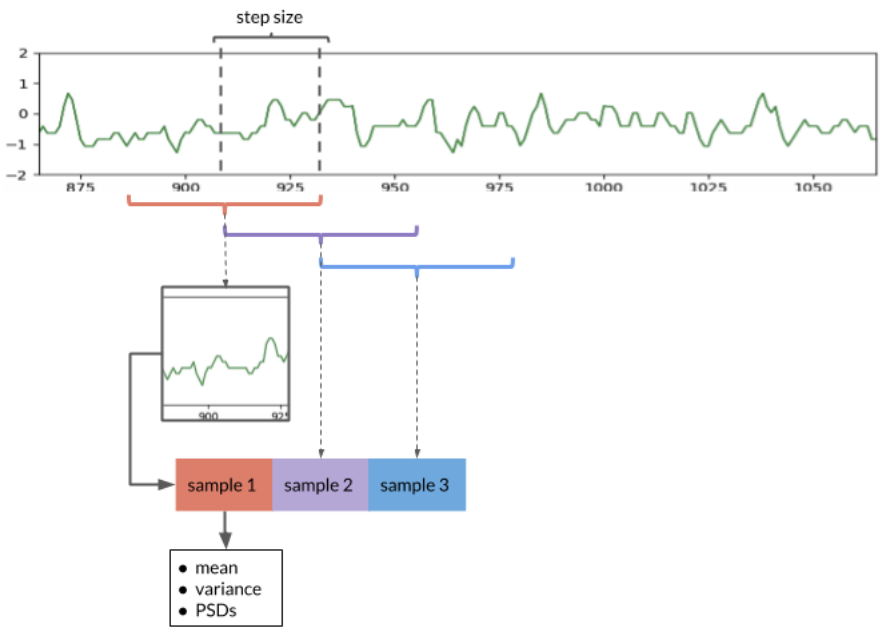


Figure 4.1: Using a sliding window approach with chosen window size, the signal is divided into samples. From each sample statistical and spectral features are extracted and they are further used for extracting complex behavioral features.

Similar to [17], statistical and spectral features are extracted from each window. As for statistical features mean and variance over the values in each window is computed. As the spectral feature, power spectral density (PSD) per window is computed using 6 logarithmically spaced bins between 0-10 Hz. Computing PSD is a method to convert a continues signal into a discrete form and shows how power of a signal is distributed over different frequencies. By dividing them into logarithmically spaced bins, the resolution at low frequencies is increased because lower frequencies contribute more to the power of the signal, but the dimensionality is kept low. Each bin gives information about the characteristic of behavior of the person at that time window, therefore each bin is treated as a single feature. Combination of the mentioned features results in 8 feature dimensions per window:

- mean
- variance
- PSD (6 bins)

Computing these 8 features for each 7 types of signal mentioned earlier results in 56 dimensions per window. While computing complex behavioral features, each dimension is treated separately and complex features are extracted from each of these dimensions. Since the complex features are computed with the sliding-window approach with 4 different window sizes, in total 224 different feature categories are extracted that will further be used to extract those behavioral coordination features.

4.3. Feature Extraction

As mentioned in the literature study, features that capture behavioral coordination and similarity between people's behavior are shown to be indicators of positive affect between people. These features can be grouped into two categories as *synchrony* and *convergence*. However there is no clear definition of these concepts in the literature and the terms have been used interchangeably. Therefore, each term is redefined here to be used in the scope of this thesis.

- **Synchrony:** In the survey on interpersonal synchrony, Delahercehe et al. [9] defined synchrony as "dynamic and reciprocal adaptation of the temporal structure of behaviors between interactive partners". Thus it can be interpreted as the coordination between two individual's behavior when they are in an interaction. Synchrony is the main feature that is used in the literature when modeling different kinds of interest based on movement features. One form of synchrony is mimicry, which is defined by Feese et al. [13] as person B following the behavior of person A by displaying the same nonverbal cue with person A just after they display it. These nonverbal cues can be face touch, arm cross, leg cross, posture change etc. As explained in Chapter 2, in the literature for quantifying mimicry either the behavior was annotated manually or a number of sensors were used to extract nonverbal cues and further quantify mimicry. Since in our case we have only one accelerometer and have no behavioral cue extracted from its signal such as head or hand gesture, we did not aim to capture mimicry in specific behaviors but rather in general movements.

- **Convergence:** While measuring the similarity in spoken dialogue of interlocutors Edlund et al. [10] refers to convergence as an increase in similarity. Thus, when two people's behavior converges it indicates that their similarity in behavior increases over time. As explained in Chapter 2, in the literature convergence of audio features of interacting partners have been used as an indicator of attraction and team cohesion. Here, we develop features to extract convergence from motion data. To our knowledge, this is the first time that motion convergence is used for estimating attraction.

In the following subsection, specific features for each category and the metrics are explained in more detail.

4.3.1. Synchrony

To measure synchrony of behavior of two interacting people, we decided to use four different metrics which are inspired from the literature. First metric aims to capture only the linear and monotonic relationship between two people's behavior whereas the second one can capture any kind of relationship. Third feature aims to capture the characteristics of mimicry behavior during the interaction by extracting the amount of mimicry that occurs. The last feature aims to capture a general idea about if one person is leading the conversation such that the other person mostly mimics him/her.

Correlation

Correlation has been used in the literature as a metric to capture similarity of overall body motion and also motion of specific body parts such as hands or head of two people [20, 38, 43, 45, 46]. Here we used Pearson correlation as the previous studies, which can be computed as follows:

$$\rho_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{\sigma(X)\sigma(Y)} \tag{4.1}$$

in which x and y refer to the low-level accelerometer features of person A and B and x_i and y_i refer to the sample-specific values. μ_x and μ_y denote the mean and $\sigma(X)$ and $\sigma(Y)$ denote the standard deviation of x and y over the samples considered.

It captures the linear correlation between two variables and returns a score between -1 and 1. In our context, correlation between two person's signals are computed and it is expected to obtain a score closer to 1 when two people have a positive feeling towards each other.

(Normalized) Mutual Information

This metric has also been used in the literature to measure the co-occurences between two people's behavior [17, 27]. It measures the mutual dependence between two variables and quantifies how much information can be obtained about one variable by observing the other variable. In our case it captures the dependence between two people's behavior on each other. It is calculated as follows:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{4.2}$$

where H(X) and H(Y) represents the entropy of input streams of random variables X and Y and H(X,Y) represents the joint entropy of these random variables.

Additionally, normalized mutual information is computed to obtain a score between 0 and 1. It is calculated as follows:

$$NI(X;Y) = \frac{I(X;Y)}{\sqrt{H(X)H(Y)}} \tag{4.3}$$

A higher score is obtained when two people have an influence on each other's behavior thus we expect it to be higher when two people has a more positive feeling for each other. For classification, both normalized and non-normalized mutual information is used.

Mimicry

This mimicry metric is inspired by the work of Nanninga et al. [31]. Since the context of interaction in our study is different than theirs, the metric is changed to adapt to our case. In their work, they had a meeting context where a group of people were interacting, thus they considered each person's mimicry with every one else.

The goal is to capture when one person imitates their partner's behavior. Figure 4.2 illustrates how this feature is computed. Each sample window of Person A's signal is compared with the consecutive window of Person B's signal. To compare these windows, the similarity between low-level features of these windows is computed using squared difference, resulting in similarity scores for the entire interaction as $D = [d_0, d_1, ...d_n]$. Following the approach of [31], from these similarity scores minimum ($min(D)$), maximum ($max(D)$), mean ($mean(D)$) and variance ($var(D)$) are computed and used as features. It should be noted that this is an asymmetric feature such that it returns different results depending on which person is chosen as Person A or Person B. Thus, it is computed for both cases.

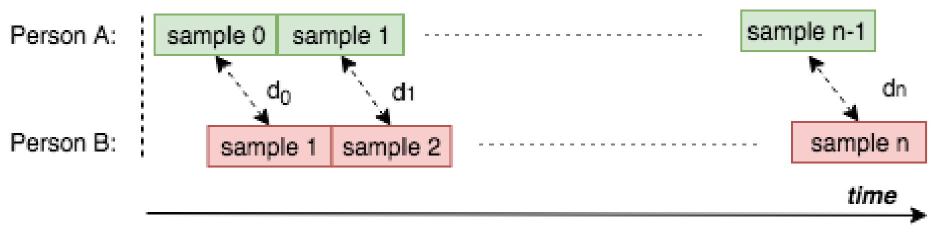


Figure 4.2: Illustrating the extraction of mimicry feature. Each time sample is compared with the other person's preceding time sample by computing a similarity score between sample features. Minimum, maximum, mean and variance of these sample scores are extracted to be used as mimicry features.

Time-lagged correlation

Additional to regular correlation of movement, in the literature lagged correlation has been used to measure the correlation of a person's movement at a given time with the interlocutor's movement at a time in the past [20, 45]. This is computed by measuring the similarity in behavior at different time lags again using Pearson correlation as follows:

$$\rho_{xy} = \frac{\sum_{i=1}^{N-\tau}(x_i - \mu_x)(y_{i+\tau} - \mu_y)}{\sigma(X)\sigma(Y)} \tag{4.4}$$

in which x and y refer to the low-level accelerometer features of person A and B and x_i and y_i refer to the sample-specific values. τ refers to the number of samples chosen to be used as positive lag. μ_x and μ_y denote the mean and $\sigma(X)$ and $\sigma(Y)$ denote the standard deviation of x and y over the samples considered.

This metric can indicate the leader-follower relationship of two people in a conversation by showing who is driving the interaction. In an example case of measuring the correlation between person A and person B's movement, if a higher score is obtained when person B's signal is positively lagged, this indicates that person B is leading the interaction. In the literature, time lags of 0 to 5 second are used. We also follow this here and use +/- 1 time step lags, considering that the time steps are of size 1, 3, 5 or 10 seconds.

4.3.2. Convergence

To measure convergence, three different metrics are developed that are inspired from various literature. These features aim to measure if two people's behavior style is diverging or converging through the interaction they have. The idea is that if two people have a more positive feeling for each other, they show a more converging behaviour. The difference between first and second feature is that the asymmetric convergence aims to capture which participant's behavior style converged to the other's therefore it can help us define a leader-follower relationship at the interaction. On the other hand, the symmetric convergence feature does not measure the direction of the convergence but only the amount of convergence. Finally, the third feature aims to capture the convergence in a more general way by only comparing the first and second half of the interaction.

Symmetric convergence

This feature is inspired by the works of [28, 31] which computed convergence in audio features. Therefore, we adapted the feature to our modality. In order to compute audio features they had to extract speaking statuses and take only speech parts but we considered the entire signal for extracting the feature. This measure compares two people's behavior at each time step and aims to capture the similarity between their behavior increasing or decreasing over time. In order to compute it, corresponding window samples of two participants' signals are compared with each other. To measure the similarity at each time step, the squared difference between these corresponding windows' low-level features are computed, as illustrated in Figure 4.3, resulting in difference scores for each sample of the interaction as $D = [d_1, d_2, ..., d_n]$. After that, the correlation of these scores with time is computed to understand if they increase or decrease using Pearson correlation formula (eq. 4.4) and a correlation coefficient is obtained. In the Pearson correlation formula, X becomes the sample index of each window sample, so $1, 2, ..., N$ and Y is the difference scores, D. Since the goal is to capture convergence, decrease in difference scores means converging behavior. Therefore, the correlation coefficient is expected to be more negative for converging interactions meaning that the participants tend to show similar behavior over the interaction.

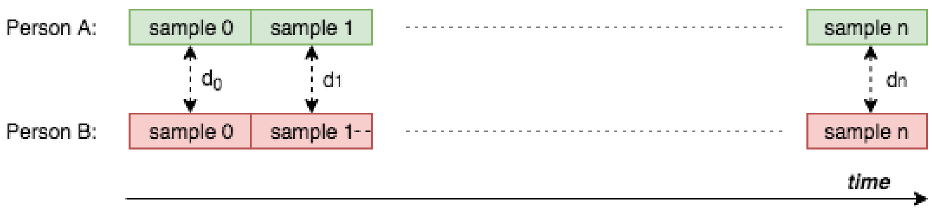


Figure 4.3: Illustrating the extraction of symmetric convergence feature. Each time sample is compared with the other person's corresponding time sample by computing a difference score between sample features. These difference scores are further correlated with time to extract one convergence score.

Asymmetric convergence

This feature has been inspired by [31] and again adapted to the context of this thesis. In their work since the setting was a meeting which lasted longer than our date interactions. Therefore we used a shorter time window for learning period which is explained further here. First two minutes of the date interaction is chosen as the learning period in which the behavior of one participant is modeled and the last one minute of the interaction is compared to this learned model. To understand if the second person's behavior converges to the first person's behavior, the low-level features of the samples from the last one minute are compared to the learning part's low-level features. To measure the similarity, squared distance between these features are computed as illustrated in Figure 4.4, resulting in difference scores for each sample in the last one minute of interaction as $D = [d_1, d_2, ..., d_n]$. Following, the correlation of these scores with time is computed to understand if they increase or decrease using Pearson correlation formula and a correlation coefficient is obtained. Since the goal is to capture convergence, decreasing difference means converging behavior. Therefore a more negative correlation coefficient indicates high convergence and more positive affect between two people. Since this feature only captures the convergence of Person B's behavior to Person A, it is also computed by changing the order of A and B.

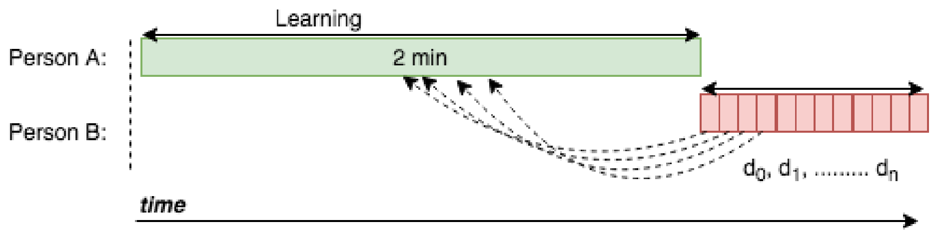


Figure 4.4: Illustrating the extraction of asymmetric convergence feature. Each time sample in the last 1 minute period is compared with the other person's first 2 minute by computing a difference score between sample features. These difference scores are further correlated with time to extract one convergence score.

