Interplay between verbal response latency and physiology of children with autism during ECA interactions

Theodora Chaspari, Chi-Chun Lee and Shrikanth Narayanan

University of Southern California (USC)
Signal Analysis and Interpretation Laboratory, USC, Los Angeles, California, USA
chaspari@usc.edu, chiclee@usc.edu, shri@sipi.usc.edu

Abstract

The affective state of children with autism is not always expressed or discernible through observational cues, a phenomenon which is further confounded by vast variability across individuals on the autism spectrum. Electrodermal Activity (EDA) is a physiological signal indicative of a person's arousal and thus affording us new insights into a child's inner affective state. In this work we study EDA cues of children with autism while interacting with an Embodied Conversational Agent (ECA). EDA is affected by both cognitive and social factors. In this paper, we consider the child's verbal response latency as the overt behavioral cue and link it with his/her physiology. A classification experiment was performed to differentiate between physiological cues of high and low verbal response latency intervals, based on the assumption that different kinds of mechanisms are triggered in each case. Our results indicate that physiological patterns between short and long verbal response latencies are more discriminative for some children than others, suggesting the existence of multiple levels of cognitive and social efforts across children. They also show variable levels of arousal response, which can provide a complementary view of the observational cues.

Index Terms: Electrodermal response, verbal response latency, affective state, cognitive and social activity, autism

1. Introduction

Social and emotional deficits are one of the main traits describing Autism Spectrum Disorders (ASD). Children with autism find it more difficult to express their conscious feelings [1] and show different patterns in perceiving and conveying emotional information [2] than their typically developing peers. In light of these findings, having a way to monitor the internal state of children with autism might afford us new insights into the mechanism of their interaction and affectivity [3, 4].

The sympathetic nervous system is associated with arousal changes, caused by emotion, cognition, or attention. Sweat is a good indicator of increased sympathetic activity and can be tracked through changes in the conductance of the skin surface. Electrodermal Activity (EDA) consists a sensitive measure of skin conductance, thus can be used to provide an estimate of emotional, cognitive and other kinds of arousal [3, 5].

Children with ASD have difficulty in exchanging information especially in social interaction, and this may not be always be apparent in the observable audio-visual cues. For instance it is hypothesized that such circumstances may be associated with high levels of (internal) stress. This inherent gap between their observable behavior and their inner affective state is not well understood, and can be potentially bridged by monitoring their physiology. In this paper, we investigate the association between children's verbal response latency and their physiological state in spoken dialog settings. The duration of verbal response latencies is reported to be very important for children, as it can be indicative of their sometimes conflicting mental procedures [6]. Since EDA reflects aspects of the underlying mental state, and specifically of the amount of socio-cognitive load [5], we explore if physiological signals of long and short verbal response latency regions should exhibit different feature-level patterns. The quantitative analysis of physiological cues during the response latency areas can provide a better understanding of a child's behavior. Similar links between verbal latency and physiology have been studied for assessing anxiety levels [7], where response latencies and the corresponding physiological data were distinct for low and high anxiety subjects. To the best of our knowledge, the link between EDA and verbal response latency has not been examined extensively, and this analysis, motivated also by observations on our data (Section 3), is an effort to study this association.

Our data come from the "Rachel ECA Interaction Corpus," containing spoken dialog recordings of a child interacting with an ECA, named Rachel, and his/her parent. The Rachel system [8] aims to encourage children to participate in social interactions and display their emotional reasoning abilities (Section 2).

