Using Electrodermal Activity to Recognize Ease of Engagement in Children during Social Interactions

Javier Hernandez[1] **Ivan Riobo**[2] **Agata Rozga**[2] **Gregory D. Abowd**[2] **Rosalind W. Picard**[1]

[1]Media Lab
Massachusetts Institute of Technology
Cambridge, MA, USA
{javierhr, picard}@media.mit.edu

[2]School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA, USA
{ivan.riobo, agata, abowd}@gatech.edu

ABSTRACT

The recent emergence of comfortable wearable sensors has focused almost entirely on monitoring physical activity, ignoring opportunities to monitor more subtle phenomena, such as the quality of social interactions. We argue that it is compelling to address whether physiological sensors can shed light on quality of social interactive behavior. This work leverages the use of a wearable electrodermal activity (EDA) sensor to recognize ease of engagement of children during a social interaction with an adult. In particular, we monitored 51 child-adult dyads in a semi-structured play interaction and used Support Vector Machines to automatically identify children who had been rated by the adult as more or less difficult to engage. We report on the classification value of several features extracted from the child's EDA responses, as well as several other features capturing the physiological synchrony between the child and the adult.

Author Keywords
Electrodermal Activity; social engagement; physiology; feature analysis; Support Vector Machines.

ACM Classification Keywords
H.5.3. Group and Organization Interfaces: Evaluation/methodology, Synchronous interaction; J.4 Social and Behavioral Sciences: Sociology.

General Terms
Experimentation, Measurement, Performance.

INTRODUCTION

A basic feature of successful social interactions is synchrony, or the tendency of social partners to modulate and coordinate their behaviors and affective states [10, 22].

UbiComp '14, September 13 - 17, 2014, Seattle, WA, USA
Copyright 2014 ACM 978-1-4503-2968-2/14/09…$15.00.
http://dx.doi.org/10.1145/2632048.2636065

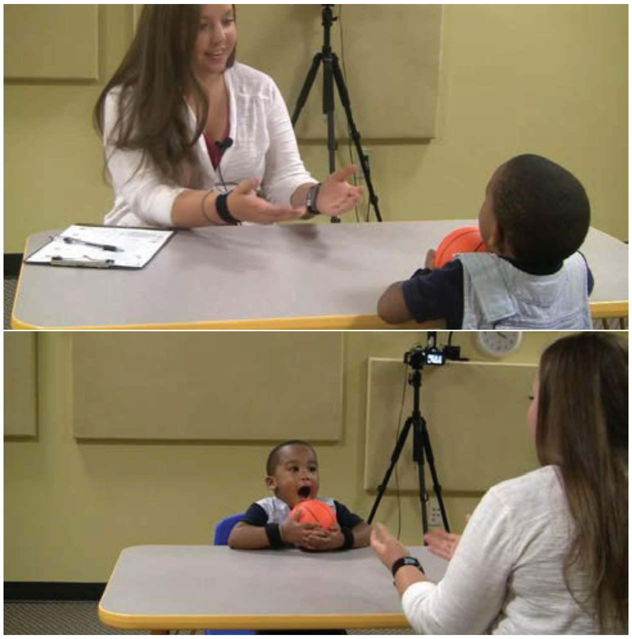

Figure 1. Interaction between child and adult during a ball play game.

Even though there is strong evidence that synchrony occurs at the physiological level as well [21, 22, 24], most approaches to assessing interactive synchrony focus on how two social partners coordinate externally observable behaviors – patterns of attention, affective expression, gesture, and speech. Until recently, explorations in recognizing the quality of social interactions for other purposes, such as augmenting behavior annotation, have been hampered by limitations of the sensing hardware: a reliance on wired and bulky sensors that precluded assessment of physiological states in the context of actual social interactions. For young children, using invisible physiological signals, such as electrodermal activity, to assess interactive synchrony is particularly appealing since their capacities for social coordination are just emerging.

With recent advances in wearable biosensing technologies, it is now feasible to develop systems that automatically monitor not only outwardly observable behaviors, but also inward physiological states of children that may serve as key markers of social engagement. Such technologies may not only help characterize qualitative aspects of children's social engagement, but may potentially also assist with the identification and quantification of developmental delays [22]. This paper explores whether we can successfully leverage modern biosensors to identify children who have been rated by an adult interactive partner as more or less difficult to engage. In particular, we monitored the electrodermal activity of 51 child-adult dyads during a short naturalistic social interaction (see Figure 1 for an example of interaction), and used Support Vector Machines to automatically differentiate children who were rated as easier versus harder to engage.

The paper is organized as follows. We begin by summarizing related research on measuring engagement and the use of physiological information in the context of social engagement. We then outline our procedure for collecting data and rating children's engagement. After explaining how the physiological signals are preprocessed, we describe a variety of features to characterize the child and adult's physiological responses. We then report our main findings in terms of classification performance of combinations of different types of features. We conclude by providing some discussion and highlighting future steps to push forward this line of research.

BACKGROUND AND PREVIOUS WORK

Measuring Engagement

Analyzing the engagement level of people has been the focus of interest in a wide variety of situations. Although the definition of engagement is very context-specific, there are three well differentiated approaches to measuring it: self-reports, external ratings, and physiological information.

Self-reports can take many forms such as interviews or surveys taken during or after a situation of study [27]. This method is arguably one of the fastest and most direct approaches to gather information but is subjective and can suffer from information recall bias. Moreover, self-reports are disruptive and are not appropriate for certain populations, such as young children who may not be able to reflect on and articulate their affective state. An alternative method involves having experienced coders review videos of recorded interactions to rate the perceived interactive experience, or to mark onsets and offsets of individual interactive behaviors or engagement states [1]. This method is very common in the field of facial expression analysis where Facial Action Unit coders [8] annotate the appearance of specific facial movements associated with basic emotions. Although this approach is useful for the development of automatic expression recognition systems, it can be relatively time intensive and laborious to train coders to reliability [5].

A less disruptive approach involves measuring physiological signals. A wide variety of signals have been used in different settings. In the context of market research, for example, researchers have shown that signals such as gaze behavior [31], heart rate [20] and EDA [19] can be used in laboratory settings as effective indicators of the interest levels of people to certain stimuli. For instance, Hernandez et al. [14] and, more recently, Silveira et al. [29] have shown that physiological metrics such as facial expressions and EDA, respectively, can be used to recognize the engagement level of TV viewers. In other social environments such as conference meetings, head pose orientation has been widely used to identify the visual focus of attention of participants [30, 32], with the underlying assumption that people pay attention to whatever they are looking at. Although facial and head gestures can be measured at a distance, they are easier to voluntarily control and may not always be congruent with the internal affective experience.