Global convergence

This feature has been inspired by the work of [28]. Since they used audio channel and measured pitch convergence, the feature is adapted to fit in the context of this thesis. The idea is to measure the similarity of two people's behavior in the beginning and at the end of their date interaction and compare these similarities. It is expected that the behavior will be more similar at the end of the interaction due to the convergence. To capture this, first and second half of the signals are taken as illustrated in figure 4.5. Similarity between both person's first half's features are computed using squared differences and saved as d_0, and similarity between their second half's features are computed and saved as d_1. After that, the difference between these similarities is computed by subtraction as:

$$c = d_1 - d_0 \tag{4.5}$$

The difference is expected to be negative, meaning the behavior is more similar in the end of the interaction due to the convergence.

Figure 4.5: Illustrating the extraction of global convergence feature. The interaction is divided into half. First half of each person's signal's features are compared and one similarity score is obtained. Same is applied for the second half of the interaction and another similarity score is obtained. First one is subtracted from the second one to extract the global convergence score.

4.3.3. Symmetric vs. Asymmetric features

Symmetric features are the features that are same for both participants of the date such as correlation, mutual information, symmetric convergence and global convergence. On the other hand, asymmetric features can capture the direction of mimicry or convergence and thus have a different meaning for each participant in a date. For example, if the male participant is mimicking the female, this can be interpreted differently for each side. These features include time-lagged correlation, mimicry and asymmetric convergence. Effect of each of these feature categories is further analysed for different tasks.

Table 4.1 summarizes all the features that are used in our experiments with the corresponding IDs. Additionally, symmetric features are colored with red and asymmetric features are colored with blue.

4.3.4. Feature preprocessing

After extracting the behavioral coordination features, we perform some preprocessing on them before using them for classification tasks with training and testing. This is necessary because when features are extracted considering all the variables mentioned in the previous section, the number of features becomes very high.

Table 4.1: Feature vector with IDs. Grouped into two feature types. Symmetric features are colored with red and asymmetric features are colored with blue.

Feature type	Feature	ID
Synchrony	Correlation	0-223
	Mutual Information	224-559
	Mimicry	560-2351
	Time-lagged correlation	2352-2799
Convergence	Symmetric convergence	2800-3135
	Asymmetric convergence	3136-3583
	Global convergence	3584-3807

Reducing the dimensionality of the feature space will help creating simpler models, reducing the training time and also overfitting.

Feature scaling
Feature scaling is a method that is used for machine learning tasks to standardize the features of the used dataset. The reason is because the range of features can vary widely and particular features can tend to dominate other features when learning models and this might cause overfitting. By scaling, variance of the features are taken in the same range. Therefore, standardization is applied to the features using the following formula:

$$z = \frac{x - \mu_x}{\sigma_x} \tag{4.6}$$

Feature selection
There are different ways to apply feature selection. We apply a feature selection method which computes the ANOVA F-value between each feature and target labels and selects the features with highest scores. As for the highest score, it is decided to select features which have a significantly high F-value ($p < 0.05$). The main point of this feature selection method is selecting a set of representative features by taking into account correlations between features and ground truth labels for each task.

Dimensionality reduction
As for the last step of feature preprocessing, further dimensionality reduction is performed by applying principal component analysis (PCA) to the selected features and the top principal components which preserve the 95% of the variance are kept.

4.4. Classification

In the experiment part of this work, we run classification tasks for predicting attraction between people. The first classification problem is predicting one-way attraction that is, to predict if a person in a date interaction is attracted or not to their partner. The second classification problem is predicting mutual attraction that is, if both people are attracted to each other or not. In the case of mutual attraction, this date interaction is labeled as a match. More detailed explanation of labeling the dates can be found in Chapter 3. The predictive performance of the extracted features that are mentioned in the previous section is assessed in both problems. In the rest of this section, an overview of the classification and evaluation methods are explained.

4.4.1. Classifier

A logistic regressor is chosen as classifier for the task of predicting interest similar to [4]. This classifier uses a logistic function to model the relationship between a binary dependent variable and independent variables. It optimizes the following cost function to predict the binary dependent variable correctly:

$$min_{w,c}\frac{1}{2}w^T w + C\sum_{i=1}^{n} log(exp(-y_i(X^T w + c)) + 1) \tag{4.7}$$

Where w are the feature weights, C is a regularization parameter, y_i are the label values (0 or 1), X_i feature vector and c is a bias term.

4.4.2. Evaluation

To evaluate the predictive performance of classifiers for each task, a nested 10-fold cross-validation is applied. To do this, initially the dataset is divided into ten different subsets and 9 of them are used for training and 1 is used for testing. While training each of these 9 sets, grid search is applied to find the optimum regularization parameter of the model. After selecting the model with the optimum hyperparameter, it is tested with the test fold to evaluate its performance. This is repeated 10 times by changing the testing fold in each iteration and average of their performances is presented. To obtain a measure that is unaffected by the class imbalance, the Area under the Receiver Operator Characteristic (AUC) was used to determine performance. Area under this curve (AUC) is computed to obtain a score between 0 and 1.

Baseline features

It is also necessary to compare the performance of our features with the state-of-the art features that have been used for attraction prediction tasks in the literature. Veenstra et al. [44] used behavioral features that are extracted from the video and Cabrera-Quiros et al. [4] adapted some of these features to make it possible to extract them from accelerometer data. They extracted the mean and variance of the magnitude of acceleration for the entire signal of each person for each date. Additionally, they also applied a sliding window of 1 second with a shift of 0.5 seconds and calculated the variance over these windows. From this variance over sliding windows, they extracted again the mean and variance which results in 4 simple features. Since their setting is the most similar to our context, we compared the performance of these features with our synchrony and convergence features.

5
Results

In this chapter, the results of the conducted experiments are outlined. First, in Section 5.1.1 performances of classification tasks for one-way attraction prediction and mutual attraction prediction using behavioral coordination features are described. The performance of one-way attraction prediction is also compared with the state-of-the-art features from the literature. In Section 5.2, features are analysed more thoroughly to understand their specific contributions and influence of different factors are discussed. Finally in Section 5.3 motion convergence is investigated more deeply.

5.1. Prediction of attraction with behavioral coordination features

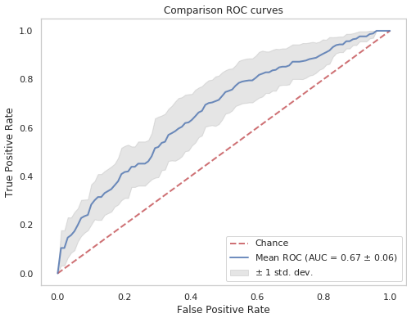
This section outlines the results for predicting one-way attraction and mutual attraction (match) by adopting the suggested methodology. First, a ROC AUC analysis is provided, comparing the results to the random baseline. Then the results are compared with the state-of-the-art features from the setting closest to ours.

5.1.1. Predicting one-way attraction

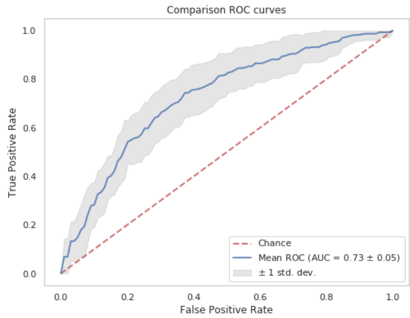
The first problem investigated in this thesis is that of predicting if a person is attracted to his or her date partner. Within this problem, we had four tasks as: predicting whether a person would like to see the other person again (*SeeAgain*), predicting whether a person has a friendly attraction to the other person (*Friendly*), predicting whether a person has a romantic attraction to the other person (*Romantic*) and predicting whether a person has a sexual attraction to the other person (*Sexual*). As explained in Chapter 4, for each task 10-fold cross-validation is applied with a logistic regressor. In Figure 5.1 the performance for predicting different attractions types are shown by the mean ROC curves for these folds with ± 1 standard variation. Performances are compared to the random baseline classifier which assigned every data point to the most-frequent class. Obtained mean AUC scores are 0.67($\pm$0.06) for the *SeeAgain* task, 0.73($\pm$0.05) for *Friendly* task, 0.68($\pm$0.04) for *Romantic* task and 0.69($\pm$0.07) for *Sexual* task.

As is seen, for all tasks our features performed better than the random baseline of 0.50 AUC score. The performance of *Friendly* task is the highest showing that it was a relatively easier type of attraction to predict than others. This can be possibly explained by men and women showing similar patterns of behavior in a social attraction situation which is discussed in more detail in Subsection 5.1.1. In addition, *SeeAgain* task is seen to be relatively more difficult and the possible reason for this is because the labels for this task are obtained using the answers to ”wanting to see the person again” question which does not target a specific attraction type thus people could interpret it differently. Therefore the ground truth might not be very clear to be modelled using our features. We can compare these results to the similar work of Veenstra et al. [44]. They aimed to predict if people want to exchange contact information in a speed-date and also to predict how physically attracted the participant is to the other person. Even though they are not exactly the same type of question, *SeeAgain* task is similar to predicting wanting to exchange contact information. In their experiments, they also achieved a higher score in predicting physical attraction than wanting to exchange contact information. Based on this, we can claim that people's answers to the specific questions about attraction can be predicted more easily.

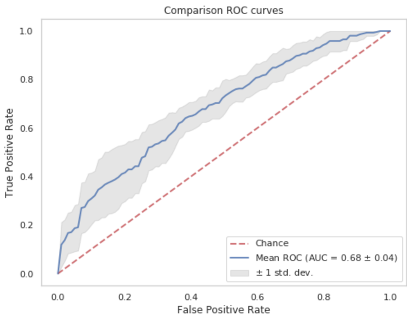
We also compared the performance of our features with the state-of-the-art features from [5], which is the closest to our setting in terms of approach and modality. The examined features were:



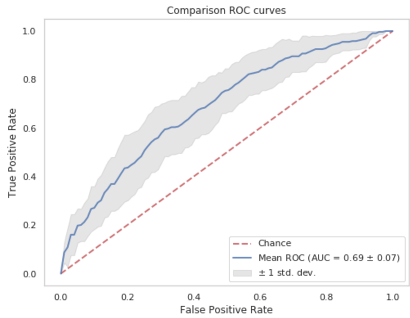
(a) SeeAgain task



(b) Friendly task



(c) Romantic task



(d) Sexual task

Figure 5.1: ROC curve plots for visualizing the performance of predicting one-way attraction using behavioral coordination features. For all tasks the performance is better than random baseline.

- Mean of the magnitude of acceleration
- Variance of the magnitude of acceleration
- Mean of the variance of acceleration over 1 second windows
- Variance of the variance of acceleration over 1 second windows

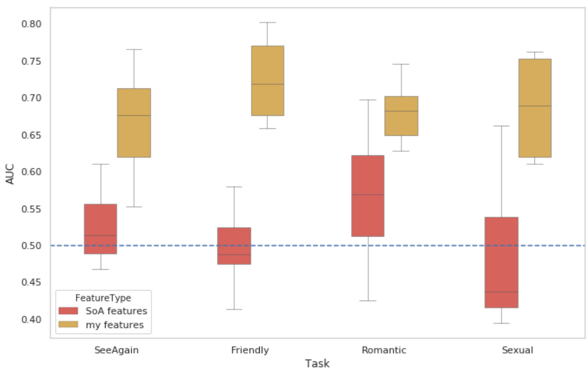


Figure 5.2: Boxplots visualizing the AUC scores for comparison. Red boxes indicate the scores obtained by using baseline features, yellow boxes indicate the scores obtained in this work. For all tasks features used in this work outperformed the baseline features.

In Figure 5.2 the results for the prediction of different attraction types are outlined using both sets of features. As is seen, the prediction power of our features is higher for all tasks. Their features take into consideration only single participant's movements but in order to extract our features we used both person's movement signals and computed behavioral coordination between them. These results indicate that while modeling attraction between two people, it is important to model the coordination of behavior rather than computing only features of individual movements.

Separating male and female data

As mentioned in the literature study, research in psychology showed that men and women have differences in mate-selection and courtship behavior [14, 18]. Therefore, this is taken into account here while creating the classification tasks as well. Previous studies over attraction prediction split the men and women data and ran classification tasks separately on each set [4, 44]. Even though, Veenstra et al. [44] obtained better accuracy by classifying each gender's data separately, results of Cabrera-Quiros et al. [4] showed better performance when they combined both data. Thus, this might be dependent on the features that are used in each study. In this thesis, separating and combining male and female data are tried and expected that separated case will perform better due to the gender behavioral differences. In Figure 5.3 mean AUC scores are plotted to compare the prediction performances for each attraction task. Performances of male and female tasks are compared with the combined dataset's performances.

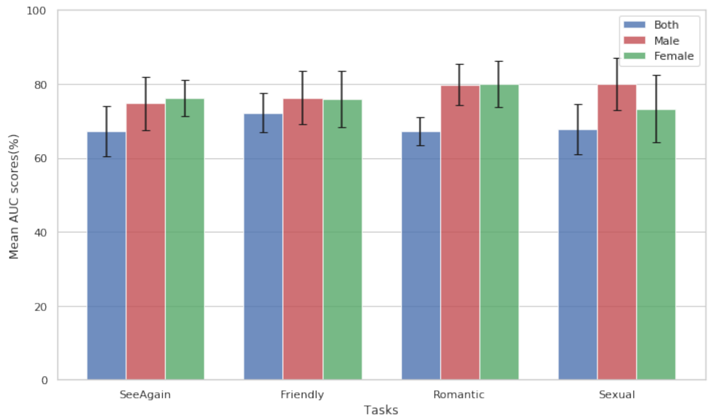
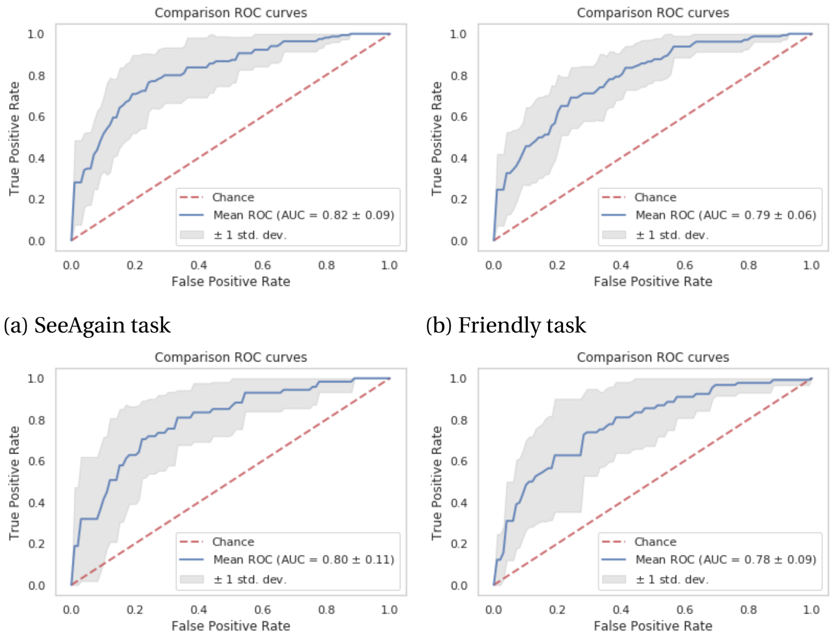


Figure 5.3: Classification performance (Mean AUC) for predicting attraction by separating male and female data analysed. Blue bars indicate the results obtained using both data as combined, red bars indicate using only male data and green bars indicate using only female data. Separation increased the prediction performance for all tasks.