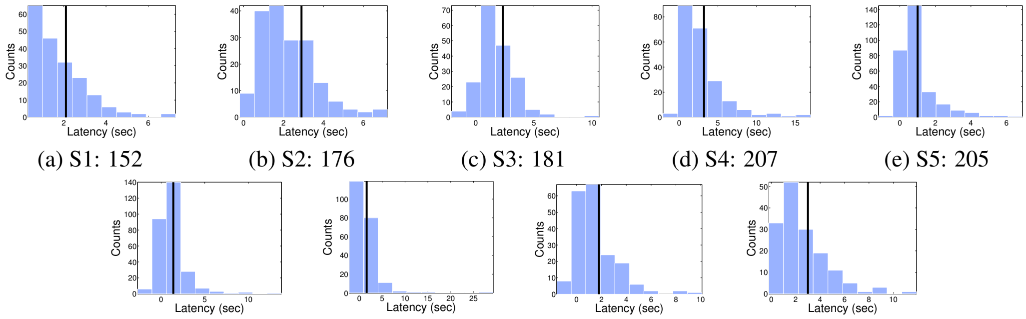
In this paper we differentiate between children's physiological patterns occurring after Rachel's turns according to whether they belong to a short or long verbal response latency interval. Based on the assumption that long latencies are likely to be caused by demanding cognitive events and/or stressful social interactions, we hypothesize that inner mechanisms produced in these intense affective activity intervals will be reflected on children's physiology. We test this by automatically classifying the child's EDA cues following Rachel's questions into whether they occur in a short or long verbal response latency region.

Our analysis through classification experiments indicates that EDA patterns convey information about a child's inner state, since they differ according to the duration of verbal response latency with respect to Rachel's turns (Section 5.2). Our findings also show that aroused affectivity is present both in low and high verbal response latency intervals, depending on the child (Section 6). This suggests that even though there may be no obvious (audible/visible) signs of arousal, EDA might give a complementary view of a child's state, as proposed in other studies [9]. This information could be incorporated in dynamically personalized dialog interfaces that are targeted for this special population.

2. Description of Data

Our experiments were based on data from the "USC Rachel ECA Interaction Corpus." This ECA used controlled interaction sce-

This research was funded by NSF and NIH.



(a) S1: 152 (b) S2: 176 (c) S3: 181 (d) S4: 207 (e) S5: 205

(f) S6: 245 (g) S7: 212 (h) S8: 57 (i) S9: 81

Figure 1: Total count (written at the legend) and distribution of the nine subjects' verbal response latencies (in seconds) with respect to Rachel's turns. The vertical solid black bar at each subfigure represents 70th percent threshold (to the left).

narios providing a structured way to elicit natural conversational data, that can be more reliably compared both between and within the subjects. The Rachel ECA was controlled in a Wizard of Oz paradigm, in which a hidden experimenter controls Rachel's behavior using a graphical interface.

The Rachel study [8] consists of four sessions recorded with a smart-room setup, having two shot-gun and two lapel (for the child and the parent) microphones, three Sony High-Definition cameras and four Affectiva Q Sensors. The sensors were worn by the child and the parent on each of their opposite-side wrist and ankle. They measure EDA, temperature and x, y, z-axis acceleration. For the purpose of this paper, we only used the EDA values measured on children's wrist.

This study analyzed data collected from nine verbally fluent subjects diagnosed to be on the Autism spectrum, seven boys and two girls, whose ages are given in Table 1. During the recordings, we made sure to keep consistent the Wizard-Of-Oz moderator for each subject. Every child participated in 4 separate sessions with Rachel, each lasting on average 25 minutes.

Table 1: *Subjects' age information.*

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9
Age (years)	12	7	10	7	7	6	8	7	8

3. Verbal Response Latencies

"Verbal response latency" is the duration of the time interval in which there is a turn-taking between two interlocutors. In this study, we examined children's verbal response latencies with respect to Rachel's turns. We chose to examine this type of response latency, because Rachel's behavior is controllable and remains consistent across subjects, thus minimizing the effect of the other interlocutor variability on the child's behavior.

Plotting the histograms of verbal response latencies for each child, we notice that they usually appear to be skewed towards the left. This means that the most common response latencies are short, while long latencies are more rare. Based on this observation, we drew a threshold at the 70^{th} percentile of latency values, which was computed from the data of each child separately, in order to distinguish between short and long verbal response latencies (Figure 1). Negative values of this measure mean that the child started talking before Rachel had finished the current turn.