Social Engagement and Synchrony

A key property of engagement during social interactions is interactional synchrony [10, 22], which is associated with the coordination of behaviors between individuals during social interactions. Traditionally, a high degree of synchrony indicates a high level of engagement, consisting of closely coordinated behaviors and contingent social responses [18].

The synchrony of physiology during social interactions (also known as physiological linkage) has been studied in a broad set of applications. For instance, Levenson and Gottman [21] monitored several physiological signals such as heart rate and EDA in 30 married couples to study marital satisfaction. They found that greater synchronization was associated with more distressed interactions. In a different study, Marci et al. [24] analyzed EDA to study the empathy between 20 patient-therapist dyads. In this case, greater synchronization was associated with higher patients' ratings of perceived therapist empathy. Physiological synchrony has also been used as a measure of the intensity of gaming and social interactions [9, 12], irrespective of the emotional valence of the interaction.

Methodological limitations such as wired and cumbersome sensors traditionally have made impractical the study of inward responses of children in the context of naturalistic interactions. However, the availability of modern wearable physiological sensors provides an opportunity to begin to study internal physiological markers of social engagement in the course of naturally-occurring scenarios. One relevant example is the work of Hedman et al [13], which monitored and visualized the EDA of 22 children with sensory challenges while they used zip lines, jumped in ball pits, and otherwise engaged in occupational therapy services. More recently, Chaspari et al [7] explored the utility of EDA of three children with Autism Spectrum Disorder and their therapists to better quantify the quality of interventions. In comparison with previous studies, the work presented in this paper considers 51 children-adult interactions and is the first to explore physiological synchrony in automating recognition of social engagement.

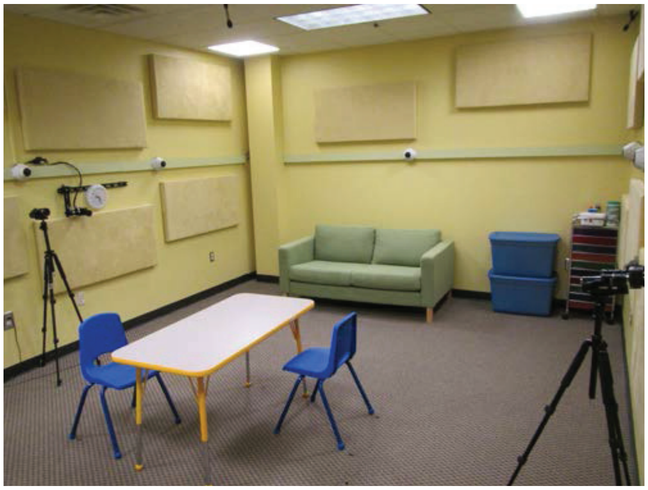

Figure 2. Experimental setting.

DATA COLLECTION

The data utilized in this analysis came from a larger dataset collected by a multi-disciplinary team with the ultimate goal of building a broad set of computational tools for measuring and analyzing child social-communication behavior [28]. This section provides details on the experimental setting, the social interaction, and the experimental procedure relevant to our study.

Procedure

Children between the ages of 15 and 30 months were recruited through advertisements distributed to daycare centers, community and parent mailing lists, and through targeted mailings to families of young children identified through a commercially available newborn mailing list. The age range was chosen to correspond to a period in development when key social-communicative behaviors are emerging and becoming consolidated. Both children and parents were invited for a 30-45 minute play session. During the first 10-15 minutes, there was a warm-up period when the child and adult examiner played together on the floor so that the child could get acclimated to the new environment and to the wearable sensors. Then, the child and the adult moved to a small table where the adult engaged the child in a 2-5 minute semi-structured play interaction (described below). The child was seated in the parent's lap or, in cases where a child preferred to sit alone, the parent remained in the room seated at a nearby sofa. Each family received a $50 gift card for their participation, and the child was given a small toy at the end of the session. The experimental protocol was reviewed and approved by the university's Institutional Review Board.

Apparatus

To capture the interaction, we equipped a laboratory room with several cameras, microphones, and physiological sensors (see Figure 2). Both child and adult wore Affectiva Q^{TM} (htttp://www.qsensortech.com) biosensors on both of their wrists. However, we only utilized the information captured by the left wrists, which corresponded to the non-dominant hand of the adults and is in accordance with standard practice [4].

The Q^{TM} sensors use dry Ag-AgCl 1cm diameter electrodes and record EDA, skin surface temperature, and actigraphy through 3-axis accelerometry (sampling rate of 32Hz). To improve signal acquisition, K-Y gel was placed on each electrode and the biosensors were attached during the warm-up period so that a baseline level was reached before the beginning of the semi-structured social interaction. To synchronize the streams of different biosensors, the adult switched on all of the devices at the same time and then simultaneously pressed the event-mark buttons on the sensors three consecutive times. Finally, the sensors were horizontally moved for a few seconds in front of one of the cameras to synch visual and accelerometer data.

Social Interaction

The interaction between the child and the adult consisted of five scripted activities during which the adult actively attempted to engage the child. These included saying hello to the child, rolling a ball back and forth, looking through pictures in a book together, putting the book on one's head as a hat, and gentle tickling (see Figures 1 and 3 for representative moments). These interactions were carefully designed to elicit behaviors that are developmentally relevant to the social and communicative growth of a young

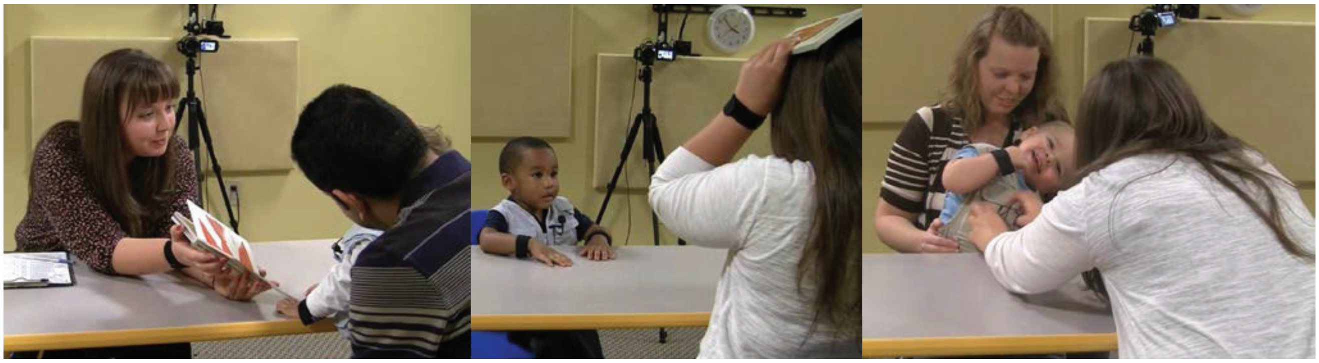

Figure 3. (From left to right) Representative moments of looking pictures in a book, book on as a hat, and tickling.

