Do They Like Me? Using Video Cues to Predict Desires during Speed-dates

Arno Veenstra and Hayley Hung
University of Amsterdam
Science Park 904 1098 XH Amsterdam The Netherlands
arno.veenstra@student.science.uva.nl, h.hung@uva.nl

Abstract

In this paper we introduce video features which are used to predict if people want to exchange contact information with the other in a speed-date, we also use these features to predict how physically attractive participants found their dates. Previous work on predicting and interpreting speed-dates has focused mainly on the audio channel. We use automatically extracted features related to position, proximity and motion. This paper shows that these features can be used to significantly outperform the baseline and have comparable performance to audio-only systems. The data used has been gathered from a real speed-date event, involving 16 participants. Experiments were carried out on 64 speed-dates lasting 5 minutes. The best performance on prediction exchanging contact information was 72% and 70% accuracy for males and females respectively, and 70% for both genders when predicting physical attraction.

1. Introduction

Finding a partner can be difficult, although a lot of people succeed, millions of people seem to have difficulties finding the right partner. Dating services are abundant and together make a multi-million industry. Ranging from online dating websites and speed-date events to mail-order brides. One of the reasons people can have a hard time finding the right partner is that interpersonal communication in general and dating in particular is often guided by misperception and misinterpretation [10, 17].

Because of this it is interesting to see if machines can potentially understand and predict human behavior when people are interacting with a potential partner. For instance Ranganath *et al*. [17] have shown that in some cases machines can even outperform humans in predicting flirtatious behavior. An application of this could be to build a system which provides humans with feedback about their own behavior, but also about the behavior of the person they are interacting with. See for instance the work of Madan *et al*. [14] who implemented such a system on a smartphone. This system gives the user immediate feedback about the probability of the other person wanting to exchange contact information. An interesting aspect of giving humans this feedback is that they might be able to discover things about their own unconscious feelings about the other. It is of course also nice to get insights into what the other person is feeling about you.

Previous research on automatically analysing speed-dates has been performed using audio data [14, 17]. However, findings from behavioral psychology indicate that there are quite a few important cues related to attraction and liking that cannot be captured from audio. These cues include for instance *proximity*, *orientation*, *gaze*, *posture* and *touch* [2, p. 88]. Therefore it is interesting to see if and how video features can be used in this context. This paper makes a step towards this by addressing the role of automatically extracted proximity-related cues.

The contribution this paper makes is that we show that the video channel can indeed be used to predict whether or not a person wants to exchange contact information with the other after a speed-date and predict whether or not a person is physically attracted to the other. We show that positional information acquired from a simple tracking method can be used to extract useful features for these tasks.

The rest of this paper is built up as follows. In section 2 we give a short overview of related work in social signal processing in general and about speed-dates in particular. Then in section 3 we explain how the data was collected. In section 4 we explain which features were extracted and how this was done. In section 5 the experimental setup will be discussed. In section 6 the results will be presented and finally we end with a conclusion and discussion of future work in section 7.

2. Related Work

Automatically analysing speed-dates fits into the broader research area of *social signal processing*, which studies the automatic recognition of social signals and social behaviors in order to equip computers with so called *social intelligence* [7, 16, 19]. Research in this area includes, amongst other things, measuring who is most dominant in a group meeting [1, 12], predicting outcomes of negotiations [6], speed-dates [14, 17] and automatically recognizing agreement and disagreement [3].

The first attempt to automatically interpret speed-dates was done by Madan, Caneel and Pentland [14]. They used four measures extracted from audio: activity, engagement, emphasis and mirroring. The audio was recorded by using PDAs with headsets. Using these measures they were able to predict if participants in a speed-date event were willing to provide contact information to the other participant. Using a two-class linear classifier they were able to predict this with an accuracy of 71%. Using a support vector machine as a classifier did not increase the accuracy significantly.