As is seen, separation increased the prediction performance for all tasks. However the improvement in *Friendly* tasks is not statistically significant ($p > 0.1$), with mean AUC score of 0.76($\pm$0.07) for males and 0.75($\pm$0.08) for females indicating that men and women show similar behavior when they have a friendly interest in each other. This could also be the explanation for *Friendly* task to show the highest prediction performance when both data is used, compared to other types of attraction because a more general pattern can be modelled since both males and females show similar movement features related to social attraction. On the other hand for *Romantic* and *Sexual* tasks we have a statistically significant increase ($p < 0.1$) with mean AUC score of 0.79($\pm$0.04) for males and 0.80($\pm$0.06)for females in *Romantic* and 0.80($\pm$0.07) for males and 0.75($\pm$0.10) for females in *Sexual*, in line with the literature [14, 18] this indicates that men and women show different behavioral characteristics when they have a romantic or sexual attraction to the another person. Similarly, for *SeeAgain* task performance significant increase is observed in the separated case with mean AUC score of 0.76($\pm$0.07) for males and 0.77($\pm$0.06) for females.

5.1.2. Predicting mutual attraction

The second problem investigated in this thesis is to predict if both people who had an interaction are attracted to each other or not. Similar to the first problem we have four tasks, each for one type of attraction. For each task again 10-fold cross-validation is applied with a logistic regressor. In Figure 5.4 the performance for predicting mutual attraction are plotted with the mean ROC curves for these folds with $\pm$1 standard variation. Resulting mean AUC scores are 0.82($\pm$0.09) for *SeeAgain* task, 0.79($\pm$0.06) for *Friendly* task, 0.80($\pm$0.11) for *Romantic* task and 0.78($\pm$0.09) for *Sexual* task. We observe that the mutual attraction prediction tasks have shown better performance than one-way attraction prediction tasks. Another interesting observation is that there is not a similar trend in task difficulties of one-way attraction and mutual attraction prediction tasks. We could not compare our results with the state-of-the-art features' performance because they did not use their features for predicting mutual attraction.



(a) SeeAgain task (b) Friendly task

(c) Romantic task (d) Sexual task

Figure 5.4: ROC curve plots for visualizing the performance of predicting mutual attraction using behavioral coordination features. For all tasks the performance is higher than random baseline.

5.2. Feature analysis

This section provides a more thorough examination on the performances of different features. First, in Subsection 5.2.1, a correlation analysis between the vectors of single features and the ground truth for each task is presented. Subsequently, in Subsection 5.2.2 and Subsection 5.2.3 the influence of different signal types used for extracting features and the influence of different window sizes for feature extraction is studied respectively. Subsection 5.2.4 elaborates on the performance differences of each behavioral coordination feature and lastly Subsection 5.2.5 compares the performances of symmetric and asymmetric features.

5.2.1. Correlation analysis

In order to have a deeper understanding about the contributions of each feature, a correlation analysis of each feature and the labels for each task is conducted. For one-way attraction prediction task, only the features with the correlation coefficients greater than 0.1 or less than -0.1 are investigated. However, for match prediction tasks there were more features with high correlation thus only the top 10 features are investigated here. As expected, the features with the highest correlation coefficients were found to vary with different tasks. This indicates that each type of attraction manifests in different behavioral characteristics. Table 5.1 summarizes the features with highest correlation coefficient for each task of one-way attraction prediction and Table 5.2 summarizes the features for each task of mutual attraction prediction.

One interesting finding of this analysis is that *Correlation* features that are computed over the Z-axis are negatively correlated with *Friendly* attraction as opposed to the expectation of positive correlation as explained before. Z-axis captures the forward-backward acceleration of the body. Therefore, negative correlation could be because of one person's backward and other person's forward movement occurring simultaneously. Considering the fact that during the interactions people were sitting opposite to each other, this might be due to people's simultaneous movement occurring along the same axis but in different directions. On the other hand, most of the *Correlation* features that are extracted using PSD bins indicating correlation in the movement frequencies of couples showed positive correlation with the *Friendly* and *Sexual* attraction. It shows that the correlation of movement did not occur necessarily in the direction of movement but also the frequency of movement of two people were similar.

It is also seen that *Mutual Information* features had high positive correlation with only the *SeeAgain* and *Friendly* tasks whereas the *Mimicry* features had high positive correlation with only the *Romantic* and *Sexual* tasks. These findings are interesting because in the literature synchrony and mimicry features have been used

Feature name	r-score	p-value
NMI 3sec PSD4 Z	0.10	0.003
GlobConv 10sec PSD1 Y	-0.10	0.004

(a) SeeAgain task

Feature name	r-score	p-value
Corr 1sec mean Z	-0.12	0.0008*
MI 5sec mean AbsZ	0.11	0.001
MI 5sec mean Z	0.11	0.001
MI 10sec mean Z	0.11	0.001
AsymConv 1-2 10sec PSD0 AbsY	0.11	0.001
NMI 3sec PSD4 Z	0.11	0.002
SymConv 5sec PSD4 X	0.11	0.002
NMI 10sec mean Z	0.11	0.002
AsymConv 1-2 10sec PSD2 Mag	0.11	0.002
AsymConv 1-2 10sec PSD5 Y	0.11	0.002
LagCorr 2-1 1sec mean Z	-0.11	0.001
LagCorr 1-2 1sec mean Z	-0.11	0.001
Corr 3sec mean Z	-0.11	0.002
NMI 5sec mean Z	0.10	0.003
MI 3sec PSD4 Z	0.10	0.003
AsymConv 1-2 5sec PSD0 X	0.10	0.004
LagCorr 2-1 1sec PSD4 Z	0.10	0.004
LagCorr 1-2 3sec mean Z	-0.10	0.003
GlobConv 10sec PSD5 Y	-0.10	0.004

(b) Friendly Task

Feature name	r-score	p-value
Mim_mean 2-1 10sec PSD0 Z	0.12	0.0006*
Mim_std 2-1 10sec PSD0 Z0	0.12	0.0007*
Mim_max 2-1 10sec PSD0 Z	0.12	0.0007*
Mim_mean 2-1 10sec PSD0 ZAbs	0.11	0.001
Mim_std 2-1 10sec PSD0 ZAbs	0.11	0.002
Mim_max 2-1 10sec PSD0 ZAbs	0.11	0.002
Mim_min 1-2 5sec PSD0 YAbs	0.11	0.002
Mim_std 2-1 5sec PSD0 Z	0.10	0.004
SymConv 10sec PSD2 ZAbs	-0.10	0.003

(c) Romantic Task

Feature name	r-score	p-value
LagCorr 1-2 10sec PSD5 Y	0.11	0.002
Mim_min 1-2 10sec PSD2 XAbs	0.11	0.002
LagCorr 1-2 10sec PSD4 Z	0.11	0.003
Mim_min 1-2 1sec var XAbs	0.10	0.003

(d) Sexual Task

Table 5.1: Features with the highest correlation coefficient with the ground truth for each task of one-way attraction prediction. All features listed are significantly correlated with $p<0.01$. * indicates that a feature is significantly correlated with $p < 0.001$. Acronyms used: NMI: Normalized mutual information, MI: Mutual information, GlobConv: Global convergence, AsymConv: Asymmetric convergence, SymConv: Symmetric convergence, LagCorr: Time-lagged correlation, Corr: Correlation, Mim: Mimicry, mag: Magnitude, Y: Y axis, Z: Z axis, X: X axis, Abs: Absolute value, PSD (with number): Power spectral density with which bin is used. In each feature name first word is the feature type, second word is the used window size, third and fourth word corresponds to the signal that the feature is computed with. In asymmetric features, 1-2 corresponds to participant leading the conversation and 2-1 corresponds to their partner leading the conversation.

Feature name	r-score	p-value
SymConv 5sec PSD4 X	0.19	0.0002*
SymConv 3sec PSD4 X	0.19	0.001
SymConv 5sec PSD5 XAbs	0.17	0.001
AsymConv f-m 5sec PSD2 YAbs	0.17	0.001
AsymConv f-m 10sec PSD2 YAbs	0.17	0.001
SymConv 10sec PSD4 X	0.16	0.002
AsymConv f-m 10sec PSD0 YAbs	0.16	0.002
SymConv 3sec PSD5 XAbs	0.16	0.002
SymConv 1sec PSD4 X	0.15	0.004
AsymConv f-m 10sec PSD4 YAbs	0.15	0.004

(a) SeeAgain task

Feature name	r-score	p-value
SymConv 5sec PSD4 X	0.19	0.0001*
SymConv 3sec PSD4 X	0.19	0.0001*
SymConv 5sec PSD5 XAbs	0.17	0.0008*
AsymConv f-m 5sec PSD2 YAbs	0.17	0.0009*
AsymConv f-m 10sec PSD2 YAbs	0.17	0.0009*
SymConv 10sec PSD4 X	0.16	0.001
AsymConv f-m 10sec PSD0 YAbs	0.16	0.001
SymConv 3sec PSD5 XAbs	0.16	0.001
SymConv 1sec PSD4 X	0.15	0.002
AsymConv f-m 10sec PSD4 YAbs	0.15	0.003

(b) Friendly task

Feature name	r-score	p-value
Corr 3sec PSD0 Mag	0.18	0.0002*
Mim_min m-f 5sec PSD2 YAbs	0.18	0.0002*
LagCorr 5sec neg PSD2 ZAbs	0.18	0.0002*
Mim_min m-f 1sec mean XAbs	0.18	0.0003*
AsymConv f-m 1sec mean YAbs	0.17	0.0005*
Corr 5sec PSD0 YAbs	0.16	0.0001*
Corr 5sec PSD5 Y	0.16	0.0001*
LagCorr 5sec neg PSD2 Z	0.16	0.0001*
SymConv 10sec var YAbs	-0.16	0.001
Corr 3sec PSD0 ZAbs	0.15	0.0001*

(c) Romantic task

Feature name	r-score	p-value
NMI 10sec mean X	0.17	0.0007*
MI 5sec PSD2 Mag	0.16	0.001
AsymConv m-f 3sec PSD0 YAbs	-0.16	0.001
Mim_min m-f 10sec var X	0.15	0.002
MI 10sec mean X	0.15	0.002
AsymConv m-f 5sec PSD1 YAbs	-0.15	0.003
Mim_min m-f 10sec PSD0 XAbs	0.14	0.003
AsymConv m-f 10sec var Y	-0.14	0.004
AsymConv m-f 5sec mean Z	-0.14	0.005
AsymConv m-f 3sec mean Z	-0.14	0.006

(d) Sexual task

Table 5.2: Features with the highest correlation coefficient with the ground truth for each task of mutual attraction prediction. All features listed are significantly correlated with $p<0.01$. * indicates that a feature is significantly correlated with $p < 0.001$. Acronyms used: NMI: Normalized mutual information, MI: Mutual information, GlobConv: Global convergence, AsymConv: Asymmetric convergence, SymConv: Symmetric convergence, LagCorr: Time-lagged correlation, Corr: Correlation, Mim: Mimicry, mag: Magnitude, Y: Y axis, Z: Z axis, X: X axis, Abs: Absolute value, PSD (with number): Power spectral density with which bin is used. For asymmetric features, m-f corresponds to male participant leading the conversation and f-m corresponds to female participant leading the conversation.

for modeling both romantic and non-romantic attraction/interest. However, out results show that people enjoying a friendly conversation show more synchronic behavior but in a flirtatious interaction mimicking behavior becomes more prominent.

Convergence features were expected to have negative correlation with the labels, because more negative convergence values indicate a higher convergence, which we hypothesize to be an indicator of positive attraction. This is obtained in *SeeAgain* and *Romantic* tasks but the opposite is observed in most features of *Friendly* task. We can conclude that actually the divergence in behavioral characteristics might be an indicator of friendly attraction to the other person.

Another observation is that there are not many features with high correlation for the *SeeAgain* task. This can be because the labels for this task are obtained by the answers to "wanting to see the person again" question and this is a vague description for any attraction. Therefore the ground truth might not be a very clear indicator of any attraction.

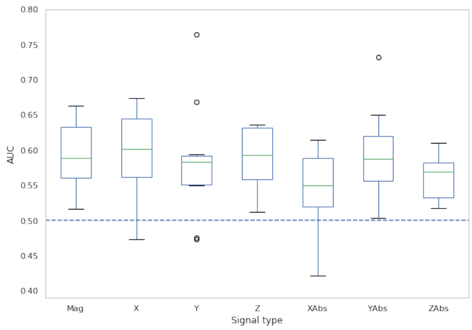
When analyzing the high correlated features with *Match* tasks, we could also pay attention to the direction of asymmetric features that will give us information about the leader-follower behavior of each gender. In *SeeAgain* and *Friendly* tasks, *male convergence to female* features have a positive correlation with matches. This is the opposite of what is expected because a positive convergence score shows a non-converging behavior and it was not expected to be correlated with attraction as indicated in [28]. On the other hand, in the *Sexual* task *female convergence to male* feature shows a negative correlation with attraction, indicating that when there is a mutual sexual attraction during an interaction it is revealed by female's behavior converging to the male's. We can discuss these findings under the light of the data analysis made in Chapter 3. As mentioned, no-matches in *Sexual* attraction occur mostly because female participant is not attracted in her date partner. Also in label distribution of *Sexual* task, male participants had a higher number of positive labels compared the females. Based on these, we can claim that the action and decisions of females are more determinant for two people to show mutual sexual attraction to each other. Additionally, *female mimicking male*-mimicry features show the highest positive correlation with mutual *Romantic* and *Sexual* attraction which shows that female's behavior is more distinctive for mutually flirtatious conditions. Interestingly even though both are synchrony features, *Correlation* shows higher correlation with *Romantic* matches whereas the *Mutual Information* shows higher correlation with *Sexual* matches.

In summary, we see that different types of attraction are indicated with different behavioral features. Therefore, depending on the task, a subset of features can be selected for better prediction performance. Even though we can not be certain that our features are capable of modelling exactly the behavior that we aim to model since we do not have ground truth about the synchrony and convergence, they are shown to be good at predicting the goal outcome which is to predict interpersonal attraction. In addition, synchrony and convergence terms are not very well defined and there is no consensus over them in the literature which makes it even more difficult to generalize our findings.