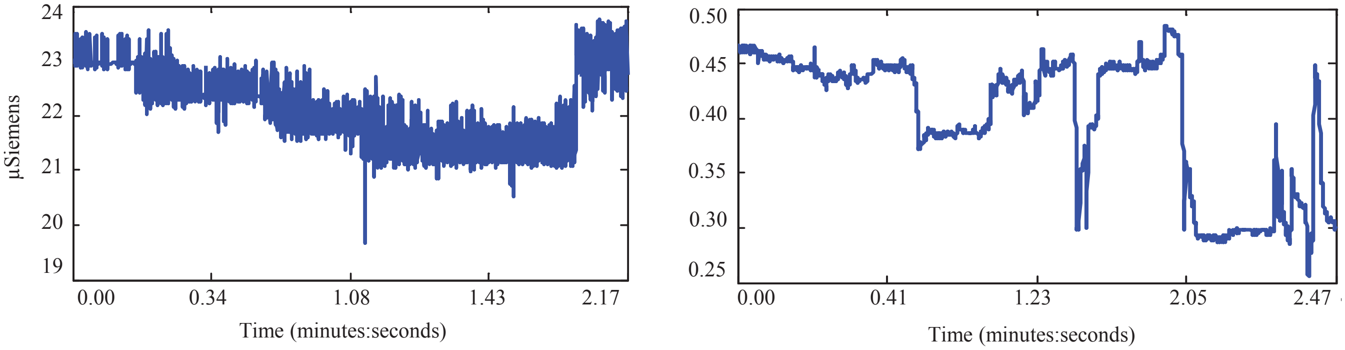

Figure 4. Examples of noisy EDA responses: quantization error (left), and movement of the sensor (right).

child [25]. The ordering of the different activities remained constant for all sessions.

Engagement Scoring

Adults were research assistants with extensive experience in interacting with young children. Each adult was trained to consistently and naturally guide children through the five activities, and score child's engagement for each of the five activities. As part of the training, adults were required to obtain over 90% agreement for 3 consecutive sets of 10 sessions previously rated by a clinical consultant.

Although the definition of engagement varies for different settings and studies, our play protocol defined engagement as the amount of effort required to engage the child. The scoring guidelines used by the adults were as follows:

- Score 0: The interaction with the child required little effort for the adult and/or the child was ready and eager to engage.
- Score 1: The interaction with the child required some effort on the part of the adult due to the child's shyness or distractibility.
- Score 2: The interaction with the child required extensive effort and/or the child was highly fussy or refused to interact.

Data Overview

In this work, we collected and synchronized information from 74 sessions. However, 7 of the sessions contained abnormally high electrodermal responses (>20 µSiemens) and the Q^{TM} sensor was unable to record the data without quantization problems (e.g., see left in Figure 4). Furthermore, 16 other sessions were discarded due to the presence of large amounts of artifacts in at least one of the sensors. These sessions could be easily characterized by long periods of flat responses (i.e., 0 µSiemens) and/or abrupt signal drops that were incongruent with the slow exponential decays of typical EDA responses [4]. Abrupt signal drops such as those observed on the right in Figure 4 are mostly due to movement of the sensor. The distribution of engagement ratings of the excluded sessions was similar to the one observed when considering all the sessions.

After excluding the sessions with quantization and artifact problems, the final subset of data contained 51 sessions (27 females), guided by 4 different adults (all female and right handed). For these 51 sessions, the average age of the children was 21 months (SD = 5.23), and the average duration of the social interaction was 2.72 minutes (SD = 1.02 minutes). Furthermore, 79.6% of the individual stage engagement rating scores were 0, indicating that most children were easy to engage.

CHARACTERIZATION OF EDA RESPONSES

Electrodermal Activity and Arousal

Electrodermal activity, often referred to in earlier work as galvanic skin response, has been one of the most widely used signals in psychophysiological research during the last century [4]. EDA has been commonly measured as skin conductance off the finger or palmar surface, which provides an indication of the activation of eccrine sweat glands. Since this type of sweat gland is purely innervated by the sympathetic nervous system, skin conductance has been considered as one of the best indicators of sympathetic arousal [4]. Increased levels of arousal typically result in sensory alertness, increased readiness to respond, and mobility. Furthermore, arousal regulates attention and emotion, which are critical for successful social interactions and daily functioning.

Preprocessing

Prior to analyzing physiological responses, the data typically undergoes several preprocessing steps. Quick sensor movements may introduce signal artifacts in the form of high frequency changes that need to be considered. This problem is very common in uncontrolled settings such as social interactions as these typically involve gestures and body movements. In order to attenuate these artifacts, we use a Hanning filter with a 1 second window [1].

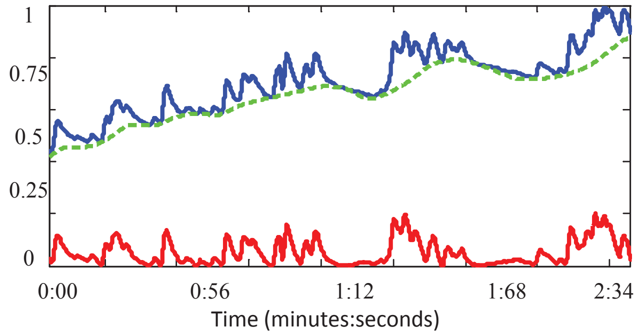

Figure 5. Normalized skin conductance (top-blue) and its tonic (green-dashed) and phasic (red-bottom) components for one of the child's sessions.