Other work on the automatic analysis of speed-dates was done by Ranganath *et al*. [17]. They also focussed on speed-dates, but their task was different. Instead of focussing on predicting if people would be willing to share contact information, they looked at the difference between intention and perception when flirting. They showed that there is quite some misperception involved and people tend to perceive signals from others more like the signals they intend to send out themselves. They showed that someone's perception of their own flirting is highly correlated (r=0.73) with their perception of the other person's flirting. An important contribution the paper made was that they showed that machines can outperform humans in recognizing flirtatious behavior. The system outperformed men in detecting female flirting (71.5% versus 56.2%) and also outperformed women in detecting male flirting (69% versus 62.2%). To achieve this they also used audio features. The features were grouped into prosodic, dialogue and lexical features. The prosodic features consisted of F0 and RMS amplitude features. Dialogue features included the number of turns, laughter, disfluencies, such as 'um' and 'uh' or sentence restarts, and overlap, the number of turns in which the two speakers spoke at the same time. The lexical features consisted of autoencoded words, so no hand labelled set of positive and negative words was needed. An important difference with the work of Ranganath *et al*. [17] and that of Madan *et al*. [14] is that they also looked at verbal behavior. While this of course might yield valuable information it also makes it language dependent.

As said, the major difference between the work of Madan *et al*. [14] and Ranganath *et al*. [17] and the work presented here is that they focused on the audio channel. While the audio channel is of course a valuable source of information it also has some disadvantages. For instance there might be privacy concerns from the participants, as everything they say is recorded. Another thing is that putting microphones on the table or even attaching them to clothes is more intrusive than hanging cameras in the room, especially in our case where the cameras were put on the ceiling. So no frontal facial features of participants were captured.

Previous work on automatically analysing opposite sex encounters from the video channel does exist, but was not in the setting of a speed-date and the task was quite different. Grammer *et al*. [8] looked into opposite sex encounters and different ways of assessing non-verbal behavior. In their experiment two persons from the opposite-sex were brought together in a room and were asked to wait while the experimenter left for ten minutes. When the experimenter came back they were separated and asked to fill out a questionnaire which included questions about the person waiting with them. They introduced what they call *automatic movie analysis* to analyse these encounters. This method uses the video channel and uses motion based features to extract a measure called movement quality. This measure consists of the number of movements, duration, size, speed and complexity. Their method uses motion images by subtracting subsequent video frames. Using this method they showed that female interest covaries with movement quality. A drawback here is that it was not approached as a classification task, so no accuracy or related performance measures are given.

Another example of work that was similar in methods, but differed in the task at hand was given by Zen *et al*. [20]. Just like the work presented here they looked at positional information. They investigated the relation between proxemics and personality traits. Proxemics is a term introduced by Edward T. Hall [9] and it describes the distance and associated interactions that can occur between a person and someone they are interacting with. He distinguished different regions around a person which play different roles when interacting. These regions were called the public space, social space, personal space and intimate space. Zen *et al*. used an informal party setting in which people interacted with each other to see how the use of personal space can be a predictor of extraversion and neuroticism. They used fixed cameras to do tracking and used Pan Tilt Zoom cameras to get higher resolution images of specific targets. They used the minimum distance towards others, the velocity (which they defined as the variation in position every 2 seconds) and the number of intimate, personal and social relationships as features. They used these features to train two SVM classifiers, one for extraversion and one for neuroticism. They were able to predict extraversion with an accuracy of 66% and neuroticism with an accuracy of 75%.

One of the reasons why it makes sense to look into proxemics when studying attraction is given by psychological research. Studies have shown that cues related to people's position in space are important when looking at attraction. For instance Heslin and Patterson [11] state that a smaller interpersonal distance is amongst the most promising cues when studying attraction. Argyle [2] notes that proximity and orientation are important cues for liking and sexual attraction. People who like each other or are sexually attracted to each other show a closer proximity and are more often leaning to each other, when seated. The orientation is often more direct, but side by side in some cases.

3. Data Collection

The data used in this research was collected during a speed-date event. In total 16 people participated in the experiment, 8 males and 8 females. All males dated with all females resulting in 64 dates. The participants had an age between 20 and 29, with an average of 23.4. Participants were recruited from the student population and were encouraged to bring friends of the same sex. Because of this most participants were students and a few were professionals. The dates lasted 5 minutes and after each date the participants were required to fill in a questionnaire regarding the interaction they just had. We asked if they wanted to exchange contact information with the other person and whether they thought the other was physically attractive. If both people wanted this, we provided them with the e-mail address of each other.

