

USING PHYSIOLOGY AND LANGUAGE CUES FOR MODELING VERBAL RESPONSE LATENCIES OF CHILDREN WITH ASD

Theodora Chaspari, Daniel Bone, James Gibson, Chi-Chun Lee, Shrikanth Narayanan

University of Southern California (USC)

Signal Analysis and Interpretation Laboratory, USC, Los Angeles, California, USA

{*chaspari,dbone,jjgibson,chiclee*}@usc.edu, *shri@sipi.usc.edu*

ABSTRACT

Signal-derived measures can provide effective ways towards quantifying human behavior. Verbal Response Latencies (VRLs) of children with Autism Spectrum Disorders (ASD) during conversational interactions are able to convey valuable information about their cognitive and social skills. Motivated by the inherent gap between the external behavior and inner affective state of children with ASD, we study their VRLs in relation to their explicit but also implicit behavioral cues. Explicit cues include the children's language use, while implicit cues are based on physiological signals. Using these cues, we perform classification and regression tasks to predict the duration type (short/long) and value of VRLs of children with ASD while they interacted with an Embodied Conversational Agent (ECA) and their parents. Since parents are active participants in these triadic interactions, we also take into account their linguistic and physiological behaviors. Our results suggest an association between VRLs and these externalized and internalized signal information streams, providing complementary views of the same problem.

Index Terms— Verbal response latency, Electrodermal Activity, Language modeling, Autism Spectrum Disorders, Generalized linear regression

1. INTRODUCTION

Modeling human behavior with signal-derived cues affords us new quantifiable measures and insights into better understanding typical and atypical behavioral patterns [1]. Turn-taking dynamics during interactions provide a window into the participants' mental state and underlying social and cognitive processes. Especially for children, temporal conversational patterns can reflect their developmental trajectory, social abilities and mental skills [2]. One of the core symptoms of Autism Spectrum Disorders (ASD) are impairments of social interaction, in that children with ASD depict limited reciprocal conversations and reduced spontaneous communicative speech [3]. Modeling their interactional dynamics might offer us new insights into the nature of their social and cognitive behaviors.

One of the main indicators of children's quality of rapport is their Verbal Response Latency (VRL), defined as the time duration between the end of someone's turn and the beginning of the child's corresponding turn. Long VRLs provide valuable information about the children's perceived cognitive and affective state, reflected on their overt observable as well as covert inner behavioral cues. Motivated by these, we examined the VRLs of children with ASD during their interactions with an Embodied Conversational Agent (ECA), named "Rachel", and one of their parents. Children had to complete

several games and story-telling tasks, designed to elicit their cognitive and emotional reasoning abilities [4] and in parallel get engaged into an interactional triad with the ECA and their parent.

Children's response latencies to Rachel's stimuli can reflect their cognitive state, since long VRLs might occur after mentally challenging questions, or before complex answers requiring more time for preparation and processing [5]. This led us to examine children's language cues following their VRLs, potentially illuminating their cognitive effort. VRLs can be also indicative of children's inner affective state, in that long VRLs could be caused by high anxiety levels. Especially for children with ASD, where a possible gap between their externally observable behavior and their inner affective state has been reported [6], examining their physiology could provide us with a complementary view of their psychological state. For this reason, we study the association of VRLs with the children's Electrodermal Activity (EDA), which serves as a sensitive measure of their emotional, cognitive and other kinds of arousal [7, 8].

Since parents have an active role in our interactional scenario, in our modeling we also take into account their linguistic and physiological behavior with respect to the children's VRLs. This choice can be justified by the inherent child-parent synchrony reported in literature, such as a consistent mother-child behavioral coordination across time [9], or a physiological coordination affected by various pathological conditions [10].

In this paper, we hypothesize that the children's short and long VRLs can be distinguished based on the their own and their parents' language and EDA patterns. An understanding of the factors that might have caused the long VRLs could provide us valuable information on the child's mental and affective state. We validate our approach through statistical tests, indicating that the two types of VRLs can be discriminated with linguistic and physiological patterns from both the child and the caregiver. We further conduct classification experiments to discriminate between short and long VRLs and regression tasks to predict VRL duration. The unweighted classification accuracy ranges from 45.40% to 81.68% and the correlation of the target and predicted latencies lies in the interval 0.05-0.95 depending on the subject, reflecting the population heterogeneity. Our results indicate that with moderate accuracy we can predict the type and value of VRL duration based on physiological and linguistic features from the child and the parent during these kinds of interactions.