A fundamental challenge when developing algorithms that rely on EDA data is the large individual differences (e.g., range of values) that appear due to multiple factors (e.g., different amount of sweat glands, variable skin thickness). This problem is especially relevant when building supervised learning tools that infer certain information (e.g., engagement levels) from a group of people (i.e., training set) with the hope that it would generalize to other unseen people (i.e., testing set). In this work, we normalize the range of values of each session to be between zero and one in accordance with standard practice (e.g., [15], [23]). This normalization not only amplifies the physiological changes associated with the session but also facilitates comparison across people (both within training set and between training and testing sets).

Finally, while common EDA measurements are a one dimensional time series signal, there are typically two distinct components: a phasic component which shows quick changes associated with stimulus-specific or non-specific responses, and a tonic level which changes more slowly and can be observed in the absence of any particular discrete environmental event or external stimuli. While some studies using EDA in the context of classification use the original 1D signal (e.g., [15]), this work explores the discriminative power of each of the components separately. This approach has the additional benefit of being able to better capture complementary aspects of the physiological responses. In order to extract the two components from the original signal, we utilized the deconvolution approach proposed by Benedek and Kaernbach [1]. Figure 5 shows an example of the decomposition.

Features

Before the physiological responses can be used for classification, it is necessary to extract representative features. In this work, we explored the utility of multiple features, which can be grouped into two categories: individual features (IF), and synchrony features (SF).

From the tonic and phasic components of each child's EDA signal, we extracted the following IF features: mean, standard deviation, area under the curve, relative positions of maximum and minimum values, slope (estimated by linear interpolation), average number of peaks, and average of the peaks amplitudes. Peaks were detected using the *findpeaks* MATLAB function and were required to have an amplitude of at least 0.01 after normalization and a minimum distance between peaks of at least 1 second. While some of the features aim to capture the temporal aspects of the responses (e.g., slope captures an overall increase or decrease of the response), other features aim to capture overall activation throughout the period (e.g., average number of peaks can be seen as an indicator of arousal).

Motivated by previous research in physiological synchrony (e.g., [21], [24]), we also explored several SF features to capture the relationship between the EDA responses of the child and the adult. One of the most effective and commonly used methods is the Pearson product-moment correlation (PC), which measures the linear dependence between two variables. Although this method works well in practice, we also evaluated the following two methods:

- Canonical Correlation (CC). This method similarly measures linear dependence between two variables but also tries to represent the information in a different dimensional space where the correlation is maximized. An important property of this method is that the result is invariant with respect to affine transformations of the variables. Therefore, we hypothesized that this method could address some of the individual differences of the signals that could not be corrected by the preprocessing steps. To the best of our knowledge, this approach has not been previously explored in the context of social engagement from EDA.
- Dynamic Time Warping (DTW). This method utilizes a dynamic programming approach to find the similarity between two signals. The main advantage of this method is that it allows some temporal flexibility in terms of signal durations and delayed responses. We hypothesized that this method would help align and compare asynchronous responses between child and adult, such as those that could be observed during turn-taking interactions (e.g., ball play interaction).

As part of the SF features, we also computed the difference between some individual features extracted from both the adult and the child's responses. In particular, we extracted mean, number of peaks, and average amplitude of these peaks. Then, we utilized the L2-norm as a distance metric to capture the difference between the pairs of features. Figure 6 shows an overview of the features we explored, grouped into the two categories. Note that the dimensionality of each feature is one.

EXPERIMENTAL SETTING

The goal of the study is to explore the feasibility of accurately and automatically identifying children rated as more or less difficult to engage by relying solely on the physiological responses of a 5-stage social interaction. While the original goal was to provide fine grained

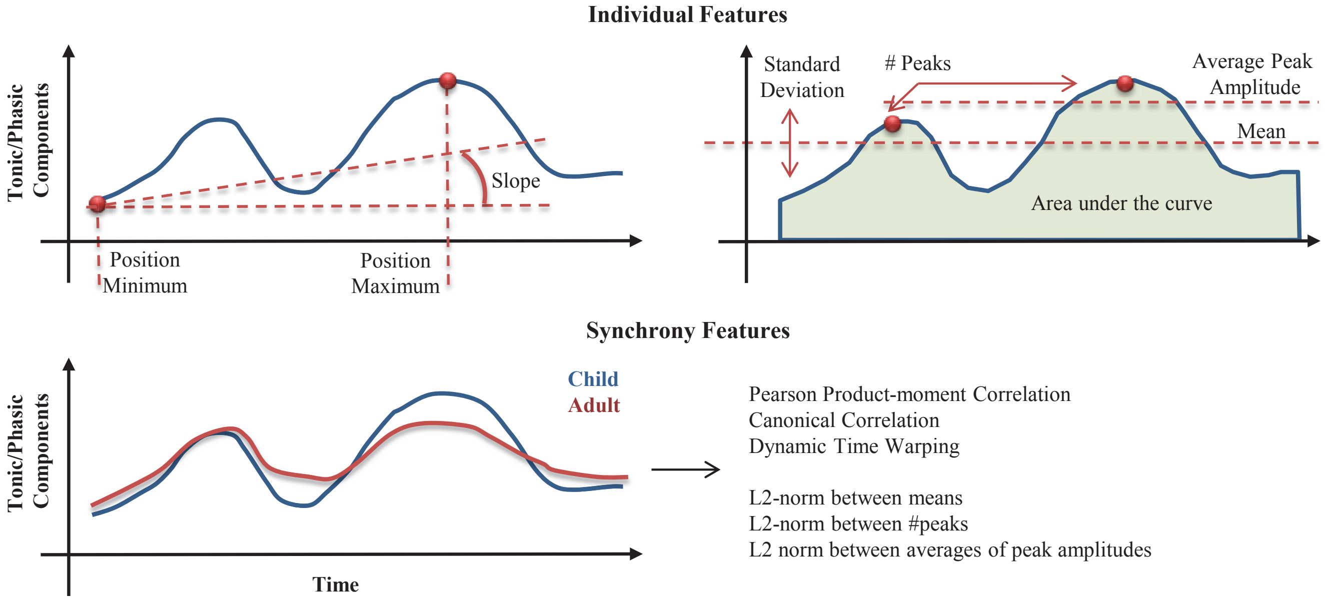

Figure 6. Summary of features extracted from the physiological responses: (top) individual features extracted from the child, and (bottom) synchrony features extracted from both child and adult's responses.

engagement scores for each of the sub-stages of the social interaction, we encountered two major complications with the dataset. First, some of the stages are too short to make meaningful physiological assessments and explore some of the synchrony-based features, especially since EDA responses may appear 1 to 5 seconds after a specific stimulus [4]. For instance, the stages of saying "hello" and putting the book on one's head as a hat lasted 6 and 9 seconds on average, respectively. Second, the distributions of engagement ratings within stages are very unbalanced, making it very challenging for the classifier to appropriately model all the classes and not ignore the least common one. For instance, 84% of the engagement ratings for rolling the ball and gentle tickle were zero (i.e., easy to engage). In order to attenuate these problems and start exploring the feasibility of automatically characterizing social engagement, we divided the sessions into the following two groups: *easier to engage* (n = 29), consisting of the children who scored zero for all five stages, and *harder to engage* (n = 22), consisting of the children who scored 1 or 2 for at least in one of the stages. This division enabled us to not only examine a longer observation window (2.7 minutes on average) but also to address the problem as a binary classification problem with relatively balanced distribution of classes.