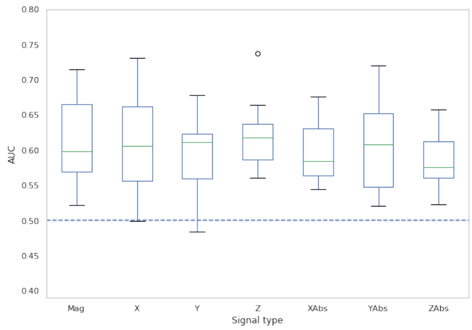
5.2.2. Influence of different signal types

In the preprocessing step, different interpretations of recorded tri-axial accelerometer signal are extracted which are raw values of each axis, absolute values of each axis and the magnitude of acceleration. Each of these signal types are expected to capture different movement types, therefore here we will analyze and discuss their effect on estimating different attraction types. Figure 5.5 and 5.6 present the performances of classification tasks using different signal types for one-way and mutual attraction prediction respectively. In the plots, each bar corresponds to the prediction scores using features extracted from one type of signal, calculated with 10-fold cross-validation. We observe that for each task, features extracted from different types of signal performed better than the rest. This indicates that different attraction types can be estimated better using features that capture different types of body movements.

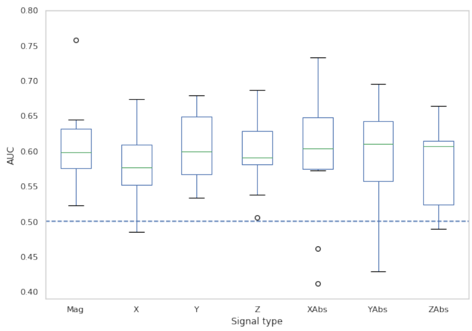
In one-way attraction estimation tasks, for *SeeAgain* task features extracted using *raw X* and *absolute Y* signals showed a relatively better performance than the others. In *Friendly* task results, one interesting observation is that *raw Z* features performed the best whereas *absolute Z* features performed the worst. Since Z-axis captures the forward-backward movement and in our setting the people were sitting opposite to each other, it is expected that *raw Z* and *absolute Z* signals can capture different interpretations of forward and backward leaning movement. Thus the large difference in performances can be explained with that. In *Romantic* task, *magnitude* of movement features showed the higher average AUC scores whereas they showed the worst performance in *Sexual* task. Based on this we can claim that, in *Romantic* attraction prediction the direction of the movement was more relevant than the other types of movements and it was less relevant for *Sexual* attraction. On the other hand, *raw Y* axis features showed the best performance in *Sexual* task im-



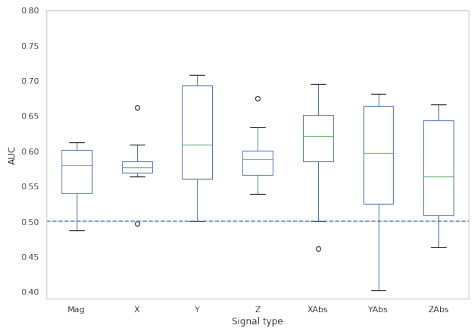
(a) SeeAgain task



(b) Friendly task

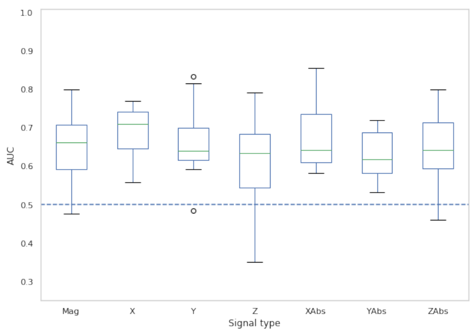


(c) Romantic task

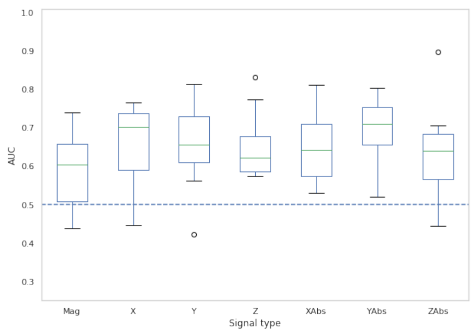


(d) Sexual task

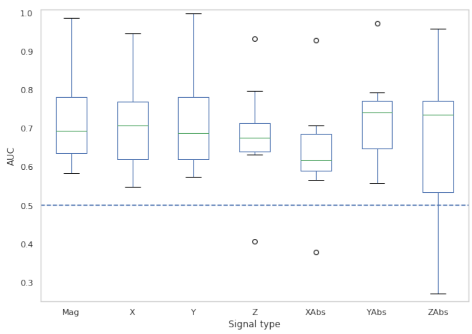
Figure 5.5: The relative contributions of the different signal types for each one-way attraction prediction task. Acronyms used: Mag: magnitude, XAbs: absolute value of X, YAbs: absolute value of Y, ZAbs: absolute value of Z



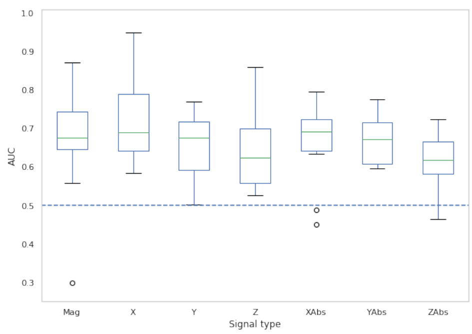
(a) SeeAgain task



(b) Friendly task



(c) Romantic task



(d) Sexual task

Figure 5.6: The relative contributions of the different signal types for each mutual attraction prediction task. Acronyms used: Mag: magnitude, XAbs: absolute value of X, YAbs: absolute value of Y, ZAbs: absolute value of Z

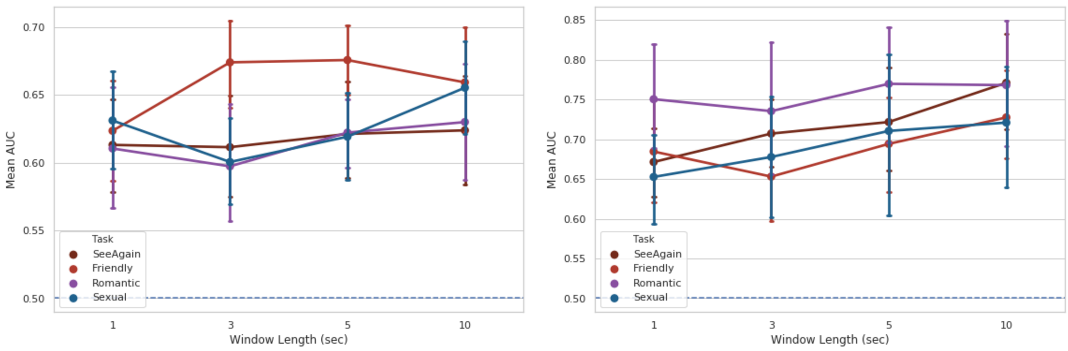
plying that upward-downward movement was a better indicator of *Sexual* attraction. This is quite interesting considering that people were sitting during the interactions and there did not occur big upward-downward movements. The main result of this analysis is that even though there was not a large difference between performance results obtained using each signal type, we observe that different type of signals showed the highest and lowest performances for each task.

In mutual attraction prediction tasks the differences in performances are observed as well. For *SeeAgain* and *Sexual* tasks *raw X* axis features performed the best implying that the features capturing right-to-left movements were better indicators for these attraction types. For *Friendly* task, both *absolute Y* and *raw Y* features achieved the highest results again implying upward-downward movement features which is an unexpected finding. In *Romantic* task almost all features showed similar performances but *absolute Z* and *absolute X* were the features that performed relatively worse than the rest.

One other important finding of this analysis is that the same attraction types were estimated better with the features extracted using different types of signal in one-way and mutual attraction tasks. This also shows that even though the attraction type is same, it is revealed with different types of movement characteristics when it is only one-way or mutual. Moreover, all feature types are shown to be predictive to some degree, indicating that there is not a single signal type clearly contains most of the information. Thus, features that are extracted using all kinds of signals seem to complement each other as the joint score is the highest.

5.2.3. Influence of different window sizes

In the preprocessing step, the sliding window approach is applied to the raw accelerometer signal to compute statistical and spectral features with varying window lengths. Different window lengths are expected to capture varying amount of information about the movement behavior of people from the raw signal. Therefore, it is worth comparing the performances of features extracted with each window length. Figure 5.7 presents the results of four different window sizes, 1, 3, 5 and 10 seconds, for both one-way and mutual attraction prediction. In the plots, each color of bar corresponds to the prediction scores of one attraction task, calculated with 10-fold cross-validation.



(a) One-way attraction

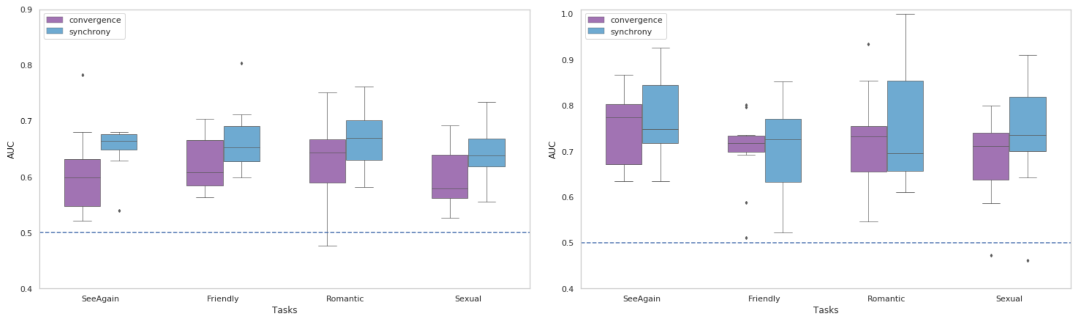
(b) Mutual attraction

Figure 5.7: Classification performance (Mean AUC) for predicting attraction using different window lengths for feature extraction analysed. Different colored lines indicate different tasks.

For one-way attraction prediction, we observe that features extracted with 10-second window size performed better than the rest for all tasks except *Friendly* task. This is an interesting behavior but as explained in Subsection 5.2.1, for predicting different types of interest different features carried more information. Thus, this behavior can be of because different features possibly capturing better information about behavior with different window sizes. For match prediction, it is seen that the features that are extracted with 10-second window sizes performed better for all tasks. From this we can conclude that behavioral coordination features can be captured with 10-second windows rather than smaller window sizes.

5.2.4. Comparison feature types

In this subsection the relative contributions of each feature type in the attraction prediction performance are studied. As Figure 5.8 shows, all types of behavioral coordination features had good estimation performances for all attraction tasks. Both with one-way and mutual attraction prediction, synchrony features performed slightly better than the convergence features for all tasks. This aligns with the literature since movement synchrony was shown to be an indicator of attraction between people whereas movement convergence has not been studied in that context but pitch convergence was seen as indicator of attraction [28]. On the other hand, it is observed that convergence features also performed better than random baseline indicating that movement convergence is also a good estimator of interpersonal attraction. In addition, the scores that are presented in Section 5.1.1 show that combination of these features always performed better than using them separately.



(a) One-way attraction (b) Mutual attraction

Figure 5.8: Comparison of the prediction performance, using convergence and synchrony features. Bar plots visualizing AUC scores. For all tasks synchrony features outperformed convergence features.

It should also be noted that these findings do not necessarily indicate that synchrony is a better measure than convergence for predicting attraction. Another possible explanation for this can be that our synchrony measures are better at capturing the synchrony behavior therefore they performed better. Synchrony features used in this study are mostly adapted from studies that computed these features using motion channel but convergence features are adapted from studies that used audio channel. Therefore, they might have required a different implementation than we did here. However, since we do not have the ground truth for synchrony and convergence behavior, it is not possible to ensure if these features capture what is intended the capture. This is further discussed in the next chapter.

5.2.5. Comparison symmetric and asymmetric features

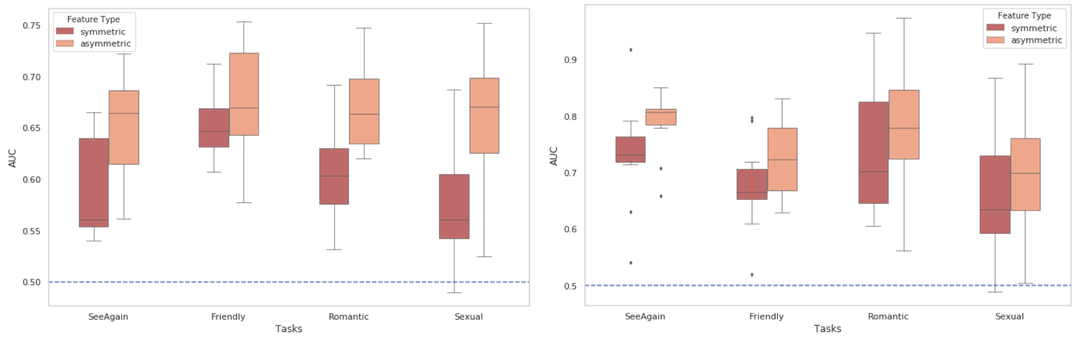
In this section, symmetric and asymmetric features are compared by examining their performances in predicting one-way and mutual attraction. As described in Chapter 4, symmetric features do not capture the direction of behavior therefore they are the same for male and female participant of the date. Symmetric features mentioned here are:

- Correlation
- Mutual information
- Symmetric convergence
- Global convergence

On other hand, asymmetric features can capture the direction of behavior and possibly give information about leader-follower characteristic of the interaction thus they have a different meaning for each participant during a date. For example, if the male participant is mimicking the female then she becomes the leader in this interaction and he becomes the follower. Therefore, in one-way attraction prediction tasks this mimicry feature obtains different values for the each participant of that date (Person A and Person B) and interpreted as "Person A is the leader" or "Person A is the follower". It should be noted that the gender is not taken into account because gender of Person A is not considered while extracting feature. On the other hand, for mutual attraction prediction tasks each date is considered as one data point and each participant do not have their own feature vector but the date interaction has one feature vector. In that case the asymmetric feature has one meaning for that date and it can be interpreted as "female participant is the leader" or "male participant is the leader". Asymmetric features mentioned here are :

- Time-lagged correlation

- Asymmetric convergence
- Mimicry



(a) One-way attraction (b) Mutul attraction

Figure 5.9: Comparison of the prediction performance, using symmetric and asymmetric features. Bar plots visualizing AUC scores. For all tasks asymmetric features outperformed convergence features.