2. RELATION TO PRIOR WORK

Many studies have underscored the importance of children's VRLs with respect to various factors, including pathological conditions. Atance et al. [11] suggested that longer VRLs for incorrect responses

Thanks to NSF and NIH for funding.

indicate conflict and provide information about cognitive processes. Weismer et al. [5] showed that, when they have to wait longer before answering, children with language impairments produce more correct, complex and longer responses. Evans et al. [12] used stochastic models to represent response latencies and verbal pauses of children with language deficiencies. Finally, Heeman et al. [13] found that children with ASD take, on average, significantly longer time to respond to a question compared to typically developing children.

The connection between physiological data and VRLs has not been extensively studied. Researchers have associated VRLs with physiological cues for studying low and high anxiety subjects [14]. Our previous work has indicated that children’s EDA patterns differ with respect to their VRLs, conveying information about their inner affective state [15]. We have observed that high arousal levels are present both during long and short VRLs, depending on the subject, highlighting the complex affectivity in the population with ASD.

Behavioral similarity between interlocutors, like the one we are finding between child and parent during ECA interactions, has been observed in other studies as well. In [16] authors have explored child-psychologist attuning with respect to their prosodic characteristics during the Autism Diagnostic Observation Schedule (ADOS) interviews. Similar findings were depicted in [17] for predicting toddlers’ task engagement based on the child’s and the clinician’s acoustic cues. Finally, prosodic entrainment was found to be related with positive and negative affect during married couples interactions [18].

3. DATABASE DESCRIPTION

The data for this study come from the “USC Rachel ECA Interaction Corpus” [4] containing recorded interactions between children, an ECA and their parents. The experiments were designed to elicit social and affective behaviors and natural conversational data from children with ASD through games and narrative tasks. Rachel presents a consistent stimuli across subjects, providing structured prompting and facilitating within- and across-subjects analysis.

Our data represent 9 verbally fluent subjects with ASD (Table 1), aged between 6 and 12 years, each participating in 4 separate sessions with duration of approximately 25 minutes. We have clinical scores from 8 of these subjects, of which 4 were diagnosed with Autism and 4 with Pervasive Developmental Disorders (PDD). A child is diagnosed with PDD when he/she displays behaviors or deficits of the Autism Spectrum but does not meet the criteria for a specific disorder, such as Autism. In section 6 we will describe some of our observations for these two groups.

Interactions are recorded with audio-visual and physiological equipment. For the purpose of this paper we used human derived transcripts and EDA signals captured from the child’s and parent’s wrist and ankle using the Affectiva Q-Sensor [19]. The Q-Sensor Curve was worn on the wrist and the Q-Sensor Pod on the ankle.

Table 1. Subjects’ age and clinical information.

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

PDD: Pervasive Developmental Disorder

4. RELATION OF VERBAL RESPONSE LATENCIES WITH LANGUAGE AND PHYSIOLOGY

We describe and analyze through statistical tests the features used to capture the linguistic and physiological cues of our subjects. Short and long VRLs were differentiated based on the 70th percentile of latency values for each child separately. This threshold was computed empirically from the data, since VRL histograms were skewed

Table 2. Description of linguistic features.

Category	Feature Description
General turn descriptor	Number of words per sentence (WPS) Number of words with more than 6 letters
Sentence structure	Percentage of pronouns, (auxiliary) verbs, adverbs conjunctions (conj.), negations, quantifiers
Psychological processes	Percentage of words describing social, affect, cognitive (cogn.) and perceptual (percept.) processes
Paralinguistic dimensions	Percentage of assent, non-fluencies (non-fluen.) and fillers

Table 3. Bootstrap on difference of means between short (S) and long (L) VRLs based on linguistic features from child’s and parent’s turns ($\dagger : p < 0.1$, $* : p < 0.01$). Each table cell notes which of the two types of VRLs has greater mean on the corresponding feature.