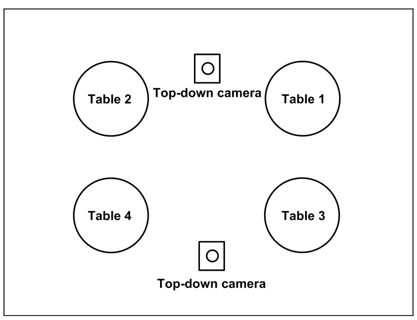

Figure 1. Overview of the room

The participants were standing round high tables during the dates. This gave them more freedom in positioning themselves as compared to a seated setting. Psychological research indicates that a standing arrangement is more sensitive to proxemic behavior than a seated arrangement [13].

The room was equipped with two top-down cameras approximately 5 meters from the floor. The cameras recorded at a frame rate of 20 fps. Four dates occurred simultaneously and each top-down camera captured half of the room, thus two dates and four people were present in each camera stream. See figure 1 for an overview of the room.

4. Feature Extraction

The overall procedure of extracting features was as follows. First positional information was extracted using a combination of background subtraction and clustering. After that several features were calculated from these positions. We describe the feature extraction process in more detail below.

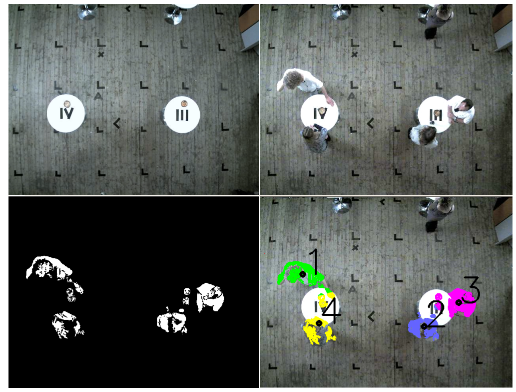

Figure 2. Overview of position extraction. Top-left: eigenbackground. Top-right: original input image. Bottom-left: Result of subtracting the eigenbackground from the input image. Bottom-right: clustered foreground with centroids (black dot). Each color represents a different detected person.

4.1. Extracting Positions

Using top-down cameras has the advantage that occlusions are rare, and tracking is therefore relatively easy. Because of this a fairly simple tracking method was used. The aerial view also preserves the privacy of the participants as their faces are not readily captured. The first step is to take a region of interest around the two tables in each frame to make sure we only see the four people we're interested in. The next step is to find the people in the image. This is done by creating an eigenbackground from hand selected frames which are empty. Using an eigenbackground instead of just a single background frame allows us to get rid of some variance and noise in our background image. When we subtract the eigenbackground from the frame and threshold it, we are left with foreground blobs. This can be seen in the bottom-right image in figure 2. This works well in most cases, although sometimes parts of people are considered as background because the color of their clothing or hair is very similar to the background. After this we cluster the segmented frame using k-means clustering. Five clusters were used; one for each person and one for the noisy pixels that should belong to the background. The centroids of the clusters were taken as the 2D position of each person in the image plane. See figure 2 for an overview of the steps.

To associate the ID of each person with their corresponding video data, hand labelled associations of the initial position of each cluster were used. Tracking was performed by looking for the closest corresponding cluster centroid in the previous frame. Also a sanity check was performed which checks if the newly found positions were very far from the previous ones. If this was the case no update was made and the previous positions were taken as the current. This is sometimes needed when k-means finds the wrong clusters because of small irregularities in the video stream. We also smoothed the resulting position to cancel out some noise in the measurements. This was done by using a sliding window with a width of 15 frames.

4.2. Positional Features

Several features were extracted using the positional information described in the previous section. These features can roughly be grouped in the following categories: position, distance, movement and synchrony. See table 1 for a summary of the extracted features. We provide further details about features that are not self-explanatory below.

4.2.1 Position

DIFANGLE is the difference in the angle both persons have with the table. It represents the angular proximity between the participants. This can range from 0 to 180 degrees. Where 0 degrees would mean both persons are standing at exactly the same side of the table and 180 degrees means the are on direct opposite sides. This feature is used because it should give us some information on how the people are positioned with respect to each other. A nice example of this can be seen in figure 3. Note that only the position of the person is used and not their body orientation.