Classification

In order to perform classification, we used the LIBSVM library [6], which provides an efficient implementation of Support Vector Machines (SVMs) [3]. We used a 10-fold-cross-validation protocol for testing and training of the algorithm. Therefore, we divided the sessions into 10 different groups and used 9 of them as a training set and the remaining one as the testing set. This process was iteratively repeated until engagement labels were automatically generated for all the groups. During the training phase, the training set was divided into 10 different groups and followed the same iterative process to gather performance for different misclassification costs ($\log_2$ C, for C = {0, 1, 2 ... 18}) of a Linear SVMs with probabilistic estimates. Once the process was completed, we used the whole training set and the best cost to obtain the final classifier model, which then was used in the testing set. The misclassification weights for each class were set to be the class priors of the other class to force the algorithm to make more balanced predictions of the labels [16]. When considering the whole dataset, for example, the misclassification weight of the smallest class (i.e., harder to engage) class was 29/51. Note that data from the same child was never used for training and testing at the same time.

For some of the experiments, we also incorporated the Sequential Forward Selection (SFS) approach [11] in the training phase, allowing us to identify some of the most discriminative combinations of features. Starting from the best single feature, this method iteratively incorporates features that improve performance. The algorithm stops iterating if there are not more remaining features or the best performance is achieved with a smaller subset.

Performance

Two of the most common methods to evaluate performance of a classifier are the area under the Receiver Operating Characteristic (ROC) curve, and the area under the Precision/Recall (PR). In this work we use the average between the areas under the two curves obtained when considering different thresholds on the probability estimates provided by SVMs (similar to [14]). This metric was used as a reference to find the optimal misclassification cost of SVMs during the training phases and will be used in the following section to report classification performance. In particular, this metric ranges from 0 (worst performance) to 100 (maximum performance) and its main advantage is that

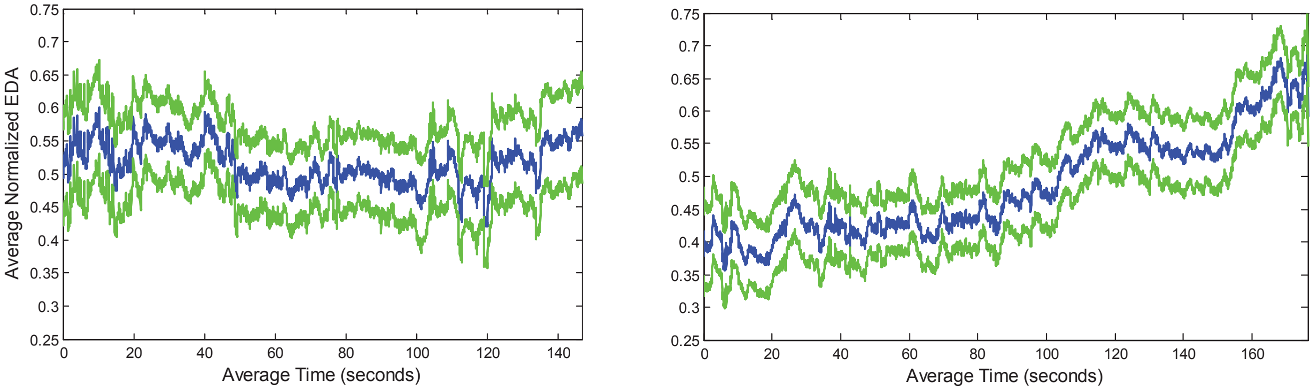

Figure 7. Average EDA (blue) and standard error (green) of children in the *easier* (left) and *harder* (right) to engage groups.

it provides a classification performance that is more invariant to unbalanced classes than traditional metrics (e.g. Accuracy, F1 score). Furthermore, this metric captures the discriminative power of the classifier for several configurations (i.e., different thresholds on the probability estimates). Note that a classifier that always predicts the most likely class will obtain a performance of 0 as none of the curves can be computed (i.e., probability estimates are always one).

RESULTS

In this section we analyze the physiological responses of 51 children during their interaction with an adult. The first two parts of the analysis provide graphical and quantitative intuition of the most relevant EDA characteristics, respectively. The final part of the analysis provides quantitative evaluation of the different types of features and recognition performance in the context of social engagement recognition.

Physiological Responses

In order to provide a preliminary graphical intuition of the physiological responses, Figure 7 shows the average normalized EDA response and standard error of children in the *easier* (left) and *harder* (right) to engage groups. Since the completion time was different for each session, we resampled each response to last the average session time of each group, and then computed the average across sessions within the same group. While the average session time of children in the first group was 2.44 minutes (STD = 0.4), the average session time of the second group was 2.93 minutes (STD = 1.25). This difference is to be expected as by definition in the latter group the adults had to spend more time trying to maintain the child's engagement. As can be observed in the graph, the average physiological responses of each group show distinctive trends. While children who were easier to engage displayed a more constant response throughout the session (around the average), children who were more difficult to engage displayed a response that continuously increased over time. Note, however, that the distribution of engagement ratings throughout the interaction stages varied for each child. Figure 8 shows the average engagement score throughout the whole interaction for children in the *harder* to engage group. The engagement scores for children in the *easier* to engage group remained constant throughout the whole interaction (zero by definition). As can be seen, the ratings fluctuate for the different parts of the interactions, indicating that children were more difficult to engage right at the beginning (corresponding to the adult saying hello to the child) and easier to engage by the end of session (corresponding to the gentle tickling). Furthermore, the lowest engagement scores were achieved between the 60 and 120 seconds of the session, corresponding to the part where the adult and the child looked through pictures in a book together. Interestingly, both positive and negative fluctuations of the engagement scores seem to be associated with EDA increases. In the following sections, we explore different EDA characteristics in terms of classification performance.