Figure 5.9 plots the AUC scores of prediction tasks using symmetric and asymmetric features separately. These results show that asymmetric features outperformed the symmetric features in all tasks for predicting one-way attraction. They also showed a slightly better performance at predicting mutual attraction in match tasks. In either cases, separating them did not outperform the combination of them. Considering the fact that attraction is an asymmetric property, these results align with our expectations that attraction is estimated better using asymmetric features.

5.3. Characteristics of Motion Convergence

As mentioned earlier, convergence in audio features of the interacting partners have been studied in the literature [28] but convergence in movement during conversations is an understudied topic. It is shown that regardless of attraction when people are engaged in a conversational interaction, their speech audio features converge during the interaction. We want to test this theory with movement features and hypothesize that during the 3-minute dates the participants' movement characteristics converge regardless of being attracted to each other or not. In order to test our hypothesis, we compare the convergence scores of interacting pairs with non-interacting pairs and expect interacting pairs to have a significantly higher convergence score. Here, we created the non-interacting pairs by taking the motion signals of people who are of opposite sexes and not having a date with each other at that moment. Specifically, we took the signals that were recorded at the same time meaning both participants were having dates at that time but with different people.

For testing our hypothesis that interacting pairs have a higher convergence score than non-interacting pairs, we performed two-sample one-tailed t-test. It should be noted that the results of these statistical tests are not conclusive and in order to make strong conclusions, more tests should be conducted with different datasets. Therefore, our aim here is only to shed some light to the motion convergence concepts. For this reason, we used only the signals that are computed using 3-second windows. The results of paired t-tests for all three convergence features (*Symmetric*, *Asymmetric*, *Global*) extracted using all types of low-level features are presented in Appendix A.

Even though we can not make a strong conclusion, there are still results which are statistically significant with $p < 0.05$ and $p < 0.001$. Interestingly, in most cases convergence features that are extracted using PSDs showed a strong significance instead of mean and variance. This can possibly be explained by convergence occurring in movement frequencies while people are having an interaction regardless of attraction. Moreover, most of the significant results are achieved in *Symmetric* convergence scores compared to *Asymmetric* and *Global* convergence. This supports the idea that different convergence features capture movement convergence in different ways. Moreover, this can also be due to these features not actually measuring the motion convergence. As discussed earlier, due to the lack of ground truth for convergence behavior we can not ensure if the features do capture the intended properties.

Another interesting observation is that most of the significant results in *Symmetric* convergence are achieved with the features extracted using *raw Z* and *absolute Z* signals. This indicates that there is convergence in the movements along Z-axis in people during conversations. These findings also support the idea that using different signals rather than only taking the magnitude is important when using accelerometer data because as is seen these signals can possibly capture different movement characteristics better.

One other possible explanation of not having more significant results is the affect of people sitting. It should be noted that for creating non-interacting pairs we took pairs who were having date interactions with different people simultaneously. This means that interaction contexts were the same for interacting and non-interacting pairs where two people had a conversation in a seating position, which most probably started with an introduction round and lasted 3 minutes (date duration). Therefore, we could expect that people began the conversations moving more and ended less active which might cause most of the interactions to show a similar pattern. Thus, we could not observe a significant difference between interacting and non-interacting pairs' convergence scores.

In conclusion, these results shed some light on the movement convergence in conversing couples showing that convergence can occur regardless of attraction. However, in order to make strong conclusions more statistical tests must be done and also features should be compared with ground truth to ensure they grasp the intended motion characteristics.

6

Discussion & Conclusion

In this section, the limitations of the proposed approach and possible opportunities for future research is discussed. Finally, the thesis is concluded.

6.1. Discussion

One of the main challenges of this research was that it was not possible to validate if our synchrony and convergence metrics do actually capture the intended behavioral concepts. Even though the metrics were selected and developed after a thorough literature research and they performed well in classification tasks, it is an indirect validation of the actual behavior that we wanted to capture. In order to overcome this challenge, it can be helpful to find a way to validate the synchrony and convergence behavior. One possible method is to annotate them by externally observing the interactions. After obtaining the ground truth, the proposed features can be tested with that data to ensure that they do measure the right properties. With this method the features can be improved and presumably increase the prediction performances.

One another point to be taken into consideration is that in the dataset each participant had a date interaction with around 7 people from opposite sex. Therefore, some of the dates have the same person as female or male participant. However, for classification tasks we did not consider this and treated each data point independently. This approach does not take into account the interpersonal differences but rather accepts attraction behavior as the same for all people. For future work, it is possible to develop classification methods that will take the interpersonal differences that people will show in courtship situation into account.

The results show that synchrony features performed better for predicting all types of attraction. This aligns with the findings of the previous literature, however as indicated previously since we can not ensure that our features capture the intended concepts, it is not safe to claim that synchrony is a better feature for estimating attraction. The results also showed that asymmetric features performed better at classification task. This also aligns with the expected findings since attraction is an asymmetric property but the same idea also applies in this case and we can not claim it as a fact.

In addition, all behavioral coordination features are extracted using the same low-level features assuming these low-level features are of equal importance for the behavioral coordination concepts. However, it is also possible that for example convergence metrics might capture the intended behavior better with PSD features whereas synchrony metrics might perform better with signals extracted with X axis. In this study, these possible differences were not considered but for future research the metrics and low-level feature types can be investigated with more detail.

To further understand the contribution of the features, we analyzed how each feature type correlates with the ground truth labels. Even though we have seen that most *Correlation* features are positively correlated with attraction, some of the *Convergence* features also showed positive correlation which was against the expected findings. Therefore, motion convergence can be investigated more deeply in future research and affect of convergence on other types of interest and in different settings can be studied. Additionally, it should be noted that this correlation analysis aims only to provide general insights and should not be treated as strong statistical conclusions.

We also investigated the motion convergence regardless of attraction and found that some of our convergence features showed significantly higher scores when they are computed for interacting people compared to the non-interacting people. Even though our results were not enough to make strong conclusions, they were aimed to shed a light on movement convergence. For future work, it will be very interesting to test these findings in different datasets and with different convergence features. Moreover, motion convergence can also be an indicator of cohesion in team settings thus it will also be interesting to investigate it in group conversations.

6.2. Main Conclusion

In this study, we have presented a method for automatically predicting attraction between people using behavioral coordination features extracted from data recorded by a single body-worn accelerometer. We used synchrony and convergence features that are inspired from literature which used mostly audio and video channel, and adapted them to motion channel. Our features could predict one-way social attraction with a 73% Area under the ROC curve (AUC), out-performing the state-of-the art [5] which was the most similar setting to our case. We could also predict mutual romantic attraction with an AUC of 80%. To our knowledge, this is the first time that motion convergence is used for estimating interpersonal attraction.

Our results also showed that prediction performance increases when male and female data is separated, indicating that men and women have different behavioral characteristics when showing attraction. Finally, we ran a further analysis to investigate the individual feature contributions and found that different attraction types are indicated by different type of features. More specifically we could predict social attraction better using movement correlation features whereas for romantic and sexual interest mimicry features were better indicators. Moreover, features extracted from different types of signals recorded from accelerometers showed varying performances for different attraction types. Additionally, asymmetric features out-performed the symmetric features aligning with the fact that attraction is an asymmetric property and also our synchrony features showed better performance than convergence features for all types of attraction prediction tasks. Finally, we have seen that convergence in movement characteristics can occur in interacting people regardless of attraction.

Bibliography

[1] Michael Ashton and Kibeom Lee. Empirical, theoretical, and practical advantages of the hexaco model of personality structure. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc*, 11:150–66, 06 2007. doi: 10.1177/1088868306294907.

[2] Konstantinos Bousmalis, Louis-Philippe Morency, and Maja Pantic. Modeling hidden dynamics of multimodal cues for spontaneous agreement and disagreement recognition. pages 746 – 752, 04 2011. doi: 10.1109/FG.2011.5771341.

[3] David M Buss and David P Schmitt. Sexual Strategies Theory : An Evolutionary Perspective on Human Mating. 100(2):204–232, 1993.

[4] Laura Cabrera-Quiros, Ekin Gedik, and Hayley Hung. Estimating self-assessed personality from body movements and proximity in crowded mingling scenarios. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI 2016, pages 238–242, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4556-9. doi: 10.1145/2993148.2993170. URL `http://doi.acm.org/10.1145/2993148.2993170`.

[5] Laura Cabrera-Quiros, Andrew Demetriou, Ekin Gedik, Leander van der Meij, and Hayley Hung. The matchnmingle dataset: a novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates. *IEEE Transactions on Affective Computing*, 2018.

[6] T. Chaspari, B. Baucom, A. C. Timmons, A. Tsiartas, L. B. Del Piero, K. J. W. Baucom, P. Georgiou, G. Margolin, and S. S. Narayanan. Quantifying eda synchrony through joint sparse representation: A case-study of couples' interactions. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 817–821, April 2015. doi: 10.1109/ICASSP.2015.7178083.

[7] Theodora Chaspari, Adela C. Timmons, Brian R. Baucom, Laura Perrone, Katherine J.W. Baucom, Panayiotis Georgiou, Gayla Margolin, and Shrikanth S. Narayanan. Exploring sparse representation measures of physiological synchrony for romantic couples. *2017 7th International Conference on Affective Computing and Intelligent Interaction, ACII 2017*, 2018-Janua(Section 2):267–272, 2018. doi: 10.1109/ACII.2017.8273611.

[8] Cynthia L Crown. Coordinated interpersonal timing of vision and voice as a function of interpersonal attraction. *Journal of Language and Social Psychology*, 10(1):29–46, 1991.

[9] Emilie Delaherche, Mohamed Chetouani, Ammar Mahdhaoui, Catherine Saint-Georges, Sylvie Viaux, and David Cohen. Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing*, 3(3):349–365, 2012.

[10] Jens Edlund, Mattias Heldner, and Julia Hirschberg. Pause and gap length in face-to-face interaction. pages 2779–2782, 01 2009.

[11] Paul Ekman and Wallace V. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1:49–98, 1969.

[12] Sally D. Farley. Nonverbal Reactions to an Attractive Stranger: The Role of Mimicry in Communicating Preferred Social Distance. *Journal of Nonverbal Behavior*, 38(2):195–208, 2014. ISSN 01915886. doi: 10.1007/s10919-014-0174-4.

[13] S. Feese, B. Arnrich, G. Tröster, B. Meyer, and K. Jonas. Quantifying behavioral mimicry by automatic detection of nonverbal cues from body motion. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pages 520–525, Sep. 2012. doi: 10.1109/SocialCom-PASSAT.2012.48.

[14] Raymond Fisman, Sheena Iyengar, Emir Kamenica, and Itamar Simonson. Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, 121:673–697, 02 2006. doi: 10.1162/qjec.2006.121.2.673.

[15] Daniel Gatica-Perez. Modelling interest in face-to-face conversations from multimodal nonverbal behaviour. *Multimodal Signal Processing*, 03 2012. doi: 10.1016/B978-0-12-374825-6.00020-4.

[16] Ekin Gedik and Hayley Hung. Personalised models for speech detection from body movements using transductive parameter transfer. *Personal and Ubiquitous Computing*, 21(4):723–737, 2017.

[17] Ekin Gedik and Hayley Hung. Detecting conversing groups using social dynamics from wearable acceleration: Group size awareness. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(4):163:1–163:24, December 2018. ISSN 2474-9567. doi: 10.1145/3287041. URL `http://doi.acm.org/10.1145/3287041`.

[18] Karl Grammer, Masanao Honda, Astrid Juette, and Alain Schmitt. Fuzziness of nonverbal courtship communication: Unblurred by motion energy detection. *Journal of personality and social psychology*, 77:487–508, 10 1999. doi: 10.1037/0022-3514.77.3.487.

[19] Nicolas Guéguen. Mimicry and seduction: An evaluation in a courtship context. *Social Influence*, 4: 249–255, 10 2009. doi: 10.1080/15534510802628173.

[20] Yuval Hart, Efrat Czerniak, Orit Karnieli-Miller, Avraham E. Mayo, Amitai Ziv, Anat Biegon, Atay Citron, and Uri Alon. Automated video analysis of non-verbal communication in a medical setting. *Frontiers in Psychology*, 7(AUG):1–8, 2016. ISSN 16641078. doi: 10.3389/fpsyg.2016.01130.

[21] Chiao-Yin Hsiao, Wan-Rong Jih, and Jane Hsu. Recognizing continuous social engagement level in dyadic conversation by using turn-taking and speech emotion patterns. 07 2012.

[22] Yuyun Huang, Emer Gilmartin, and Nick Campbell. Conversational engagement recognition using auditory and visual cues. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 08-12-Sept(September 2016):590–594, 2016. ISSN 19909772. doi: 10.21437/Interspeech.2016-846.

[23] Hayley Hung and Daniel Gatica-Perez. Estimating Cohesion in Small Groups Using. *IEEE Transactions on Multimedia*, 12(6):563–575, 2010.

[24] Hayley Hung, Gwenn Englebienne, and Jeroen Kools. Classifying social actions with a single accelerometer. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 207–210. ACM, 2013.

[25] Johan C. Karremans and Thijs Verwijmeren. Mimicking attractive opposite-sex others: The role of romantic relationship status. *Personality and Social Psychology Bulletin*, 2008. ISSN 01461672. doi: 10.1177/0146167208316693.

[26] Anmol Madan, Ron Caneel, and Alex Pentland. Voices of attraction, 2004.

[27] Claudio Martella, Ekin Gedik, Laura Cabrera-Quiros, Gwenn Englebienne, and Hayley Hung. How was it?: Exploiting smartphone sensing to measure implicit audience responses to live performances. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, pages 201–210, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3459-4. doi: 10.1145/2733373.2806276. URL `http://doi.acm.org/10.1145/2733373.2806276`.

[28] Jan Michalsky and Heike Schoormann. Pitch convergence as an effect of perceived attractiveness and likability. In *INTERSPEECH*, 2017.

[29] Philipp M. Muller, Sikandar Amin, Prateek Verma, Mykhaylo Andriluka, and Andreas Bulling. Emotion recognition from embedded bodily expressions and speech during dyadic interactions. *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, pages 663–669, 2015. doi: 10.1109/ACII.2015.7344640.

[30] Iftekhar Naim, Md. Iftekhar Tanveer, Daniel Gildea, Ehsan Hoque, Mohammed, Hoque, and Ehsan Hoque. Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing*, 9(2):1–14, 2016. ISSN 1949-3045. doi: 10.1109/TAFFC.2016.2614299.