Subj.	Child					Parent			
	WPS	Verb	Conj.	Affect	Non-Fluen.	Verb	Adverb	Cogn.	Percept.
S1	L*	L†	L*	S*	-	L*	L*	L*	L*
S2	L*	-	L*	S†	L*	L*	S*	L*	L*
S3	L*	L†	L†	-	S†	L*	S*	L*	L*
S4	-	-	-	-	L*	L*	L†	L*	L*
S5	L†	L†	-	-	-	L*	L*	L*	L*
S6	L*	L†	-	S*	-	L*	L*	L*	-
S7	L†	-	-	S†	-	L†	-	-	-
S8	L*	-	-	-	-	L*	L*	L*	L*
S9	-	L†	-	-	S†	L*	L*	L*	L*

towards the left [15]. All statistical tests were performed using bootstrap on the means of the two VRL types.

4.1. Linguistic features

Linguistic features aim to describe individuals’ word usage and sentence content. We used the Linguistic Inquiry and Word Count Toolbox (LIWC) [20] designed to highlight the various emotional, cognitive and structural components present in speech. We extracted linguistic features for the child’s turns that followed Rachel’s questions, and the parent’s turns that were uttered between Rachel and child speech. The features, described in detail in Table 2, include general turn descriptors, sentence structure information, word categories depicting psychological constructs and paralinguistic dimensions, and result in a 16-dimensional vector. In case the parent did not speak after Rachel, we set the corresponding feature sample equal to zero. Also in order not to add bias to our data, we did not include the count of words for the parent features, which has an inherent-time dependence.

We report the results of the statistical tests for the features that are more intuitively related to each person (Table 3). Children’s sentence complexity, depicted in words per sentence, verb and conjunction use, is related with their VRLs, in that most of the children tend to produce significantly more complex sentences after long VRLs. This is in agreement with previous research findings [5]. Also some of the children tend to wait for significantly less time before uttering a sentence containing affect. Interestingly subjects S3 and S9 display significantly more fillers after short VRLs, suggesting that the presence of a non-verbal vocalization allows them to have more time before responding to a question. Speech delays were shown to be associated with fillers in other studies as well [21].

From the statistical analysis of parent’s linguistic features (Table 3), higher sentence complexity, indicated by use of verbs and adverbs, can be associated with long VRLs. Also the presence of cognitive and perceptual words, like “think” or “know”, is related with long VRLs. This could indicate that during long VRLs, parent tend to provide their insight, in an effort to motivate their children.

4.2. Physiological features

EDA is a slow varying signal, with sampling frequency of 8Hz and amplitude measured in micro-Siemens (μS). To remove noise, we

Table 4. Bootstrap on difference of means between short (S) and long (L) VRLs based on mean EDA, peak height, and 1st order absolute EDA difference. Each table cell notes which of the two types of VRLs has greater mean on the corresponding feature.

Subject	Child			Parent		
	Mean EDA	Mean Peak Height	Mean 1 st EDA Diff	Mean EDA	Mean Peak Height	Mean 1 st EDA Diff
S1	-	-	-	-	-	-
S2	L*	-	-	L*	-	-
S3	S†	-	-	S*	-	-
S4	S†	-	-	L*	-	L†
S5	S*	-	-	S*	S†	S†
S6	L*	-	L†	S*	S*	S*
S7	-	-	-	L*	S†	-
S8	L*	S*	S*	S*	S†	S†
S9	L*	-	-	S*	S†	-

performed low-pass filtering with Hanning windows of two different lengths of 40 and 200 points (5 and 25 seconds). Features were computed over two different time window lengths of 4 and 8 seconds with 1 second time step. This multiple time-scale analysis enables us to keep both local variations and global trends. A similar approach was followed in [22], where they have computed features based on slow (0 – 0.2Hz) and very slow (0 – 0.08Hz) EDA response.

EDA features are extracted from all four sensor streams, i.e. the child’s and parent’s wrist and ankle. They include the first 4 moments of the windowed signal, number of peaks, mean peak height and width and their ratio, and the first 4 moments of absolute and relative 1st and 2nd order time-difference, as described in [15]. This results in a 212-dimensional feature vector; 53 features calculated over 4 different combinations of filter-length and time-scale.

All EDA feature samples were extracted after Rachel had stopped speaking and over a 7 seconds time interval. In case Rachel started speaking again before the end of these 7 seconds, features were extracted until the beginning of Rachel’s next turn. These long effect intervals taken into account after the stimuli, can be justified because of the slow varying nature of the EDA signal. In our experiments we did not include subject 1’s EDA signals from child’s wrist and ankle, since we observed that these were distorted from clipping.