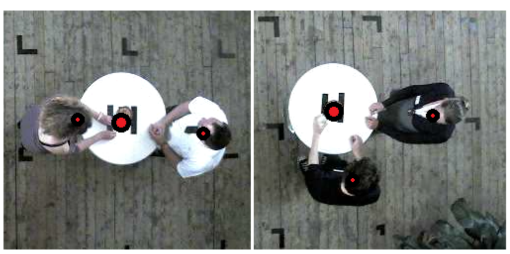

Figure 3. An example of two different values of DIFANGLE. The left images correspond to an angle of about 173 degrees. The right correspond to an angle of about 80 degrees. The two centroids (indicated by black and red circles) were found using k-means clustering and the table centers were hand labelled.

4.2.2 Movement

DECRDIS is the difference between the average euclidean distance in the first n frames and the last n frames. We expect that people who wanted to exchange contact information had a nice date and were therefore standing closer to each other at the end of the date. We also look at the beginning of the date to take into account how close they were standing when the date just started and they are most likely to not have made a decision yet. Figure 4 shows how this is calculated. x1 represents person x at the beginning of the date, x2 represents person x at the end of the date. The same goes for person y. To calculate DECRDIS we subtract d2 from d1. We chose n to be 250, which is about 12.5 seconds. In practice varying n did not appear to effect the performance significantly.

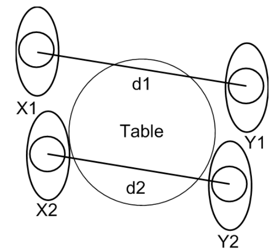

Figure 4. Decrease in distance is calculated as the difference between the distance in the beginning of the date (d1) and the distance at the end (d2). x1 is the mean position of person x in the first n frames. x2 is the mean position of person x in the last n frames.

Position	
AVG-DIFANGLE	Average angle between participants with respect to table
Movement	
VARDIS	Variance in distance
VAR-DIFANGLE	Variance in angle between participants with respect to table
VARPOS	Variance in position
VARPOS-OTHER	Variance in position of the other
DECRDIS	Decrease in distance
MOVDISTR	Movement distribution
MOVDISTR-OTHER	Movement distribution of other person
Distance	
AVGDIS	Average distance
Synchrony	
MOTIONSYNC	Synchrony in motion
MOTION-REACTION	Distribution of motion reaction
MOTION-REACTION-OTHER	Distribution of motion reaction of the other

Table 1. Overview of the used features

MOVDISTR represents how often someone moves in a particular direction. This direction is taken relative to the other person. Figure 5 shows how this is calculated using two consecutive frames. x1 is the person of interest at the current frame, y1 is the other person in the date at the current frame. x2 is the person of interest in the next frame. The distance between x1 and x2 is the distance travelled and the angle between the vectors x1-x2 and x1-y1 represent the direction with respect to the other person. This angle is the absolute angle, so it ranges from 0 to 180 degrees, where 0 degrees means moving towards the other person, and 180 degrees means moving away from the other person. A histogram is accumulated by calculating this angle at each frame, weighted by the euclidean distance between x1 and x2. Finally the mean and variance are taken of this histogram. The mean represents whether or not someone moves more towards or away from the other. The hypothesis is that if a person wants to exchange contact information with the other person, he/she is more likely to move towards the other. The variance gives an idea about whether or not someone is moving mostly in the same direction. For instance if someone would move only towards and away from the other, then the histogram would have high values in the bins round 0 and 180 degrees and low values round 90 degrees. This would result in a higher variance compared to someone who would move to all sides equally, as this person's histogram would be relatively flat. Note that this feature is dependent on the person. We also include the movement distribution of the other person. So we used information about how the person of interest moved, but also about how the person he/she was dating with moved. The same goes for the other asymmetric features VARPOS and MOTION-REACTION (described in section 4.2.3).