[31] Marjolein C. Nanninga, Yanxia Zhang, Nale Lehmann-Willenbrock, Zoltán Szlávik, and Hayley Hung. Estimating verbal expressions of task and social cohesion in meetings by quantifying paralinguistic mimicry. *Proceedings of the 19th ACM International Conference on Multimodal Interaction - ICMI 2017*, pages 206–215, 2017. doi: 10.1145/3136755.3136811. URL `http://dl.acm.org/citation.cfm?doid=3136755.3136811`.

[32] L. S. Nguyen, D. Frauendorfer, M. S. Mast, and D. Gatica-Perez. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Transactions on Multimedia*, 16 (4):1018–1031, June 2014. ISSN 1520-9210. doi: 10.1109/TMM.2014.2307169.

[33] Catharine Oertel, Stefan Scherer, and Nick Campbell. On the use of multimodal cues for the prediction of degrees of involvement in spontaneous conversation. pages 1541–1544, 01 2011.

[34] Shogo Okada, Laurent Son Nguyen, Oya Aran, and Daniel Gatica-Perez. Modeling dyadic and group impressions with inter-modal and inter-person features. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018.

[35] Andrea Palazzi, Simone Calderara, Nicola Bicocchi, Loris Vezzali, Gian Antonio di Bernardo, Franco Zambonelli, and Rita Cucchiara. Spotting prejudice with nonverbal behaviours. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*, pages 853–862, 2016. doi: 10.1145/2971648.2971703. URL `http://dl.acm.org/citation.cfm?doid=2971648.2971703`.

[36] M Pantic, Roddy Cowie, Francesca D'Errico, D Heylen, Marc Mehu, Catherine Pelachaud, Isabella Poggi, M Schroeder, and Alessandro Vinciarelli. Social signal processing: The research agenda. *Visual Analysis of Humans*, pages 511–538, 01 2011.

[37] Lars Penke and Jens Asendorpf. Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of personality and social psychology*, 95:1113–35, 12 2008. doi: 10.1037/0022-3514.95.5.1113.

[38] Fabian Ramseyer and Wolfgang Tschacher. Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. *Journal of consulting and clinical psychology*, 79 (3):284, 2011.

[39] Rajesh Ranganath, Dan Jurafsky, and Dan McFarland. It's not you, it's me: detecting flirting and its misperception in speed-dates. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 334–342. Association for Computational Linguistics, 2009.

[40] Gillian M Sandstrom, Vincent Wen-Sheng Tseng, Jean Costa, Fabian Okeke, Tanzeem Choudhury, and Elizabeth W Dunn. Talking less during social interactions predicts enjoyment: A mobile sensing pilot study. *PloS one*, 11(7):e0158834, 2016.

[41] June Tangney, Roy Baumeister, and Angie Luzio Boone. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of personality*, 72:271–324, 05 2004. doi: 10.1111/j.0022-3506.2004.00263.x.

[42] Adela C. Timmons, Theodora Chaspari, Sohyun C. Han, Laura Perrone, Shrikanth S. Narayanan, and Gayla Margolin. Using Multimodal Wearable Technology to Detect Conflict among Couples. *Computer*, 50(3):50–59, 2017. ISSN 00189162. doi: 10.1109/MC.2017.83.

[43] Wolfgang Tschacher, Georg M. Rees, and Fabian Ramseyer. Nonverbal synchrony and affect in dyadic interactions. In *Front. Psychol.*, 2014.

[44] A. Veenstra and H. Hung. Do they like me? using video cues to predict desires during speed-dates. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)(ICCVW)*, volume 00, pages 838–845, Nov. 2012. doi: 10.1109/ICCVW.2011.6130339. URL `doi.ieeecomputersociety.org/10.1109/ICCVW.2011.6130339`.

[45] Andrea Stevenson Won, Jeremy N. Bailenson, and Joris H. Janssen. Automatic detection of nonverbal behavior predicts learning in dyadic interactions. *IEEE Transactions on Affective Computing*, 5(2):112–125, 2014. ISSN 19493045. doi: 10.1109/TAFFC.2014.2329304.

[46] Andrea Stevenson Won, Jeremy N. Bailenson, Suzanne C. Stathatos, and Wenqing Dai. Automatically detected nonverbal behavior predicts creativity in collaborating dyads. *Journal of Nonverbal Behavior*, 38(3):389–408, Sep 2014. ISSN 1573-3653. doi: 10.1007/s10919-014-0186-0. URL `https://doi.org/10.1007/s10919-014-0186-0`.

[47] Bo Xiao, Panayiotis G Georgiou, Chi-Chun Lee, Brian Baucom, and Shrikanth S Narayanan. Head motion synchrony and its correlation to affectivity in dyadic interactions. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.

[48] Bo Xiao, Panayiotis Georgiou, Brian Baucom, and Shrikanth S. Narayanan. Head motion modeling for human behavior analysis in dyadic interaction. *IEEE Transactions on Multimedia*, 2015. ISSN 15209210. doi: 10.1109/TMM.2015.2432671.

[49] Z. Yang and S. Narayanan. Modeling mutual influence of multimodal behavior in affective dyadic interactions. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2234–2238, April 2015. doi: 10.1109/ICASSP.2015.7178368.

[50] Zhaojun Yang, Angeliki Metallinou, and Shrikanth Narayanan. Analysis and predictive modeling of body language behavior in dyadic interactions from multimodal interlocutor cues. *IEEE Transactions on Multimedia*, 16(6):1766–1778, 2014.

A

Overview convergence tests

Results of two-sample one-tailed t-test to test the hypothesis of interacting pairs having higher convergence score than non-interacting pairs, are presented.

Signal type	**p_value**
psd_Mag_0	0.478
psd_Mag_1	0.363
psd_Mag_2	0.270
psd_Mag_3	0.028*
psd_Mag_4	0.291
psd_Mag_5	0.026*
mean_Mag	0.186
var_Mag	0.258
psd_X_0	0.00003**
psd_X_1	0.233
psd_X_2	0.010
psd_X_3	0.035
psd_X_4	0.172
psd_X_5	0.497
mean_X	0.053
var_X	-
psd_Y_0	0.269
psd_Y_1	0.0002**
psd_Y_2	0.437
psd_Y_3	0.007*
psd_Y_4	0.494
psd_Y_5	0.021*
mean_Y	0.018*
var_Y	0.498
psd_Z_0	0.162
psd_Z_1	0.144
psd_Z_2	0.008*
psd_Z_3	0.077
psd_Z_4	0.386
psd_Z_5	0.261
mean_Z	-
var_Z	-
psd_XAbs_0	0.194
psd_XAbs_1	0.285
psd_XAbs_2	0.026
psd_XAbs_3	0.0008**
psd_XAbs_4	0.001*
psd_XAbs_5	0.126
mean_XAbs	-
var_XAbs	-
psd_YAbs_0	0.007*
psd_YAbs_1	0.233
psd_YAbs_2	0.0005**
psd_YAbs_3	0.002*
psd_YAbs_4	0.00007**
psd_YAbs_5	0.070
mean_YAbs	0.321
var_YAbs	0.032*
psd_ZAbs_0	0.354
psd_ZAbs_1	0.362
psd_ZAbs_2	0.314
psd_ZAbs_3	0.233
psd_ZAbs_4	0.338
psd_ZAbs_5	0.247
mean_ZAbs	-
var_ZAbs	-

Table A.1: Results of two-sample one-tailed t-test to compare *Asymmetric convergence* scores of interacting pairs with non-interacting pairs computed using all low-level feature types. * indicates that a score is statistically significant with $p < 0.05$ and ** indicates significance with $p < 0.001$. - indicates that t-test could not be conducted due to 0 variance.

Signal type	p_value
psd_Mag_0	0.042*
psd_Mag_1	0.096
psd_Mag_2	0.369
psd_Mag_3	0.476
psd_Mag_4	0.283
psd_Mag_5	0.011*
mean_Mag	0.054
var_Mag	0.126
psd_X_0	0.266
psd_X_1	0.0004**
psd_X_2	0.026*
psd_X_3	0.374
psd_X_4	0.055
psd_X_5	0.051
mean_X	0.0002**
var_X	0.325
psd_Y_0	0.285
psd_Y_1	0.179
psd_Y_2	0.136
psd_Y_3	0.166
psd_Y_4	0.184
psd_Y_5	0.231
mean_Y	0.185
var_Y	0.449
psd_Z_0	0.281
psd_Z_1	0.110
psd_Z_2	0.0004**
psd_Z_3	0.222
psd_Z_4	0.185
psd_Z_5	0.377
mean_Z	0.053
var_Z	0.018*
psd_XAbs_0	0.457
psd_XAbs_1	0.001*
psd_XAbs_2	0.247
psd_XAbs_3	0.473
psd_XAbs_4	0.374
psd_XAbs_5	0.232
mean_XAbs	0.114
var_XAbs	0.475
psd_YAbs_0	0.117
psd_YAbs_1	0.158
psd_YAbs_2	0.458
psd_YAbs_3	0.213
psd_YAbs_4	0.198
psd_YAbs_5	0.105
mean_YAbs	0.223
var_YAbs	0.247
psd_ZAbs_0	0.242
psd_ZAbs_1	0.252
psd_ZAbs_2	0.001*
psd_ZAbs_3	0.180
psd_ZAbs_4	0.189
psd_ZAbs_5	0.455
mean_ZAbs	0.009*
var_ZAbs	0.014*

Table A.2: Results of two-sample one-tailed t-test to compare *Global convergence* scores of interacting pairs with non-interacting pairs computed using all low-level feature types. * indicates that a score is statistically significant with $p < 0.05$ and ** indicates significance with $p < 0.001$. - indicates that t-test could not be conducted due to 0 variance.

Signal type	p_value
psd_Mag_0	0.030*
psd_Mag_1	0.009*
psd_Mag_2	0.046*
psd_Mag_3	0.080
psd_Mag_4	0.143
psd_Mag_5	0.006*
mean_Mag	0.389
var_Mag	0.082
psd_X_0	0.144
psd_X_1	0.065
psd_X_2	0.207
psd_X_3	0.329
psd_X_4	0.028
psd_X_5	0.120
mean_X	0.152
var_X	0.257
psd_Y_0	0.009*
psd_Y_1	0.458
psd_Y_2	0.407
psd_Y_3	0.157
psd_Y_4	0.061
psd_Y_5	0.034
mean_Y	0.086
var_Y	0.382
psd_Z_0	0.045*
psd_Z_1	0.016*
psd_Z_2	0.012*
psd_Z_3	0.243
psd_Z_4	0.004*
psd_Z_5	0.001*
mean_Z	0.011*
var_Z	0.016*
psd_XAbs_0	0.028*
psd_XAbs_1	0.066
psd_XAbs_2	0.130
psd_XAbs_3	0.470
psd_XAbs_4	0.028*
psd_XAbs_5	0.021*
mean_XAbs	0.089
var_XAbs	0.194
psd_YAbs_0	0.454
psd_YAbs_1	0.316
psd_YAbs_2	0.297
psd_YAbs_3	0.166
psd_YAbs_4	0.142
psd_YAbs_5	0.095
mean_YAbs	0.125
var_YAbs	0.415
psd_ZAbs_0	0.059
psd_ZAbs_1	0.006*
psd_ZAbs_2	0.020*
psd_ZAbs_3	0.344
psd_ZAbs_4	0.006*
psd_ZAbs_5	0.0001**
mean_ZAbs	0.025*
var_ZAbs	0.017*

Table A.3: Results of two-sample one-tailed t-test to compare *Symmetric convergence* scores of interacting pairs with non-interacting pairs computed using all low-level feature types. * indicates that a score is statistically significant with $p < 0.05$ and ** indicates significance with $p < 0.001$. - indicates that t-test could not be conducted due to 0 variance.

B

Workshop paper

In this appendix the paper submission is included, which was written based on the thesis. The paper is submitted to the Workshop on Emotions and Emergent States in Groups (EMERGent) in 8th International Conference on Affective Computing & Intelligent Interaction. The paper is accepted and will be presented on September 3th, 2019.

2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)

Estimating Romantic, Social, and Sexual Attraction by Quantifying Bodily Coordination using Wearable Sensors

Öykü Kapcak
Delft University of Technology
Delft, Netherlands
Kapcak@tudelft.nl

Jose Vargas-Quiros
Delft University of Technology
Delft, Netherlands
J.D.VargasQuiros@tudelft.nl

Hayley Hung
Delft University of Technology
Delft, Netherlands
h.hung@tudelft.nl

***Abstract*—In this paper we introduce a novel method of estimating romantic, social and sexual attraction between two people by quantifying their bodily coordination using wearable sensors in a speed-date setting. We developed simple synchrony and convergence features, inspired from the literature and specifically adapted to be extracted from accelerometer data. To our knowledge, this is the first time that motion convergence is used for estimating attraction. Our features could predict one-way social attraction with a 73% Area under the ROC curve (AUC), out-performing previous work in a similar setting. We also showed that prediction performance increased when the male and female data are separated. We could also predict mutual romantic attraction with an AUC of 80%. Finally, we found that social attraction could be predicted better from movement correlation features whereas for romantic and sexual interest mimicry features were better indicators. Additionally, we found that "mimicking of female to male" and "convergence of female's movement to male's" were indicators of sexual and romantic mutual attraction in our data.**

***Index Terms*—Attraction, synchrony, convergence, wearable acceleration, dyadic interactions, speed-dates.**

I. Introduction

Gatica-Perez defines the term interest as "people's internal states related to the degree of engagement displayed, consciously or not, during social interaction" [1]. He also notes that this engagement arises because of different factors such as interest in the topic of a conversation, attraction to other person or social rapport. The goal of this study is to investigate the automatic detection of attraction in dyadic interactions using movement features that are automatically extracted from single body-worn accelerometers in an in-the-wild setting.

The advances in sensing technologies and the possibilities of sensing human behavior have brought interest in the automatic assessment of human behavior in several research communities. Assessing human behavior makes it possible to automatically analyze human-human interactions. This in turn makes it possible to build tools that improve the time and possibly quality of psychological and sociological research. Additionally, automatic assessment is of interest for the creation of more naturally behaving socially-aware computing systems. A further application is the creation of tools that can help people assess their own behavior in their relationships, enabling them to receive feedback about their behavior during social interactions which would increase the quality of their relationships with other people.

Recent promising advances in this field give insights into the relationship between little-understood phenomena like physical and emotional attraction and measurable human behavior. Attraction has been found to affect the way in which couples behave towards one another during interactions, affecting other known social phenomena like the level of synchrony in their movements [2], the degree to which they mimic one another [3]–[5] and the adaptation to one another's behavior [6]. Our work aims to investigate how we can automatically estimate interpersonal attraction by quantifying body coordination using wearable sensors. Concretely, the contributions of this study are three-fold. First, we proposed novel behavioral coordination features that can be extracted from a single body-worn accelerometer. We show that behavioral features such as synchrony and convergence can be extracted from motion and used to predict attraction between two people. Second, we used these features to model interpersonal attraction and tested them in a real life in-the-wild setting with a less intrusive approach. Finally, we obtained experimental evidence that supports the existing theories from psychological literature about behavioral differences between men and women in a courtship setting.