Going through various turn taking instances, we made several interesting observations. There were examples during the dialog when Rachel asked a yes/no question to the child, such as whether he/she had played on a computer before, and then to follow up she asked an open question, e.g. to describe the kinds of games he/she plays on the computer. In the first case, where the question is simple and elicits low cognitive effort, short verbal response latencies occurred, while long latencies occurred in the second case, where the child had to be more mentally alert in order to provide a description. Long verbal response latencies also happened at the beginning of the joint story-telling task with Rachel, probably because the child tried to adapt to Rachel's questions. There were also many examples of children who took a long time to respond after having given a wrong answer once and were asked to try again, which might have worked as a stressful stimuli for them. In these long latency examples, it is reasonable to assume that high cognitive activity or stressor events occurred. Since physiological signals, and particularly EDA, are linked with these phenomena, we hypothesize that physiological patterns belonging in regions of long latency will be different than the corresponding patterns of short latency regions.

An example of interplay between verbal response latencies and physiological signals is shown in figure 2. The blue solid line in the figure represents the EDA signal of the child in a given moment of a session. The red dashed and black dashed-dotted vertical lines are positioned at the end of Rachel's turns and the corresponding beginning of child's turns respectively. We notice that in instances where the dashed and dashed-dotted vertical lines are very close to each other, the EDA signal seems more constant. In cases however where there is a long interval between Rachel's and the child's turns, there appears a change in the EDA signal. It is noteworthy that the change of EDA might not occur instantly, but can happen a few seconds after the stimulus. Similar slow variation of physiological signals, including EDA, have been observed in a study recognizing human frustration [10]. Motivated by this, we analyze the child's EDA signals after Rachel has stopped speaking over a time interval of duration ranging from 5 to 10 seconds. We call these "Rachel's turns effect intervals" and a graphical representation of three such consecutive intervals is shown in figure 3.

4. Extraction of Physiological Features

The sampling frequency of EDA signals was 8Hz and their values were measured in micro-Siemens (μS). To remove noise artifacts, EDA signals were filtered with a Hanning window of length 200 points (25 seconds). Three groups of features were extracted: i) time-, ii) extrema- and iii) frequency-based features. These were inspired from other behavioral studies [10, 11]. Mo-

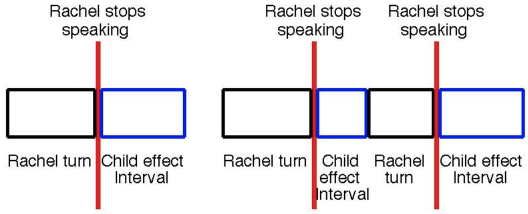


Figure 3: *Plot of Rachel's turns effect intervals.*

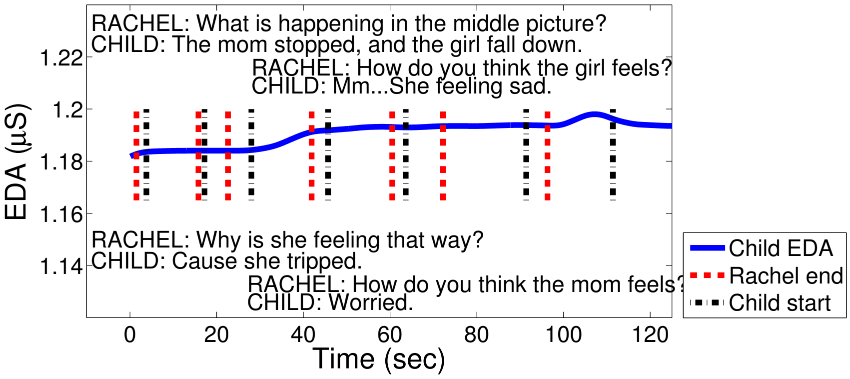


Figure 2: *Child's raw physiological signal (solid line), ending of Rachel's turns (dashed vertical line), starting of child's turn (dashed-dotted vertical line), and corresponding transcripts for three consecutive verbal response latency intervals.*

tivated by our observation that the EDA response might not occur instantly (Section 3), we computed all features over varying duration windows of 2,4,6,8 and 10 seconds and with 1 second step.

Let x denote the original filtered EDA signal measured from the child's wrist and $x(n)$ its n^{th} sample.