We examine the effect of VRLs on mean EDA, mean peak height and mean of 1st order absolute difference with statistical tests. We report results on values from the child’s wrist and parent’s ankle, since these two streams gave better performance (see Section 5.2). Mean EDA from both the child and the parent are significantly distinct between short and long VRLs for most of the subjects (Table 4). It is also noteworthy that mean peak height and absolute 1st order difference from the parent’s EDA signal is different for more subjects compared to the same measure from the child’s ankle, providing an indication that parents’ cues can convey valuable information.

5. EXPERIMENTS

5.1. Methods

We performed classification and regression tasks to predict short and long VRLs and the VRL durations, respectively. All results are reported based on a leave-one-instance-out cross-validation scheme, where “instance” denotes a VRL instance, and feature selection was done on the respective train set and then applied on the test data.

5.1.1. Classification

Classification was performed with a K-NN classifier with $K = 15$ nearest neighbors, a value which was empirically found to give better performance. For feature selection, we kept those features with Fisher Discriminant Ratio (FDR) [23] greater than the 85th per-

centile FDR of the total feature set. This was performed in an effort to maximize the between-class and minimize the within-class distances. Furthermore, if a pair of features had correlation higher than $r = 0.7$, we omitted the feature with the lowest FDR. This resulted, on average, in 30- and 3-dimensional features for the physiological and the linguistic cues, respectively. For the final decision, we concatenated the class probabilities of short VRLs from each modality into a feature vector and fed them as an input into a K-NN classifier.

5.1.2. Regression

Motivated by the fact that VRLs are essentially waiting times with respect to a stimuli, we were able to model them with a Gamma distribution and predict their value using Generalized Linear Regression (GLR). Negative VRL duration instances, constituting 2.52% of our original data, were omitted in this task, since a random variable following the Gamma distribution can only take positive values.

Generalized linear models are an extension of linear models and contain a random and a systematic component [24, 25]. The random component has a vector of n observations \mathbf{y} which is assumed to be a realization of a random variable $\mathbf{Y} = [y_1, \dots, y_n]$ with mean μ , such that $E(\mathbf{Y}) = \mu$. The systematic component of the model contains p covariates $\mathbf{X} = [x_1, \dots, x_p]$ that produce a linear predictor $\eta = \mathbf{X} \cdot \beta = \sum_{j=1}^p x_j \cdot \beta_j$. The link between the random and systematic components is done by the link function g , such that $\eta_i = g(\mu_i)$, which for the Gamma distribution is $g(\mu_i) = 1/\mu_i$.

We tested the goodness of fit of VRL values to a Gamma and an exponential distribution, since the later is a special case of Gamma with unity shape parameter, using the Kolmogorov-Smirnov test. For subjects 4-9 we got a significant fit for both distributions ($p < 0.05$), while for subjects 1-3 only for the exponential one ($p \leq 0.01$), as shown in Table 5. We performed feature selection by keeping only those features for which fitting the GLR model resulted in significant weights ($p \leq 0.07$, value which was empirically found to yield better results). To come up with the final prediction, we concatenated the prediction values from each sensory stream into a feature vector and used these as covariates to a GLR model.

Table 5. P -values for goodness of fit of VRL values to Gamma and Exponential probability distributions according to the Kolmogorov-Smirnov test.

Distribution	S1	S2	S3	S4	S5	S6	S7	S8	S9
Gamma	0.10	0.31	0.53	0.04	<0.01	<0.01	0.01	0.02	0.06
Exponential	<0.01	<0.01	0.01	0.03	<0.01	0.02	0.01	0.07	0.03

5.2. Results

Our experimental results are reported for each modality separately, the combination of all four EDA streams, both linguistic streams and finally all modalities (Tables 6, 7). EDA signals from children’s wrist seem to be more indicative for our task compared to signals from their ankle, while the opposite held for the parent. This might be due to the different sensitivity of the corresponding body parts for the two groups or because of the different kinds of sensors (Curve and Pod) worn on the wrist and ankle respectively. In many cases, like subjects 3,5 and 9 for the EDA stream, combining child and parent signals increases the classification accuracy compared to each stream separately, suggesting the presence of informative cues elicited from both people participating in the ECA interaction. Also combining all physiological and linguistic streams can improve our results, indicating the existence of complementary information in these channels. The results from the final fusion of all modalities for most of the subjects appear to have similar trends for both the classification and

Table 6. Unweighted classification accuracies (%) for classifying short and long VRLs based on child’s and parent’s physiological and linguistic cues and their combinations.