II. Related Work

Interpersonal interest and its associated non-verbal behavior have been studied by psychology researchers. The automatic quantification of this behavior has also been of interest to computer scientists. Therefore, studies from both fields are reviewed here. First, related work about romantic interest is discussed. Research on other forms of interest is mentioned briefly. Finally measures of coordination that are used in the literature are summarized.

Most of the existing work that studied romantic interest conducted experiments in speed-date scenarios. The reason for using these events in attraction studies is that the responses to the questionnaires that are filled after the dates can be used as ground truth for the prediction tasks. Madan et al.

978-1-7281-3891-6/19/$31.00 ©2019 IEEE

investigated romantic, friendly and business interest between people by extracting four types of social signal measures from audio: activity, engagement, emphasis and mirroring and successfully predicted each type of interest using these features [7]. Michalsky et al. investigated pitch convergence and found that speech of interactants became more similar over the course of conversation when perceived attractiveness and likability is higher [6]. Ranganath et al. used prosodic, dialogue, and lexical features extracted from audio recordings to investigate the participant's flirtation behavior and could predict both flirtation intention and flirtation perception [8].

Veenstra et al. found that positional features extracted from video such as position, distance, movement and synchrony are indicators of attraction. Their results also indicated that addressing male and female behavior as two different tasks for prediction increased the task performances [2]. With the aim to recreate similar results, Cabrera-Quiros et al. attempted to classify attraction levels between participants using motion features extracted from accelerometer data [9]. Even though they only used the mean and variance of the magnitude of acceleration as features, they obtained good classification results. An interesting finding of their study is that separating male and female data did not improve their prediction performance unlike [2].

Research has been done in psychology about attraction focusing on the mimicry behavior [3]–[5]. Instead of doing an automatic feature extraction, they manually annotated nonverbal mimicry events of the interacting partners by watching the recordings of the interactions. They found that mimicry was positively correlated with romantic interest. Research from psychology also showed that people use different mechanisms and strategies when it comes to searching for short-term and long-term partners [10]. Moreover, these strategies are different for men and women. It is indicated also in other research that men and women show differences in mate selection [11] and courtship behavior [12] such as women flipping their hair and moving their shoulders and men uncrossing their legs often. These research suggest that men and women should be treated separately in attraction prediction tasks.

In research about other types of inter-personal interest, researchers studied head motion synchrony of spouses during interactions [13], [14], estimated team cohesion in meeting settings using audio-visual cues and mimicry features [15], [16] and used behavioral synchrony and correlation features to predict interaction quality and outcome [17], [18].

In conclusion, previous literature shows that features that capture behavioral coordination and similarity between people's behavior are indicators of affect between people and used for modeling interest by extracting them from different modalities and settings. These features can be grouped into two categories as *synchrony* and *convergence*. In this study we also used these types of features for modeling attraction.

III. Data

We used *MatchNMingle*, a multimodal and multisensor dataset recorded with the aim to be used in research about automatic analysis of social signals and interactions for both social and data sciences [9]. It was collected in an indoor in-the-wild setting instead of a lab setting. Therefore the social interactions between participants were as natural as possible.

A. Experiment context

The whole dataset was recorded during a set of activities taking place over 3 days in total in a local bar. Each day the event started with a speed dating round where participants of opposite sex had a 3 minute date with each other, followed by a mingle party. In this research, only the data from the first part of the event is used. Participants were recruited from a university and expected to fit the criteria of being single, heterosexual and between the ages of 18 and 30. In total of 92 participants attended the event, with equal number of men and women and most of them did not know each other. During the event, participants were asked to wear devices around their necks, which record tri-axial acceleration and proximity. After removing malfunctioning devices, in total 72 participants had sufficient data recorded by wearable devices.

B. Data collection

Acceleration data was collected using tri-axial accelerometers at a frequency of 20 Hz. After each 3-minute date with the participant of opposite sex, participants were given 1 minute to fill a booklet with a questionnaire about their date partner indicating their interest in each other. Responses for these questionnaires constitute the ground truth for the tasks in this study. After removing the dates which at least one of the participants have a malfunctioning device and unreadable booklet responses, a total number of 398 date interactions were left. Since each participant had their own label for each date, male and female participants of one date interaction were treated as separate samples, resulting in total of 796 samples.

C. Defining the ground truth

The questionnaire that participants filled after their dates consisted of following questions with responses on a 7-point Likert scale (low = 1, high = 7):

- How much would you like to see this person again?
- How would you rate this person as a potential friend?
- How would you rate this person as a a short term sexual partner?
- How would you rate this person as a long term romantic partner?

Each of these questions was used to define different tasks for the interest prediction problem as respectively *See Again*, *Friendly*, *Sexual* or *Romantic*. The problem was treated as a binary classification problem, meaning each date of a participant would have binary labels for each one of these tasks. For clarification, a *date* refers to the information from a single person during a speed date, whereas a *date interaction* refers to the interaction between two participants during a speed date. These two concepts are used in two challenges of this study. The first one is to predict if one participant is *attracted* or *not attracted* to his/her date partner. This would require labeling a person's *date* thus for each speed date interaction male and female participant have their own labels. The second challenge is to predict if a *date interaction* ends with a *match* or *no match*. To obtain labels for these situations, first the responses on Likert scale need to be binarized. Initially, each person's scores for all of his/her dates are normalized with z-score normalization. Following this, *dates* that have positive score are labeled as *attracted* and negative score as *not attracted*. Following this, *date interactions* are labeled and a *match* label is given to a *date interaction* if both participants have *attracted* labels for their date and for all other cases a *no match* label is assigned.

D. Data Analysis

Distribution of labels over each class showed that *SeeAgain* and *Friendly* tasks have a balanced class distribution with 49% positive labels. On the other hand, *Romantic* and *Sexual* tasks have a bias on *not attracted* class with 40% and 42% positive labels respectively. With the match labels, it is observed that for all tasks the datasets are highly biased towards *no match* class with 30% positive labels for *SeeAgain* and *Friendly*, 13% for *Romantic* and 19% for *Sexual* task. Additionally, the class distribution difference between genders is also analyzed and observed that *Friendly* and *Romantic* tasks have similar number of positive labels in female group and male group (f:52% m:48% for Friendly, f:38% m:42% for Romantic) whereas *SeeAgain* task has significantly more positive labels in female group and *Romantic* task has significantly more positive labels in male group (f:52% m:45% for SeeAgain, f:36% m:48% for Sexual). From these we can claim that female and male participants did not differ in their inclinations toward friendship and long term romantic relationship but male participants had a higher tendency towards a short term sexual relationship.

IV. Features

Our method aims to model the coordination of behavior between two people having a date, using nonverbal behavioral features extracted from accelerometer readings.We describe the feature extraction process in more detail below.

A. Preprocessing

The accelerometer data consists of 3-dimensional readings with the X axis capturing the left-right movements; the Y axis up-down movements and Z axis forward-backward movements. Initially each axis of each person's recordings is normalized by computing the z-score within itself to remove interpersonal differences in movement intensity. Then, these normalized raw recordings are treated in multiple ways: raw values of each axis, absolute values of each axis and the magnitude of the acceleration which is computed as $\sqrt{(x^2+y^2+z^2)}$. Each of these 7 signals is divided into *n*-second windows using a sliding-window approach, with *n/2* second shifts between each window. Since the optimal window size that captures necessary information is not known, the possible values of n are chosen as 1, 3, 5 and 10 seconds.

Similar to [19], statistical (mean, variance) and spectral (power spectral density) features are extracted from each window. Power spectral density (PSD) per window is computed using 6 logarithmically spaced bins between 0-10 Hz, to increase the resolution at low frequencies. Each bin gives information about the characteristic of behavior of the person at that time window, therefore each bin is treated as a single feature. Combining these features results in 8 feature dimensions per window.

Computing these 8 features for each 7 types of signal mentioned earlier and for 4 different window-sizes results in 224 low-level features that will further be used to extract behavioral coordination features that are explained in the following subsection. An illustration of the pre-processing steps is shown in Figure 1.

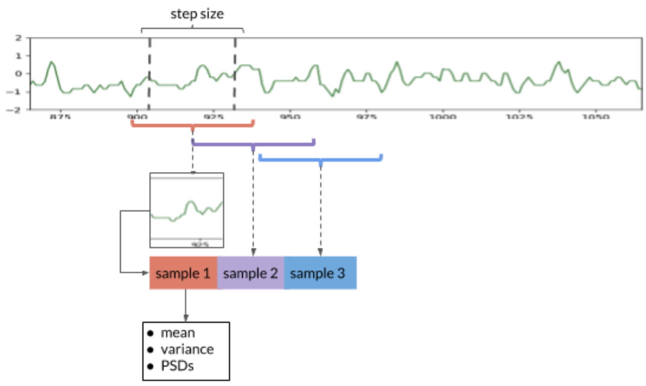


Fig. 1. Pre-processing step: Using a sliding window approach, the signal is divided into samples from which the statistical and spectral features are extracted.

B. Feature Extraction

The aforementioned low-level features are used to extract more complex behavioral coordination features that are grouped into two categories.

1) Synchrony: To measure the synchrony of behavior of two interacting people, two different measures are used.

a) Correlation: Correlation has been used in the literature as a measure of similarity of overall body motion and also motion of specific body parts such as the hands or head of two people [17], [18], [20]–[22]. Here, as in the previous studies, we used Pearson correlation:

$$\rho_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{\sigma(X)\sigma(Y)} \tag{1}$$

In our context, it captures the liner correlation between two person's signals and it is expected to give a score closer to 1 when two people have positive feelings towards each other.

b) (Normalized) Mutual Information: This measure has also been used in the literature to capture the dependence between two people's behavior [19], [23]. In our case it captures the dependence of two people's behavior on each other. It is calculated as follows:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{2}$$

where H(X) and H(Y) represents the entropy of random variables X and Y and H(X,Y) represents the joint entropy of X and Y.

Additionally, normalized mutual information is computed by dividing by $\sqrt{H(X)H(Y)}$ to obtain a score between 0 and 1. A higher score is expected when two people have an influence on each other's behavior.

c) Mimicry: This mimicry measure is inspired by the work of Nanninga et al. [16]. The goal is to capture when one person imitates their partner's behavior. Figure 2 illustrates how this feature is computed. Each sample window of Person A's signal is compared with the consecutive window of Person B's signal. To compare these windows, the distance between low-level features of these windows are computed, resulting in distance scores $D = [d_0, d_1, ...d_n]$ for the entire interaction. From these distance scores, minimum ($min(D)$), maximum ($max(D)$), mean ($mean(D)$) and variance ($var(D)$) are computed and used as features. Since this feature only captures the mimicry of Person B to Person A, the reverse is also computed.

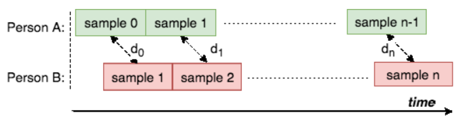


Fig. 2. Mimicry feature. Each time sample is compared with the other person's preceding time sample.

d) Time-lagged correlation: Correlation with a time lag has also been used to measure the linear relation of a follower's movement with the interlocutor's movement [20], [21]. The following formula computes the correlation between X and Y signals at a positive lag of τ samples, following formula is used:

$$\rho_{xy} = \frac{\sum_{i=1}^{N-\tau}(x_i - \mu_x)(y_{i+\tau} - \mu_y)}{\sigma(X)\sigma(Y)} \tag{3}$$

This metric can indicate the leader-follower relationship of two people in a conversation by showing who is driving the interaction. In an example case of measuring the correlation between person A and person B's movement, if a higher score is obtained when person B's signal is positively lagged, this indicates that person B is leading the interaction. Following the literature, we use +/- 1 time step lags.

2) Convergence/Divergence: To measure convergence, three different metrics are developed that, inspired in various literature. These features aim to measure if two people's behavior style is diverging or converging through their interaction. The idea is that people's feelings for each other would be more positive if they show a more converging behavior.

a) Symmetric convergence: This feature is inspired by the works of [6], [16]. It compares two people's behavior at each time step and aims to capture if the difference between their behavior decreases over time. In order to compute it, corresponding windows of two participants' signals are compared with each other. To measure the similarity at each time step, the distance between these corresponding samples' low-level features are computed as illustrated in Figure 3, resulting in distance scores $D = [d_1, d_2, ..., d_n]$, for each sample. After that, the correlation of these scores with time is computed to understand if they increase or decrease using Pearson correlation formula (eq. 3) and a correlation coefficient is obtained. Since the goal is to capture convergence, a decreasing distance indicates converging behavior. Therefore, the correlation coefficient is expected to be more negative for converging interactions where participants show similar behavior over the interaction.

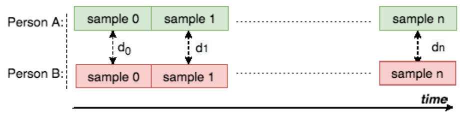


Fig. 3. Symmetric convergence feature. Each time sample is compared with the other person's corresponding time sample.

b) Asymmetric convergence: This feature has also been inspired by [16]. The first two minutes of the date interaction are taken as the learning period in which the behavior of one participant is modeled and the last one minute of the interaction is compared to this learned model. To understand if the second person's behavior converges to the first person's behavior, the low-level features of the samples from the last one minute are compared to the learning part's low-level features. To measure the similarity, distances between these features are computed as illustrated in Figure 4, resulting in distance scores $D = [d_1, d_2, ..., d_n]$, for each sample in the last one minute of interaction. Then, the correlation of these scores with time is computed using Pearson correlation. A smaller and negative correlation coefficient indicates high convergence and more positive affect between two people. Since this feature only captures the convergence of Person B's behavior to Person A, it is also computed by changing the order of people.