Time-based features were extracted using the original filtered signal and its dynamic representation, and are:

1. $\{\hat{\mathbf{x}}_\mathbf{i}\}_{\mathbf{i=1}}^{\mathbf{4}}$: the first four moments of x.
2. $\left\{\hat{\delta}_\mathbf{i}^\mathbf{1}, \hat{\delta}_\mathbf{i}^\mathbf{2}\right\}_{\mathbf{i=1}}^{\mathbf{4}}$: the first four moments of the absolute 1st and 2nd order difference of x, defined as $\delta^1(n) = |x(n) - x(n-1)|$ and $\delta^2(n) = |x(n) - x(n-2)|$.
3. $\{\hat{\epsilon}_\mathbf{i}^\mathbf{1}, \hat{\epsilon}_\mathbf{i}^\mathbf{2}\}_{\mathbf{i=1}}^{\mathbf{4}}$: the first four moments of the relative absolute 1st and 2nd order difference of x, defined as $\epsilon^1(n) = |(x(n) - x(n-1))/x(n)|$ and $\epsilon^2(n) = |(x(n) - x(n-2))/x(n)|$.

Time-based features 2 and 3 were extracted in order to examine both absolute and relative values of the dynamic representation of the EDA signal, since the first reflects the amount of arousal, while the second shows the signal's trends.

The second group of features was related to extrema values. We separated the extrema into peaks and valleys and computed their height and width (in seconds). In a given window interval we extracted the following: number of peaks (Pn), number of valleys (Vn), mean peak height (Pmh), mean peak width (Pmw), mean valley height (Vmh), mean valley width (Vmw), ratio of mean peak height to mean peak width (Phwr), ratio of mean valley height to mean valley width (Vhwr).

The third group involved frequency domain characteristics. Since EDA is a low frequency signal, we only examined frequencies in the interval $[0, 4]$ Hz. We computed the spectrogram of x using $K = 64$ sample Fourier Transform and for all mentioned window lengths. Let $F(m, k)$ be the spectrogram value of the m^{th} window ($m = 1, \ldots, N$) at the k^{th} frequency bin ($k = 1, \ldots, K/2+1$). In order to smooth this detailed frequency resolution, we separated the 33 frequency bins into 7 bands of 5 bins (except the last band with 3 bins) named b^i, $i = 1, \ldots, 7$. Based on these we computed the following:

1. $\left\{\hat{\mathbf{b}}^\mathbf{i}\right\}_{\mathbf{i=1}}^{\mathbf{7}}$: the mean over all bins in each band i.
2. $\left\{\hat{\mathbf{d}}^\mathbf{i}\right\}_{\mathbf{i=1}}^{\mathbf{7}}$: The absolute 1st order difference of $\hat{b}_i$ across windows, i.e. $\hat{d}^i(m) = \left|\hat{b}^i(m) - \hat{b}^i(m-1)\right|$.
3. $\{\hat{\mathbf{e}}^\mathbf{i}\}_{\mathbf{i=1}}^{\mathbf{6}}$: The absolute 1st order difference of $\hat{b}_i$ across frequency bands, i.e. $\hat{e}^i(m) = \left|\hat{b}^{i+1}(m) - \hat{b}^i(m)\right|$.

This feature extraction process resulted in a 240-dimensional feature vector (48 features for each window length).

5. Experiments

The purpose of our experiments is to show that there is a direct link between levels of socio-cognitive demand and affective mechanisms. Through a classification task we will show that children's physiological patterns differ between regions of long and short verbal response latency and that there exist multiple arousal levels across subjects.

5.1. Methods

5.1.1. Feature Selection

To reduce the dimensionality of the original feature vector we sorted the features (Section 4) according to the Fisher discriminant ratio criterion. We then computed the correlation of every pair and out of each pair with correlation larger than 0.5, we omitted the feature with the lower Fisher discriminant ratio.

5.1.2. Classification

We used K-nearest neighbor (KNN) to classify between short and long verbal response latency. The experiments were done using leave-one-instance-out cross-validation, where "instance" denotes a verbal response latency instance. We experimented with number K of nearest neighbors between 1 to 40 with step 5. In each fold, we performed feature selection based on the train data and then selected the same features on the test data. We did these experiments for each child separately, because we wanted to study the unique individual trends of each child with respect to their behavior and their physiology.