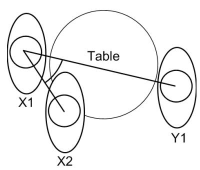

Figure 5. Information used to create the movement histogram. x1 and y1 are the two persons in the date in the current frame. x2 is the position of the person of interest in the next frame. The euclidean distance between x1 and x2 is the distance travelled, the angle represents the direction of movement with respect to the other person. Note that the body orientation of each person is shown for illustrative purpose only, but is not taken into account in the measurements.

4.2.3 Synchrony

Synchrony has often been found to be an important aspect of interactions. A well known paper on this is written by Chartrand and Bargh [5] who showed that nonconscious mimicry of one's interaction partner facilitates the smoothness of the interaction and increases liking between the people interacting. Mirroring has also been used by researchers in social signal processing [14, 18]. In order to measure mirroring in our context we extracted two features. One captures how often the two people move at the same time, the other captures how they react to each other.

We expect that if people actively mirror each other we should be able to find peaks of movement at roughly the same time. To calculate this we first determine the amount of motion per frame. This is done by looking at the difference in a person's position between consecutive frames. We then accumulate this over a window of 1 second. We slide this window over the whole date, one frame at a time, so we get the amount of motion per second. We then make a 2 dimensional histogram of this information from both persons in the date. This histogram has four bins, one which counts how often seconds of low activity occur at the same time, two bins which check how often a second of high activity in one person co-occurs with low activity in the other, and one bin that checks how often seconds of high activity in one person co occur with seconds of high activity in the other. We expect that people who wanted to exchange contact information will have higher values in the last bin. For our experiments all bins were used and correspond to the feature called MOTIONSYNC.

Besides looking if people move at the same time we also are interested in how they react to each other. We do this by looking at how the distance with a previous position of the other varies. This is done in a window of 1 second, so 20 frames are used. This gives us a distribution of 20 bins. The first bin represents the distance between x and y at time *t*. The second bin represent the distance between x at time *t+1* and y at time *t* and so on. Finally this is averaged over all 20-frame windows in the date. By fixing the position of y, we make sure that the change in distance is only based on the movement of x, so this way we know how x reacts. We called this feature MOTION-REACTION, as illustrated in figure 6.

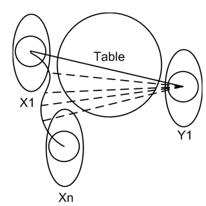

Figure 6. MOTION-REACTION is calculated by looking at the distribution of distances that exist between x1-y1 and xn-y1.

5. Experiments

In the experiments we wanted to predict if a participant wanted to exchange contact information with the other. We chose specifically this task for a couple of reasons. Firstly this task stayed really close to what the participants were asked to do during the experiment. They really got the contact information of the other in the case of a match, so because of this there was a good incentive for the participants to make a careful decision. Secondly it's a good thing to be able to compare our result with the work of Madan *et al*. [14] who ran a similar experiment with the same classification task. We also investigated perceptions of physical attraction of their date. To do this we used the interpersonal attraction scale by McCroskey and McCain [15]. We used three statements of this scale to get a score of how physically attracted the person was. The statements were: *He/she is somewhat ugly*, *I don't like the way he/she looks* and *I find* *him/her very attractive physically*. Since a 7-point Likert scale was used the score could range from 3 to 21, we binarized this by making every score of 11 and above a negative example (not physically attractive) and making everything below 11 a positive example (physically attractive).

To make our predictions we took a supervised learning approach and used two methods: the Support Vector Machine (SVM) and k-nearest-neighbor (kNN). SVMs are considered one of the best off-the-shelf classifiers and kNN is very popular because of its simplicity and performance. It is important to note that the purpose of this study is to investigate what type of automatically extracted proxemic cues can be used, rather than studying different classification methods.

Since we have 64 dates and we want to make predictions about individuals, this gives us 128 datapoints. Similar to [17] we split the classification task by gender. This allows us to create different models for each gender. The main reason to do this is that psychological and biological research indicates there are quite some differences in mate-selection and courtship behavior in males and females [4, 8].

During the experiments we used leave-one-out cross validation. This way all but one of the data points are used as training data. Also we scaled all features to have zero mean and unit variance, this way features with higher absolute values are treated as equally important with features with lower absolute values. For the SVM a radial basis function kernel was used and for the kNN classifier k was set to 3.