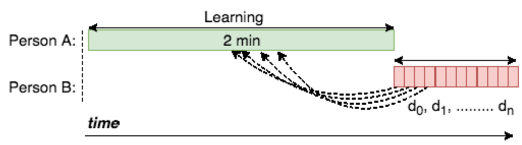


Fig. 4. Asymmetric convergence feature. Each time sample in the last 1 minute period is compared with the other person's first 2 minutes by computing a distance score between sample features.

c) Global convergence: This feature has been inspired by the work of [6]. The idea is to measure the similarity of two people's behavior in the beginning and at the end of their date interaction and compare these similarities. It is expected that the behavior will be more similar at the end of the interaction due to convergence. To capture this, the first and second half of the signals are taken as illustrated in figure 5. The similarity d_0 between the first half's features of the two

TABLE I
FEATURE VECTOR

Feature type	Feature	ID
Synchrony	Correlation	0-223
	Mutual Information	224-559
	Mimicry	560-2351
	Time-lagged correlation	2352-2799
Convergence	Symmetric convergence	2800-3135
	Asymmetric convergence	3136-3583
	Global convergence	3584-3807

persons is computed. An equivalent similarity d_0 is computed for the second half. The difference between these similarities is computed by subtraction as: $c = d_1 - d_0$. This difference is expected to be negative when convergence occurs.

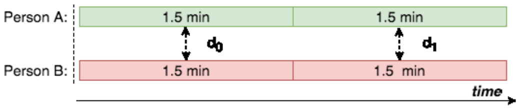


Fig. 5. Global convergence feature. The difference between both persons' features is computed for each half of the interaction.

Table I summarizes all the features that are used in our experiments with the corresponding IDs. Additionally, symmetric features are colored with red and asymmetric features are colored with blue.

C. Feature pre-processing

After extracting the features, they were pre-processed with the objective of reducing the dimensionality of the feature space. The features were first normalized to zero mean and unit standard deviation. Second, we selected a set of representative features by computing the ANOVA F-value between each feature and target labels and taking the features with significantly high F-value ($p < 0.05$). Finally, we applied principal component analysis (PCA) and the top principal components preserving 95% of the variance were kept.

V. RESULTS

A logistic regressor was chosen as classifier for the task of predicting interest, as in [24]. To evaluate the predictive performance of classifiers for each task, a nested 10-fold cross-validation was applied. To obtain a measure that is unaffected by the class imbalance, the Area under the Receiver Operator Characteristic (AUC) was used to determine performance.

The first problem investigated in this study was that of predicting if a person is attracted to his or her date partner. Performances for different attraction type predictions are compared to the random baseline classifier which assigned every data point to the most frequent class. Obtained mean AUC scores are shown in the first column of Table II. For all tasks our features performed significantly better than the random baseline of 50% AUC.

We also compared the performance of our features with the features from [9], the closest to our setting in terms of approach and modality. The examined features were: mean

TABLE II
MEAN AUC SCORES OBTAINED IN ONE-WAY INTEREST PREDICTION TASKS

	Our features	SOA features	Only female	Only male
SeeAgain	0.67($\pm$0.06)	0.53($\pm$0.05)	0.77($\pm$0.06)	0.76($\pm$0.07)
Friendly	0.73($\pm$0.05)	0.50($\pm$0.06)	0.75($\pm$0.08)	0.76($\pm$0.07)
Romantic	0.68($\pm$0.04)	0.57($\pm$0.08)	0.80($\pm$0.06)	0.79($\pm$0.04)
Sexual	0.69($\pm$0.07)	0.50($\pm$0.08)	0.75($\pm$0.10)	0.80($\pm$0.07)

TABLE III
MEAN AUC SCORES OBTAINED IN MUTUAL INTEREST PREDICTION TASKS

	Our features
SeeAgain	0.82($\pm$0.09)
Friendly	0.79($\pm$0.06)
Romantic	0.80($\pm$0.11)
Sexual	0.78($\pm$0.09)

of the magnitude of acceleration, variance of the magnitude of acceleration, mean of the variance of acceleration over 1 second windows, variance of the variance of acceleration over 1 second windows. Obtained mean AUC scores using these features are shown in the second column of Table II. As is seen they are out-performed by our features for all tasks. Even though the same dataset is used in [9], since to compute our features it is required to have valid data from both of the participants in a date, we had to discard a larger amount of dates. Moreover, we used a different method for obtaining the ground truth from the questionnaires, resulting in a dataset with different statistics.

The second problem investigated in this study was predicting if both people who had an interaction are attracted to each other or not (ie. if they are a match). Obtained mean AUC scores are shown in Table III. We observe that the mutual attraction prediction tasks have shown better performance than one-way attraction prediction tasks. We could not compare our results with the state-of-the-art features' performance because they did not use their features for predicting mutual attraction.

As in previous literature [2], [9], we also experimented separating and combining male and female data. The third and fourth columns of Table II shows the scores obtained by using male and female data separately. The results showed that separation increased the prediction performance for all tasks compared to the combined dataset's performance. The least amount of improvement is seen in the *Friendly* tasks indicating that men and women show similar behavior when they have a friendly attraction towards each other. On the other hand for *Romantic* and *Sexual* tasks we have a larger increase, in line with the literature, suggesting that men and women show different behavioral characteristics when experiencing romantic or sexual attraction. Similarly, for the *SeeAgain* task a performance increase is observed in the separated case.

VI. DISCUSSION

A. Correlation analysis

In this analysis, features are correlated with the label of each task, in order to have a deeper understanding of the contribution of each one. The features with the highest correlation coefficients were found to vary with different tasks. This indicates that each type of attraction manifests in different behavioral characteristics. One interesting finding of this analysis is that *Correlation* features that are computed over the Z-axis are negatively correlated with *Friendly* attraction as opposed to the expectation of positive correlation as explained before. Z-axis captures the forward-backward acceleration of the body. Therefore, negative correlation could be because of one person's backward and other person's forward movement occurring simultaneously. Considering the fact that during the interactions people were sitting opposite to each other, this might be due to people's simultaneous movement occurring along the same axis but in different directions. On the other hand, most of the *Correlation* features that are extracted using PSD bins indicating correlation in the movement frequencies of couples showed positive correlation with the *Friendly* and *Sexual* attraction. It shows that the correlation of movement did not occur necessarily in the direction of movement but also the frequency of movement of two people were similar.

It is also seen that *Mutual Information* features had high positive correlation with only the *SeeAgain* and *Friendly* tasks whereas the *Mimicry* features had high positive correlation with only the *Romantic* and *Sexual* tasks. From that, we hypothesize that people enjoying a friendly conversation show more synchronic behavior. On the other hand, in a flirtatious interaction mimicking behavior becomes more prominent.

Convergence features were expected to have negative correlation with the labels, because more negative convergence values indicate a higher convergence, which we hypothesize to be an indicator of positive attraction. This is obtained in *SeeAgain* and *Romantic* tasks but the opposite is observed in most features of *Friendly* task. We can conclude that actually the divergence in behavioral characteristics might be an indicator of friendly attraction to the other person.

Another observation is that there are not many features with high correlation for the *SeeAgain* task. This can be because the labels for this task are obtained by the answers to "wanting to see the person again" question and this is a vague description for any attraction. Therefore the ground truth might not be a very clear indicator of any attraction.

When analyzing the high correlated features with *Match* tasks, we could also pay attention to the direction of asymmetric features that will give us information about the leader-follower behavior of each gender. In *SeeAgain* and *Friendly* tasks, *male convergence to female* features have a positive correlation with matches. This is the opposite of what is expected because a positive convergence score shows a non-converging behavior and it was not expected to be correlated with attraction. On the other hand, in the *Sexual* task *female convergence to male* feature shows a negative correlation with attraction, indicating that when there is a mutual sexual attraction during an interaction it is revealed by female's behavior converging to the male's. Additionally, *female mimicking male*-mimicry features show the highest positive correlation with mutual *Romantic* and *Sexual* attraction. Interestingly even though both are synchrony features, *Correlation* shows higher correlation with *Romantic* matches whereas the *Mutual Information* shows higher correlation with *Sexual* matches.

In summary, we see that different types of attraction are indicated with different behavioral features. Therefore, depending on the task, a subset of features can be selected for better prediction performance. Even though we can not be certain that our features are capable of modelling exactly the behavior that we aim to model since we do not have ground truth about the synchrony and convergence, they are shown to be good at predicting the goal outcome which is to predict interpersonal attraction. In addition, synchrony and convergence terms are not very well defined and there is no consensus over them in the literature which makes it even more difficult to generalize our findings.

B. Comparison symmetric and asymmetric features

As is known attraction is an asymmetric property, meaning that it might not be reciprocal. Therefore, it is important to consider that symmetric and asymmetric features may have different meanings. Symmetric features are the same for both participants of the date. On the other hand, asymmetric features can capture the direction of mimicry or convergence and thus have a different meaning for each participant in a date. For example, if the male participant is mimicking the female, this can be interpreted differently for each side.

We ran classification tasks using symmetric and asymmetric features. AUC scores of one-way attraction prediction tasks using only symmetric features are 0.59($\pm$0.05) for SeeAgain, 0.65($\pm$0.03) for Friendly, 0.61($\pm$0.05) for Romantic, 0.58($\pm$0.06) for Sexual tasks and using only asymmetric features 0.64($\pm$0.05) for SeeAgain, 0.67($\pm$0.05) for Friendly, 0.68($\pm$0.05) for Romantic and 0.65($\pm$0.06) for Sexual. In mutual-attraction prediction tasks both feature groups showed similar performance. Therefore we did not add their results here. Our results show that asymmetric features outperformed the symmetric features in all tasks for predicting one-way attraction. Considering the fact that attraction is an asymmetric property, these results align with our expectations that attraction is estimated better using asymmetric features.

VII. Conclusion

In this study, we have presented a method for automatically predicting attraction between people using behavioral coordination features extracted from data recorded by a single body-worn accelerometer. We used synchrony and convergence features and our proposed approach out-performed the state-of-the art [9] which was the most similar setting to our case. To our knowledge, this is the first time that motion convergence is used for estimating interpersonal attraction. Our results also showed that prediction performance increases when male and female data is separated, indicating that men and women have different behavioral characteristics when showing attraction. Finally, we ran a further analysis to investigate the individual feature contributions and found that different attraction types are indicated by different type of features.

REFERENCES

[1] D. Gatica-Perez, "Modelling interest in face-to-face conversations from multimodal nonverbal behaviour," *Multimodal Signal Processing*, 03 2012.

[2] A. Veenstra and H. Hung, "Do they like me? using video cues to predict desires during speed-dates," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Nov 2011, pp. 838–845.

[3] N. Guguen, "Mimicry and seduction: An evaluation in a courtship context," *Social Influence*, vol. 4, pp. 249–255, 10 2009.

[4] J. C. Karremans and T. Verwijmeren, "Mimicking attractive opposite-sex others: The role of romantic relationship status," *Personality and Social Psychology Bulletin*, 2008.

[5] S. D. Farley, "Nonverbal Reactions to an Attractive Stranger: The Role of Mimicry in Communicating Preferred Social Distance," *Journal of Nonverbal Behavior*, vol. 38, no. 2, pp. 195–208, 2014.

[6] J. Michalsky and H. Schoormann, "Pitch convergence as an effect of perceived attractiveness and likability," in *INTERSPEECH*, 2017.

[7] A. Madan, R. Caneel, and A. Pentland, "Voices of attraction," 2004.

[8] R. Ranganath, D. Jurafsky, and D. McFarland, "It's not you, it's me: Detecting flirting and its misperception in speed-dates," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1*, ser. EMNLP '09. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 334–342.

[9] L. Cabrera-Quiros, A. Demetriou, E. Gedik, L. van der Meij, and H. Hung, "The matchnmingle dataset: a novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates," *IEEE Transactions on Affective Computing*, 2018.

[10] D. Buss and D. Schmitt, "Sexual strategies theory: An evolutionary perspective on human mating," *Psychological review*, vol. 100, pp. 204–32, 05 1993.

[11] R. Fisman, S. Iyengar, E. Kamenica, and I. Simonson, "Gender differences in mate selection: Evidence from a speed dating experiment," *The Quarterly Journal of Economics*, vol. 121, pp. 673–697, 02 2006.

[12] K. Grammer, M. Honda, A. Juette, and A. Schmitt, "Fuzziness of nonverbal courtship communication: Unblurred by motion energy detection," *Journal of personality and social psychology*, vol. 77, pp. 487–508, 10 1999.

[13] B. Xiao, P. G. Georgiou, C.-C. Lee, B. Baucom, and S. S. Narayanan, "Head motion synchrony and its correlation to affectivity in dyadic interactions," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[14] B. Xiao, P. Georgiou, B. Baucom, and S. S. Narayanan, "Head motion modeling for human behavior analysis in dyadic interaction," *IEEE Transactions on Multimedia*, 2015.

[15] H. Hung and D. Gatica-Perez, "Estimating Cohesion in Small Groups Using," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 563–575, 2010.

[16] M. C. Nanninga, Y. Zhang, N. Lehmann-Willenbrock, Z. Szlávik, and H. Hung, "Estimating verbal expressions of task and social cohesion in meetings by quantifying paralinguistic mimicry," *Proceedings of the 19th ACM International Conference on Multimodal Interaction - ICMI 2017*, pp. 206–215, 2017.

[17] F. Ramseyer and W. Tschacher, "Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome," *Journal of consulting and clinical psychology*, vol. 79, no. 3, p. 284, 2011.

[18] W. Tschacher, G. M. Rees, and F. Ramseyer, "Nonverbal synchrony and affect in dyadic interactions," in *Front. Psychol.*, 2014.

[19] E. Gedik and H. Hung, "Detecting conversing groups using social dynamics from wearable acceleration: Group size awareness," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 4, pp. 163:1–163:24, Dec. 2018.

[20] Y. Hart, E. Czerniak, O. Karnieli-Miller, A. E. Mayo, A. Ziv, A. Biegon, A. Citron, and U. Alon, "Automated video analysis of non-verbal communication in a medical setting," *Frontiers in Psychology*, vol. 7, no. AUG, pp. 1–8, 2016.

[21] A. S. Won, J. N. Bailenson, and J. H. Janssen, "Automatic detection of nonverbal behavior predicts learning in dyadic interactions," *IEEE Transactions on Affective Computing*, vol. 5, no. 2, pp. 112–125, 2014.

[22] A. S. Won, J. N. Bailenson, S. C. Stathatos, and W. Dai, "Automatically detected nonverbal behavior predicts creativity in collaborating dyads," *Journal of Nonverbal Behavior*, vol. 38, no. 3, pp. 389–408, Sep 2014.

[23] C. Martella, E. Gedik, L. Cabrera-Quiros, G. Englebienne, and H. Hung, "How was it?: Exploiting smartphone sensing to measure implicit audience responses to live performances," in *Proceedings of the 23rd ACM International Conference on Multimedia*, ser. MM '15. New York, NY, USA: ACM, 2015, pp. 201–210.

[24] L. Cabrera-Quiros, E. Gedik, and H. Hung, "Estimating self-assessed personality from body movements and proximity in crowded mingling scenarios," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ser. ICMI '16. New York, NY, USA: ACM, 2016, pp. 238–242.