5.2. Results

Our classification results range between 50.30% and 70.30%, suggesting that EDA signals contain information relevant to the amount of verbal response latency. We notice a great difference in performance across subjects, reflective of the heterogeneity prevalent in this population. Specifically EDA cues of subjects S3, S4, S6 and S9 seem more reflective of the kind of verbal response latency than the corresponding patterns of subjects S1, S2 and S7. We will discuss more about this in section 6. The unweighted classification accuracy (with chance 50%) for $K = 15$ nearest neighbors is shown in Table 3. $K = 15$ was empirically found to give better performance

Examining the classification performance for the different lengths of Rachel's turns effect intervals, we notice that the most discriminative duration lies between 5 and 8 seconds. The performance usually drops for effect intervals larger than 9 seconds, suggesting that the effect of Rachel's turn stimuli to the child's inner response degrades after this time. We also notice that the duration of the best effect interval is not consistent across children, enhancing the original assumption of variability across subjects.

We examined the most frequently selected features over all lengths of Rachel's turns effect intervals for each child (Table 2). It is noteworthy that for all children the most usually selected features are the first three order moments of the physiological signal. This suggests that time domain features, mainly associated with the arousal levels, are important for our task. The next most frequent features differ across children and belong to all three groups of features described in Section 4. This finding supports the assumption of uniqueness of personal traits across children.

Table 2: *Most frequently selected EDA features for classifying verbal response latency.*

Subj.	Most frequently selected EDA features
S1	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, Pmh, $\hat{b}_2$, Vhwr
S2	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, Pmh, Vmw, $\hat{b}_6$
S3	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, Pmh, $\hat{b}_2$, δ_3^2, ϵ_3^1
S4	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, Pn, $\hat{d}_1$, $\hat{e}_6$
S5	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, Vmh, Vhwr, δ_1^1
S6	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, Vmh, $\hat{d}_1$, δ_1^1, ϵ_1^2
S7	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, $\hat{b}_6$, δ_1^1, δ_1^2
S8	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, Pmh, Vhwr, $\hat{b}_6$
S9	$\hat{x_1}$, $\hat{x_2}$, $\hat{x_3}$, Pmh, $\hat{b}_6$, δ_1^1,δ_1^2

Table 3: *Unweighted success percentages of verbal response latency type classification based on physiological cues for different length of Rachel's turns effect intervals.*

Subject	Length of Rachel's turns effect intervals (sec) 5	6	7	8	9	10
S1	**50.30**	49.00	47.92	47.41	46.75	47.54
S2	**51.90**	44.43	47.21	47.55	48.20	48.85
S3	56.04	56.37	59.09	**60.54**	59.57	59.00
S4	**63.36**	62.30	59.89	59.16	55.16	57.09
S5	50.85	50.33	51.67	54.44	53.78	**55.53**
S6	58.52	**58.97**	57.95	57.37	58.29	57.80
S7	**50.83**	50.09	50.02	50.43	49.53	49.91
S8	**54.91**	53.95	50.93	48.16	46.07	45.70
S9	67.35	66.90	68.98	**70.30**	69.36	68.50

6. Discussion

In section 5.2, we saw that there is wide variability across subjects with respect to our task. This suggests that there might be mechanisms triggered in subjects with high classification accuracy, reflected onto their physiological signals, which are not present in subjects with low classification accuracy.

Looking at the original EDA signals, we notice that there is a difference in the arousal levels between the two types of verbal response latencies, depending on the child. For each subject we perform a bootstrap hypothesis test on the difference of means of x measured over the intervals of short and long verbal response latencies. Subjects S1 and S2 have significantly lower arousal levels during short response latency intervals. This might suggest that EDA of these children is consistent with the task effect, since long latency instances are usually associated with more demanding socio-cognitive tasks. However, this is not the case for subjects S3, S5, S8 and S9, for which the mean arousal at short latency instances is greater than the corresponding arousal at long latency instances. This high arousal during a generally low mental load situation could indicate the underlying presence of intense affectivity for subjects with ASD, not directly triggered by environmental stimuli and not always obvious though traditional observational cues. This is an important observation, with implications for developing analytical tools of stratifying social interaction behavior of children, and in informing interface designs that are sensitive to the child's true affective state.