6. Results

Experiments were carried out using all the features described in table 1. We also tested the different categories by fusing all the features belonging to a category. The performance using male and female data on the task of predicting if people want to exchange contact information is shown in table 2. The performance on predicting physical attraction is shown in table 3. The baseline reported was calculated by labelling all test data points as the most frequent class. For the task of predicting exchanging contact information this meant labelling them as not wanting to exchange contact information. For the task of predicting physical attraction this resulted in labelling them as finding the other physically attractive. Note that the baseline for the females was slightly higher than for the males in the exchanging contact information task, as more females did not want to exchange contact information. For the physical attraction task it is the other way around, as more males found their female date physically attractive.

6.1. Exchanging Contact Information

Looking at the results for exchanging contact information task, one of the first things to notice is that in numerous cases the baseline is outperformed. In a couple of cases the baseline is outperformed by more than 10% and in one case with almost 19%. This shows that positional information from the video channel can be used to predict if someone wanted to exchange contact information in our speed-date data. Even when this positional information is gathered automatically with relatively simple tracking methods and is therefore quite noisy.

	Male		Female	
	SVM	kNN	SVM	kNN
All	0.44	0.48	0.50	0.56
Positional	0.44	0.44	0.48	0.45
AVG_DIFANGLE	0.44	0.44	0.48	0.45
Movement	**0.59**	0.48	0.48	0.44
VARDIS	**0.72**	**0.58**	0.47	**0.66**
VARDIFANGLE	**0.61**	0.47	0.45	0.45
VARPOS	**0.58**	0.48	0.56	0.47
VARPOS-OTHER	0.50	0.47	0.36	0.48
DECRDIS	**0.58**	**0.55**	0.55	0.50
MOVDISTR	0.52	0.42	**0.59**	0.52
MOVDISTR-OTHER	0.44	0.50	**0.70**	**0.58**
Distance	0.45	0.45	0.53	0.50
AVGDIS	0.45	0.45	0.53	0.50
Synchrony	0.48	0.41	0.53	0.55
MOTIONSYNC	0.45	0.53	0.55	0.55
MOTION-REACTION	0.45	0.45	0.53	0.50
MOTION_REACTION-OTHER	0.45	0.45	0.53	0.53
Baseline	0.53		0.56	

Table 2. Results on predicting exchanging contact information for males and females using a support vector machine and k-nearest-neighbors. The baseline reported was calculated by labelling all test data points as not wanting to exchange contact information. Boldfaced results outperform the baseline.

Another thing to notice is that the performance for men is better. For men the baseline is outperformed in 7 cases, while for females this is only the case in 4 cases. One reason for this might be that the naive baseline for the female case is higher than for the male case. Another explanation might come from the theory described by Grammer *et al.* [8] that states that females perceive a higher risk during opposite-sex encounters. This risk stems from the asymmetry in investment males and female have with respect to the offspring. It can be argued that female decision making is therefore more complex, females have to select a good father, while males have to compete for females and are therefore less picky.

When looking at the different categories of features we see that the movement features outperform the baseline in both the female and the male case. Although this is only statistically significant for the variance in distance for men ($p < 0.05$), this finding is in concordance with the results of Grammer *et al.* [8].

The feature which gives the best performance is the variance in distance for men, which performs significantly better than the baseline. Closer inspection of the data shows that men want to exchange contact information more often in dates where there is a high variance in distance.

Madan *et al*. [14] reported an accuracy of 64% for males and 72% for females. So our results are similar in performance. Note that the results of Madan *et al*. were obtained by using individual body-attached sensors for each person, while our results are based on only one sensor which was positioned approximately 3 meters away from the participants.

We asked participants if they thought their date wanted to exchange contact information with them. It turns out that people are quite uncertain about this and answer with *maybe* in 97 of the 128 cases. In the cases where they do answer this question with *yes* or *no*, they are only correct in 45% of the cases. So even in the cases where they are quite certain they still perform worse than chance.

6.2. Physical Attraction

For the physical attraction task the baseline is also outperformed in numerous cases (see table 3), even more often than when predicting exchanging contact information. It is interesting to see that for this task females are easier to predict than males. In the male case the baseline is outperformed 6 times while for females this happens 16 times. An explanation for this might be the difference in baseline, though if the same baseline is used for the female case 9 features are still above this.