Subject	Physiological				Linguistic			All	
	Child		Parent		All	Child	Parent		
	wrist	ankle	wrist	ankle					
S1	–	–	51.01	54.65	67.48	59.76	50.00	62.22	67.48
S2	55.68	45.40	48.04	54.05	53.54	55.99	45.23	58.10	45.40
S3	58.13	51.05	51.40	60.66	61.08	64.95	50.00	72.53	71.03
S4	52.96	49.37	50.84	54.03	55.95	52.00	59.43	62.76	50.59
S5	58.35	50.94	47.74	54.86	61.49	49.61	65.83	63.39	71.95
S6	54.47	49.94	51.62	57.35	57.75	58.35	50.00	63.01	59.81
S7	48.68	57.08	53.63	45.38	55.97	54.35	51.72	50.00	57.11
S8	63.82	49.69	45.73	61.63	61.51	65.44	69.64	67.40	81.68
S9	53.72	50.59	53.42	52.77	55.10	49.69	69.36	65.37	68.37

Table 7. Correlations of target VRL and predicted values from regression based on child’s and parent’s physiological and linguistic cues and their combinations ($\dagger : p < 0.1$, $*$: $p < 0.01$).

Subject	Physiological				Linguistic			All	
	Child		Parent		All	Child	Parent		
	wrist	ankle	wrist	ankle					
S1	–	–	0.03	0.06 \dagger	-0.09	0.30*	0.14 \dagger	0.27*	0.32*
S2	0.14*	0.14*	-0.06 \dagger	-0.06 \dagger	-0.09	-0.15 \dagger	0.10	0.07	0.05
S3	0.15*	0.00	-0.05 \dagger	-0.11*	0.17	-0.00	0.69*	0.50*	0.95*
S4	-0.06 \dagger	0.00	0.07*	0.03	-0.22*	-0.18*	0.39*	0.33*	0.22*
S5	0.10*	0.04 \dagger	-0.01	0.10*	0.15\dagger	0.13 \dagger	-0.01	0.06	0.17\dagger
S6	0.11*	0.04	0.06 \dagger	0.21*	0.16 \dagger	0.20*	0.12 \dagger	0.17*	0.24*
S7	0.00	0.01	-0.01	0.06 \dagger	-0.16 \dagger	0.15 \dagger	0.18 \dagger	0.13 \dagger	0.13 \dagger
S8	0.14*	-0.05 \dagger	-0.08 \dagger	0.06 \dagger	-0.06	0.15 \dagger	0.44*	0.37*	0.46*
S9	0.16*	0.04	0.13*	0.04	0.18\dagger	0.14 \dagger	0.15 \dagger	0.07	0.18 \dagger

regression task and to give improvement over the separate streams, suggesting that linguistic and physiological features can give us consistent information about the corresponding VRLs. The occurring negative correlations in the regression task might be caused by lack of informative cues from the corresponding signals.

6. DISCUSSION

We observe some interesting effects through error analysis. There were many cases in which we noticed a lag between the actual VRL instance and the correct decision from the EDA stream. For example, the posterior probability from the EDA for some long VRL instances was a bit less than 0.5, classifying the corresponding VRL as short, but in the next instance it increased above 0.5, signifying a long VRL. This could suggest that the physiological response can happen after a variable time interval and the internal state might be more slowly varying than the verbal state. Finding efficient ways to model this could give us more insights into a person’s affectivity.

Errors occurred in ambiguous instances where the presence of a long VRL couldn’t have been justified by a high cognitive or social demand. Such examples were simple tasks in which children had to count the number of objects in a picture and children usually showed long VRLs. However the cognitive load was obviously small in these tasks, therefore the EDA signal could not capture indices of long VRLs, and the child’s verbal response was just a number, making it very difficult for the linguistic stream also to predict this long VRL. Long VRLs after Rachel’s statements or affirmative turns sometimes also failed to be correctly classified. For instance a long VRL after Rachel’s turn “I get really excited with a new puppy” could not have been explained by any signs from the child.