Linking our classification results with the above statistical test we could assume that subjects' S1 and S2 EDA responses are more similar to the typical population. In fact, going through the audiovisual recordings, we observed that these subjects seemed comfortable in responding to Rachel's turns no matter the task, consistent with the reasoning why their EDA cues were not very informative with respect to the verbal response latency type. In contrast, subjects S3 and S9 sometimes appeared troubled and confused, indicating high stress and cognitive activity levels reflecting on their EDA response as well.

Table 4: *EDA means (in μS) over short and long verbal response latency intervals and p-values of bootstrap hypothesis test on the difference of means for each child (0* denotes <0.01).*

Subject	S1	S2	S3	S4	S5	S6	S7	S8	S9
short	0.027	0.7	0.52	0.54	7.26	0.81	0.64	0.55	1.29
long	0.029	0.93	0.37	0.53	7.06	0.85	0.63	0.51	1.01
p-value	0*	0*	0*	0.42	0*	0.11	0.64	0.01	0*

7. Conclusions and Future Work

This study provides a novel analysis of physiological signals of children with autism in association with their expressive behavioral cues. The results suggest that EDA response of children with autism can reflect the amount of their verbal response latency with respect to a stimuli and can be further related to the amount of underlying socio-cognitive activity, not always obvious through traditional observational methods.

One limitation of this study is that it relies on observational cues concerning only turn-taking duration measures. Future work plans to examine expressive cues with a more detailed analysis of children's acoustic, lexical and visual gestural features, in order to see if these can be linked with their inner physiological signals. In our future work we will also explore if the concurrent parent's physiology can give additional information on observational cues and if it can be associated with the child's physiological state.

8. References

[1] Shalom, D. B., Mostofsky, S. H., Hazlett, R. L., et al.,l "Normal Physiological Emotions but Differences in Expression of Conscious Feelings in Children with High-Functioning Autism,' JADD, 36(3):395-400, 2006.

[2] Rieffe, C., Terwogt, M. M., and Stockmann, L., "Understanding Atypical Emotions Among Children with Autism,' JADD, 30(3):195-203, 2000.

[3] Picard, R. W., "Future affective technology for autism and emotion communication,' Phil. Trans. R. Soc., B 364:3575-84, 2009.

[4] Welch, K. C., "Physiological Signals of Autistic Children Can be Useful,' IEEE Instr. & Meas. Magazine, 15:28-32, 2012.

[5] Critchley, H. D., "Electrodermal responses: what happens in the brain,' Neuroscientist, 8:132-42, 2002.

[6] Atance, C. M., Bernstein D. M. and Meltzoff A. N., "Thinking about false belief: Its not just what children say, but how long it takes them to say it,' Cognition, 116:297-301, 2010.

[7] Schwartz, G. E., "Psychobiology of repression and health: a systems approach,' In Singer, J. L., "Repression and Dissociation: Implications for Personality Theory, Psychopathology and Health,' (pp. 405-470), The University of Chicago Press, 1995.

[8] Mower, E. P., Black, M., Flores, E., Williams, M. and Narayanan, S., "Rachel: Design of an emotionally targeted interactive agent for children with autism,' ICME, Barcelona, Spain, 2011.

[9] Goodwin, M. S., Groden, J., et al., "Cardiovascular Arousal in Individuals With Autism, ' Focus on autism and other developmental disabilities, 21(2):100-23, 2006.

[10] Fernandez, R. and Picard, R.W., "Signal processing for recognition of human frustration,' ICASSP, Seattle, USA, 1998.

[11] Koelstra, S., Muhl, C., Lee, J. S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A. and Patras, A., "DEAP: A Database for Emotion Analysis using Physiological Signals,' IEEE Trans. on Affective Computing, Special Issue on Naturalistic Affect Resources for System Building and Evaluation, 2011.