Another thing to notice is that for this task movement features perform well. Also the synchrony based features result in above baseline performance in a few cases. For females the average difference in angle with respect to the table also gives good results. Inspection of the data tells us that in cases where the female found the male physically attractive the average angle between them was smaller. Another individual feature which gives good results is the variance in position for women. This is the only case in predicting physical attraction in which the result is significantly better than the baseline. It seems that women move less when they are physically attracted to the other (45 pixels vs 76 pixels in variance, where the body width was approximately 80 pixels). It is interesting to see that for males the variance in position of their female counterpart is also a good predictor. What is surprising however is that in the cases where males are physically attracted the variance in position of the female is higher (67 pixels vs 47 pixels). So females who are physically attracted tend to move less, while males tend to be physically attracted to females who move more!

6.3. Comparison between both

There are some similarities between the results on predicting exchanging contacting information and predicting physical attraction. In both cases the best accuracy is around 70% and in both cases features from the movement category give good results. However there are quite some differences. While synchrony related features don't seem to work when predicting exchanging contact information, they do perform well when predicting physical attraction. Another apparent difference is that predicting attraction is a lot easier when looking at females. An explanation for these differences can be found when looking at what we are exactly trying to predict. If you are physically attracted to someone you are probably more inclined to want to exchange contact information, however in wanting to exchange contact information there are of course more things that come in to play than physical attraction. When looking at the correlation between the physical attraction score and wanting to exchange contact information we see that this is indeed the case. The Pearson correlation coefficient between those variables is 0.30 for females and 0.46 for males. This also supports the view that the decision of whether or not they want to exchange contact information is more complex for females. This also is a possible explanation of why the performance on females is better on predicting physical attraction. While the decision on exchanging contact information might be complex for females and therefore difficult to predict, whether or not they are physically attracted might be just as simple (or even simpler) than with men. Wanting to exchange contact information could be an indication of romantic attraction, instead of physical.

	Male		Female	
	SVM	kNN	SVM	kNN
All	0.48	0.55	**0.59**	0.47
Positional	0.50	0.38	**0.66**	**0.58**
AVG-DIFANGLE	0.50	0.38	**0.66**	**0.58**
Movement	0.55	0.50	0.55	**0.67**
VARDIS	0.59	0.50	0.39	0.53
VARDIFANGLE	0.33	**0.61**	**0.59**	**0.58**
VARPOS	0.59	0.38	**0.70**	**0.61**
VARPOS-OTHER	**0.70**	**0.61**	**0.61**	0.55
DECRDIS	0.55	0.39	0.42	**0.56**
MOVDISTR	0.59	0.44	**0.63**	0.50
MOVDISTR-OTHER	0.55	0.55	0.44	**0.56**
Distance	0.42	0.59	0.30	0.45
AVGDIS	0.42	0.59	0.30	0.45
Synchrony	**0.70**	**0.69**	0.31	0.47
MOTIONSYNC	0.53	**0.66**	**0.63**	**0.63**
MOTION-REACTION	0.42	0.58	0.30	0.45
MOTION-REACTION-OTHER	0.42	0.59	0.30	0.45
Baseline	0.59		0.55	

Table 3. Results on predicting physical attraction for males and females using a support vector machine and k-nearest-neighbors. The baseline reported was calculated by labelling all test data points as finding the other physically attractive. Boldfaced results outperform the baseline.

Due to a relatively small data set many results presented here are not statistically significant at conventional levels.

The findings are however supported by previous work and intuitively make sense. Future work should validate our findings on larger data sets.

7. Conclusion and Future Work

We have presented automatically extracted features which can be used to predict if people want to exchange contact information in a speed-date and which can be used to predict how physically attractive the other person is perceived. These features are calculated based on positional information gathered with an automatic tracking method. The use of top-down cameras allowed us to use a simple background subtraction method to do the tracking.