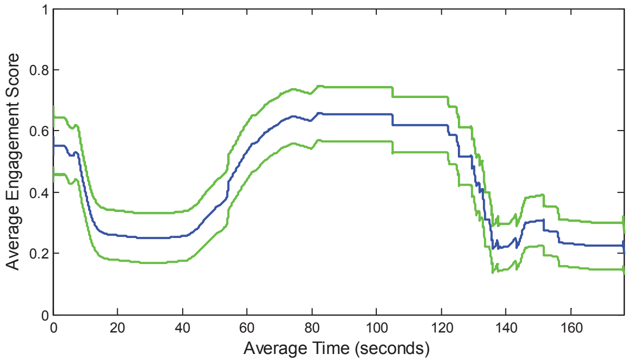

Figure 8. Average engagement score (blue) and standard error (green) for children in the *harder* to engage group. Low scores indicate easier interactions and high scores indicate more difficult interactions.

EDA Characteristics

In order to further characterize the physiological responses, this section quantifies the utility of different features in terms of classification performance. In particular, we consider each of the individual and synchrony features and extract them from both the tonic and the phasic components.

Component	Mean	STD	Area	Position min.	Position max.	Slope	#Peaks	Avg. Peak Amplitude
Tonic	66.23	44.85	66.23	53.91	**69.61**	41.59	45.19	58.12
Phasic	55.33	**64.67**	50.33	49.50	57.75	53.83	42.10	47.68
Mixed	60.92	47.34	60.92	49.92	47.18	40.19	46.59	**62.30**

Table 1. Classification scores (%) of individual features. (STD: Standard deviation)

Component	PC	CC	DTW	Mean	#Peaks	Avg. Peak Amplitude
Tonic	**59.38**	54.13	53.12	50.32	45.73	50.45
Phasic	45.06	55.30	50.52	46.35	**69.29**	47.09
Mixed	**63.18**	50.36	45.64	48.54	43.21	51.79

Table 2. Classification scores (%) of synchrony features. (PC: Pearson product-moment correlation; CC: Canonical correlation; DTW: Dynamic Time Warping)

Table 1 shows the classification performance obtained for each of the IF features extracted from the child's physiological responses. As can be seen, the relative position of the maximum tonic value yielded the highest classification performance (69.91%), followed by the tonic mean and area under the curve (66.23%). For the phasic component, the standard deviation yielded the highest performance (64.67%). Similarly, Table 2 shows the classification performance for each of the SF features. While PC yielded the best performance among the tonic features (59.38%), the difference between the number of peaks of the dyad yielded the best performance above all the features, which is 10% higher than the best tonic feature. The best result achieved by IF features is comparable to the best result achieved by SF features, demonstrating that both types are relevant in the context of engagement recognition. However, while the tonic component may be more relevant when only analyzing the responses of children (in accordance with Figure 7), the phasic component provides more discriminative information when capturing the synchrony of the dyad.

In order to assess if the decomposition of EDA into the two components provided meaningful information, we also included a mixed component, which corresponds to the original one dimensional EDA response. As can be seen on the tables (bottom rows), the best performance achieved with this approach is below the best performance achieved by the best feature of the other two components. Furthermore, extracting the two components can help better interpret the results. For instance, higher average peak amplitude in the mixed component may correspond to a case where there is a high tonic level and small phasic peaks, or another case where there is low tonic level and large phasic peaks. By looking at features from the two components separately, such as mean and average of peak amplitudes of the tonic and phasic components, we can infer that the tonic component provides more discriminative information in this case. These results suggest the decomposition of EDA responses is recommended for this type of analysis.

Engagement Recognition

While the previous section focused on the discriminative power of each feature independently, this section explores combining several features to determine the best possible performance. Furthermore, in order to assess the benefit of monitoring both interactive partners, we applied Sequential Forward Selection to IF and SF features separately, and then combined the two types.

When considering the IF features extracted from the child responses, incorporating SFS during the training phase yielded a performance of 74.35%. While different subsets of features were selected at each training fold, the top 3 most selected features were the position of maximum value (6 times) and the average peak amplitude (5 times) from the tonic component, and the standard deviation from the phasic component (4 times). When considering the SF features, incorporating SFS yielded a performance of 76.16%, which is slightly higher than using only individual features. In this case, the top 3 features were: DTW (8 times) and the differences between the number of peaks from the phasic components (6 times), and DTW from the tonic components (7 times). Interestingly, DTW were not among the best features when considered separately (53.12% and 50.52% from tonic and phasic components, respectively), but still provided relevant complementary information when combined with other features.

Finally, we examined the recognition value of combining both IF and SF features. In this case, incorporating SFS yielded a classification performance of 81.03%,

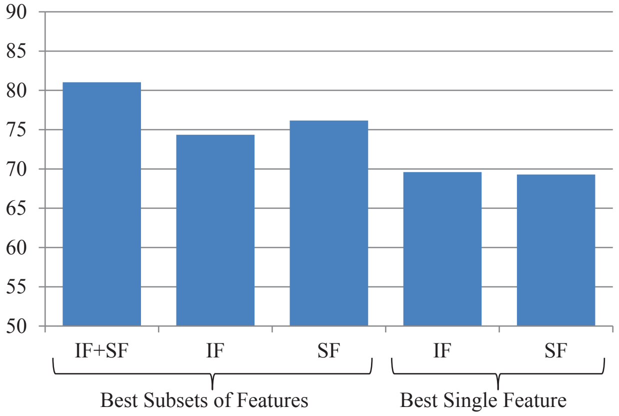

Figure 9. Top classification performance (%) achieved by subsets (left) and single (right) feature types. (IF: individual features; SF: synchrony features).

outperforming any of the previous experiments. In this case the top 3 most selected features were the difference between the peaks of the dyad (7 times) from the phasic component, and the STD and DTW from the tonic components (5 times each).

Figure 9 provides an overall overview of the best results achieved by each feature type with and without incorporating SFS during the training phase. By combining different types of features, we were able to improve the recognition performance by 11.42%. Figure 10 shows in more detail the traditional ROC and Precision/Recall curves of the best subsets of features. Note that the results reported above correspond to the average between the two curves.

DISCUSSION

Physiological signals have been extensively measured and analyzed in short controlled interactions that usually have not included spontaneous social interaction. This work is novel in examining continuous EDA activation during a spontaneous and structured social interaction, and using this information to build automated tools for recognizing how easy or hard a person is to engage.