We have shown that the video channel can be a source of valuable information in speed-dates. Numerous features yield a performance which is better than the baseline. The performance on predicting exchanging contact information is comparable with audio-only systems, despite the fact that we use unintrusive sensor data. For males the variance in distance is a good predictor of whether or not they want to exchange contact information, while for females information on the movement of their male counterpart gives good results. When trying to predict physical attraction we see that there are quite some differences in the results. Other features work well and for females physical attraction seems easier to predict than exchanging contact information, which is in concordance with biological findings. The results also indicate that there are important differences in the classification task between men and women and that addressing this as two different tasks increases performance.

An important question that needs to be answered is how the results presented here generalize to other groups. It would be really informative to look at a bigger dataset of speed-dates. Future work should investigate if a combination of the video features presented here with audio features would yield even better results. This could be combined with making a system which gives direct user feedback.

Other interesting video features that could be addressed include body and head orientation and more fine-grained motion based features, such as the motion of individual limbs.

8. Acknowledgements

The authors would like to thank their colleagues from the MatchMakeRS team for helping organize and set up the experimental part of this study. We would also like to thank cafe Polder for letting us use their facilities. This research was supported in part by a Marie Curie Research Training Network fellowship in the project "AnaSID" (PIEF-GA-2009-255609)

References

[1] O. Aran, H. Hung, and D. Gatica-Perez. A multimodal corpus for studying dominance in small group conversations. In *Proc. LREC workshop on Multimodal Corpora, Malta*, 2010. 2

[2] M. Argyle. *Bodily communication*. Taylor & Francis, 1988. 1, 2

[3] K. Bousmalis, L. Morency, and M. Pantic. Modeling Hidden Dynamics of Multimodal Cues for Spontaneous Agreement and Disagreement Recognition. *Mind*, 1005:9, 2011. 2

[4] D. Buss and D. Schmitt. Sexual strategies theory: An evolutionary perspective on human mating. *Psychological review*, 100(2):204, 1993. 6

[5] T. Chartrand and J. Bargh. The chameleon effect: The perception-behavior link and social interaction. *Journal of personality and social psychology*, 76:893–910, 1999. 5

[6] J. Curhan and A. Pentland. Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first five minutes. *Journal of Applied Psychology*, 92:802–811, 2007. 2

[7] D. Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: a review. *Image and Vision Computing*, 27(12):1775–1787, 2009. 1

[8] K. Grammer, M. Honda, A. Juette, and A. Schmitt. Fuzziness of Nonverbal Courtship Communication Unblurred by Motion Energy Detection* 1. *Journal of personality and social psychology*, 77(3):487–508, 1999. 2, 6

[9] E. Hall. The hidden dimension . 1966. 2

[10] D. Henningsen. Flirting with meaning: An examination of miscommunication in flirting interactions. *Sex Roles*, 50(7):481–489, 2004. 1

[11] R. Heslin and M. Patterson. *Nonverbal behavior and social psychology*. Plenum Publishing Corporation, 1982. 2

[12] H. Hung, Y. Huang, G. Friedland, and D. Gatica-Perez. Estimating Dominance in Multi-Party Meetings Using Speaker Diarization. *Audio, Speech, and Language Processing, IEEE Transactions on*, (99):1, 2010. 2

[13] W. Ickinger and S. Morris. Psychological characteristics and interpersonal distance, 2001. 3

[14] A. Madan, R. Caneel, and A. Pentland. Voices of attraction. 2004. 1, 2, 5, 7

[15] J. McCroskey and T. McCain. The measurement of interpersonal attraction. *Communication Monographs*, 41(3):261–266, 1974. 5

[16] A. Pentland. Socially aware, computation and communication. *Computer*, 38(3):33–40, 2005. 1

[17] 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-Volume 1*, pages 334–342. Association for Computational Linguistics, 2009. 1, 2, 6

[18] D. Reidsma, A. Nijholt, W. Tschacher, and F. Ramseyer. Measuring Multimodal Synchrony for Human-Computer Interaction. In *Cyberworlds (CW), 2010 International Conference on*, pages 67–71. IEEE, 2010. 5

[19] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12):1743–1759, 2009. 1

[20] G. Zen, B. Lepri, E. Ricci, and O. Lanz. Space speaks: towards socially and personality aware visual surveillance. In *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis*, pages 37–42. ACM, 2010. 2