Among some of the main findings, we found that the relative position of the maximum tonic value was the most discriminative feature to automatically identify easy to engage children. Furthermore, we compared several methods that captured the physiological synchrony of the dyad, and found the difference between the number of peaks of phasic components to yield similar performance. This finding indicates that both individual and synchrony features are relevant to modeling engagement during social interaction. However each component captures different aspects. While the tonic component is the most relevant information when only monitoring the child, the phasic component is especially helpful when capturing synchrony. Furthermore, we have shown that decomposing EDA responses into the tonic and phasic components provides some benefits for the analysis, such as improved recognition performance (>6%) and increased interpretability of the findings. Finally, we showed that the combination of two feature types (IF+SF) yielded the highest classification performance (11% higher than using only the best single feature). In other words, using just the child's physiology to predict the ease of engagement score assigned by the adult was not as accurate as when the automatic system used both the child's physiology and its synchrony with the adult's physiology.

Despite the significant effort to maximize the utility of the EDA data – use of gel, a warm-up period to get children acclimated to the sensor and to gather extensive baseline data – data from 31% participants had to be excluded from the analysis due to the presence of large artifacts and

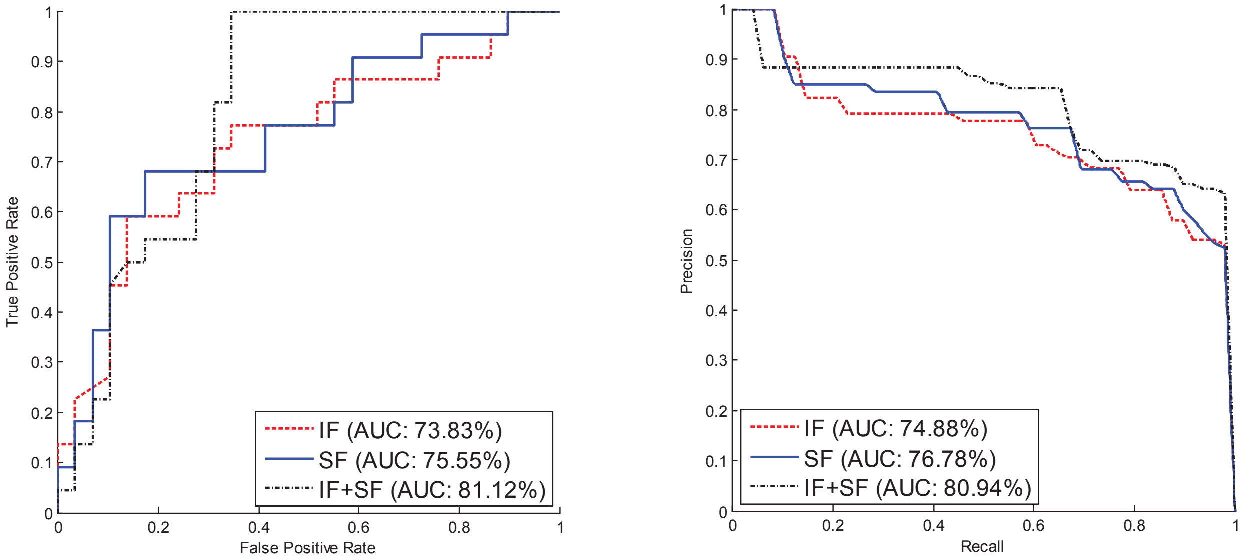

Figure 10. Receiver Operating Characteristic (left) and Precision/Recall (right) curves for the best subsets of features when considering different types of features (individual features, synchrony features, and the combination of both).

quantization problems in at least one of the sensors. Although we preprocessed the data to reduce short and high frequency motion artifacts, there were several cases where this approach was not enough. This challenge points out a limitation of our study but also highlights the great challenge of monitoring physiological information in uncontrolled environments, especially with small children. In our case, one of the main sources of artifacts was due to the loss of contact of the sensor to the skin. Although this problem can be easily fixed by tightening the Velcro band that comes with the Q^{TM} sensor, we noticed that doing so increased the child's awareness of the sensor, leading to fiddling with the sensor and thus additional motion artifacts. Finding the right trade-off between sensor tightness and awareness as well as finding the best location for the sensor (e.g., wrist vs. ankle) will probably be dependent on the age and skin of each person, and promises to be a relevant research area in the field of wearable computing.

Our final dataset consists of 51 sessions in which we monitored the time-varying physiological signals of both the child and the adult. Using this information, the present study has taken several steps towards maximizing the generalizability of the findings. First, we analyzed the recognition value of one single feature at a time, which ensures a high ratio between the number of samples and the dimensionality of the data. Second, we used an SVM in its linear form, which significantly reduces the complexity of the model. Third, we used a 10-fold-cross validation approach, which is typically used to avoid overfitting of the learned models, and ensured that we did not use data from the same people for training and testing simultaneously. Finally, we incorporated SFS as a feature selection method, which tends to find a small subset of features (i.e., reduced dimensionality) to help avoid overfitting. In the future, we plan to incorporate additional sessions and validate our method in similar conditions.

CONCLUSIONS

This work extends the possibility of recognizing the ease of engagement of children from physiological data during naturalistic social interactions that can take place almost anywhere. Leveraging modern wearable biosensors, we monitored the electrodermal responses of 51 child-adult dyads in a semi-structured social interaction. We proposed several physiological features to characterize the responses of the children and their synchrony with the responses of the adults. We found that a combination of features extracted from the child's EDA activity and features capturing the physiological synchrony between the child and the adult resulted in the highest classification accuracy in distinguishing children who had been rated as more or less difficult to engage by the adult.

Future efforts may focus on analyzing the correlation of the EDA responses with features from other modalities (e.g., head pose, voice pitch) as they can provide relevant information to further understand and better recognize children's engagement. However, these modalities require additional sensors that may not be readily accessible in daily life situations. We may also analyze each of the different interactive stages independently, as well as taking into account the influence of preceding stages and large variability of stage durations.

This study has shown that new biosensor technology can be used to capture unobtrusively, in a playful spontaneous social interaction, objective physiological time-series data that is informative about an individual child's outwardly rated engagement. As such, this work takes an important first step towards providing better measures to reliably and objectively quantify interactive social behavior, an important advancement for the study of human development.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. NSF CCF-1029585.

REFERENCES

1. Adamson, L.B., Early Interests and Joint Engagement in Typical Development, Autism, and Down Syndrome, Journal of Autism Developmental Disorders, 40 (6), (2010), 665–676.
2. Benedek, M., and Kaernbach, C. Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology, 47 (4), (2010), 647-658.
3. Boser, B.E., Guyon, I., and, Vapnik V. A training algorithm for optimal margin classifiers. In Proc. of the Fifth Annual Workshop on Computational Learning Theory, (1992), 144-152.
4. Boucsein W. Electrodermal Activity (2nd Ed). New York: Springer, (2012).
5. Calvo, R. A., and DMello, S. Affect detection: An interdisciplinary review of models, methods, and their applications. In IEEE Transactions on Affective Computing, 1(1), (2010), 18-37.
6. Chang, C. C., and Lin, C.J. LIBSVM: a library for Support Vector Machines, 2001. Software: http://www.csie.ntu.edu.tw/~cjlin/libsvm
7. Chaspari T., Goodwin M., Wilder-Smith O., Gulsrud A., Mucchetti C., Kasari C., and Narayanan S. A Non-Homogeneous Poisson Process Model of Skin Conductance Responses Integrated with Observed Regulatory Behaviors for Autism Intervention. In Proceedings of IEEE International Conference on Audio, Speech and Signal Processing (2014).
8. Ekman P., and Friesen W. Facial Action Coding System: a technique for the measurement of facial movement, Palo Alto, CA: Consulting Psychologists Press, 1978.
9. Ekman, I., Chanel, G., Kivikangas, J.M., Salminen, M., Jarvela, S., and Ravaja, N. Social interaction in games:

measuring physiological linkage and social presence. Journal Simulation and Gaming, 43(3), (2012), 321–338.

10. Feldman, R. Parent-infant synchrony and the construction of shared timing: physiological precursors, developmental outcomes, and risk conditions. Journal of Child Psychology and Psychiatry, 48 (3-4), (2007), 329-354.

11. Guyon I., and Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research, 3, (2003), 1157-1182.

12. Hatfield, E., Cacioppo, J., and Rapson, R. Emotional contagion. New York, Cambridge University Press, (1994).

13. Hedman E., Miller L., Schoen S., Nielsen D., Goodwin M., and Picard R. W. Measuring autonomic arousal during therapy. In Proc. of Design and Emotion, (2012), 11-14.

14. Hernandez, J., Liu, Z., Hulten G., DeBarr, D., Krum, K., and Zhang, Z. Measuring the engagement level of TV Viewers. In Automatic Face and Gesture Recognition, (2013), 1-7.

15. Hernandez, J., Morris, R.R., and Picard, R.W. Call Center Stress Recognition with Person-Specific Models. In Proc. of the Affective Computing and Intelligent Interaction, (2011), 125-134.

16. Huang, Y.M., and Du, S.X. Weighted support vector machine for classification with uneven training class sizes. International Conference on Machine Learning and Cybernetics, 7, (2005), 4365–4369.

17. Jo J., Lee, S.J., Jung H.G., Park, K.R., and Kim J. Vision-based method for detecting driver drowsiness and distraction in driver monitoring system. In Optical Engineering, 50 (12), (2011), 127202/1-24.

18. Kühn S. S., Müller B.C.N., Van der Leij A., Dijksterhuis A., Brass M., and Van Baaren R.B. Neural correlates of emotional synchrony. In Social Cognitive and Affective Neuroscience, 6 (3), (2011), 368-374.

19. LaBarbera, P., and Tucciarone J. GSR reconsidered: a behavior based approach to evaluating and improving the sales potency of advertising, In Journal of Advertising Research, 35 (5), (1995), 33-53.

20. Lang A. Involuntary attention and physiological arousal evoked by structural features and mild emotion in TV commercials. In Proceedings of Communication Research, 17 (3), (1990), 275-299.

21. Levenson, R.W., and Gottman, J.M. Marital interaction: physiological linkage and affective exchange. Journal of Personality and Social Psychology, 45 (3), (1983), 587–597.

22. Levenson, R.W., and Ruef, A.M. Physiological aspects of emotional knowledge and rapport. In W. Ickes (Eds.), Empathic Accuracy, 44-72. New York: Guilford Press, (1997).

23. Lykken, D.T., and Venables, P.H. Direct measurement of skin conductance: A proposal for standardization. Psychophysiology, 8 (5), (1971), 656–672.

24. Marci, C. D., J. Ham, E. Moran, and Orr. S. P. Physiologic Correlates of Perceived Therapist Empathy and Social-emotional Process during Psychotherapy. Journal of Nervous and Mental Disease, 195 (2), (2007), 103-111.

25. Mundy, P. Joint attention and social-emotional approach behavior in children with autism. Development and Psychopathology, 7 (1), (1995), 63–82.

26. Mundy P., and Burnette C. Joint attention and neurodevelopment. In F. Volkmar, A. Klin, and R. Paul, editors, Handbook of Autism and Pervasive Developmental Disorders. 3, 650–681. Hoboken, NJ: John Wiley, 2005.

27. Poels K., and Dewitte S. How to capture the heart? Reviewing 20 years of emotion measurement in advertising. Journal of Advertising Research, 46 (1), (2006).

28. Rehg J. M., Abowd G. D., Rozga A., Romero M., Clements M. A., Sclaroff S., Essa I., Ousley O. Y., Li Y., Kim C., Rao H., Kim J. C., Presti L. L., Zhang J., Lantsman D., Bidwell J., and Ye Z. Decoding Children's Social Behavior. In Proceedings of Computer Vision and Pattern Recognition, (2013), 3414-3421.

29. Silveira F., Eriksson B., Sheth A., and Sheppard A. Predicting audience responses to movie content from electro-dermal activity signals. In Proceedings of Pervasive and Ubiquitous Computing, (2013), 707-716.

30. Voit M., and Stiefelhagen R. Deducing the visual focus of attention from head pose estimation in dynamic multi-view meeting scenarios, In Proceedings of the International Conference on Multimodal Interfaces, (2008), 173-180.

31. Wedel M., and Pieters R. Eye fixations on advertisements and memory for brands: a model and findings. Marketing Sciences, 19 (4), (2000), 297-312.

32. Zhang Z., Hu Y., Liu M., and Huang T. Head pose estimation in seminar room using multi view face detectors. In Multimodal Technologies for Perception of Humans, (2007), 299-304.