Examining the posterior probabilities from each modality with respect to the various tasks, resulted in interesting observations about the interplay between the physiological and linguistic information streams. During the emotional matching games, in which a specific emotion from a set of images has to be identified, children usually ut-

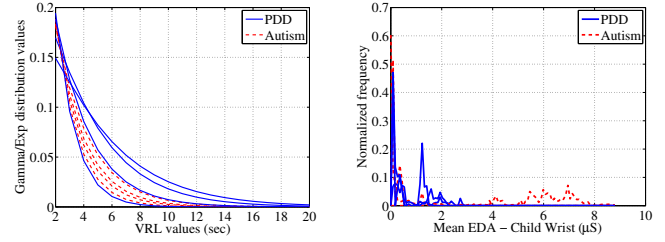


Fig. 1. (a) Gamma/Exponential VRL distributions with parameters computed using Maximum Likelihood Estimation (MLE) for subjects with PDD and Autism. (b) Distribution of mean EDA during VRL instances for subjects with PDD and Autism.

tered a letter corresponding to the image they had selected. The linguistic stream accounted for a short VRL, since there was no sign of language complexity or hesitation. The only way we could capture a long VRL in these cases was through the physiological modality, in which high cognitive load was manifested. During story-telling tasks, the child usually depicted long VRLs, for which linguistic cues in many cases gave the correct answer, since the child’s sentence involved more complex linguistic features. Motivated by these, task-specific analysis could be worth exploring in the future.

In an effort to obtain further insight about the VRLs of children with ASD, we looked for possible differences between the groups with Autism and PDD with respect to their VRLs. We fitted the VRL values of each child to a Gamma (or an exponential for subjects 1,2,3) distribution using Maximum Likelihood Estimation (MLE). We observe that most of the children with PDD tend to have higher VRL values, and that there is a visual differentiation between subjects with Autism and PDD with respect to the resulting distributions (Figure 1a). We also notice differences on the mean EDA distributions of each subject from the child’s wrist during VRL instances (Figure 1b). One subject with Autism has very high and another one very low EDA levels during VRLs, while two subjects with PDD seem to have EDA values in the middle. Similar findings exist concerning the presence of two groups with different EDA levels (high and low) within the autism spectrum, without this being related to an ASD outcome [26]. Although some of these observations might be due to person dependencies because of the small number of our subjects, they could indicate the presence of quantitative differences expressed during VRLs, which can possibly enhance our understanding about the clinical behavior of these two populations.

7. CONCLUSIONS AND FUTURE WORK

This study provides a novel analysis of VRLs of children with ASD while interacting with an ECA and their parents based on physiological and linguistic cues of the child and the parent. Our results demonstrate that the child’s physiological and linguistic behavior can indicate the nature of VRL. Parent’s cues contain complementary information for our task, suggesting that the parent is an active member in these types of triadic ECA-child-parent interactions.

In our future work, we plan to examine the specific types of ECA stimuli, for example open and closed questions, that result in long VRLs by performing speech act tagging of Rachel’s turns. We will further investigate different types of linguistic measures, like sentence structural coherence, since long VRLs might be followed by children’s unstructured sentences, as observed in our data. Finally, we plan to study possible time-dependencies across VRLs by using dynamic models, that can more appropriately capture the time evolution of these interactive conversations.

8. REFERENCES

- [1] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Invited paper. Accepted. Proc. IEEE*, 2012.
- [2] S. E. Tripp, "Children's verbal turn-taking," *Developmental Pragmatics*, pp. 391–414, 1979.
- [3] C. Lord, S. Risi, L. Lambrecht, E. H. Cook Jr., B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "Autism Diagnostic Observation Schedule - Generic: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of Autism and Developmental Disorders*, vol. 30, pp. 205–223, 2000.
- [4] E. Mower, M. Black, E. Flores, M. Williams, and S. Narayanan, "Rachel: Design of an emotionally targeted interactive agent for children with autism," *ICME, Barcelona, Spain*, July 2011.
- [5] E. Weismer and S. Schraeder, "Discourse characteristics and verbal reasoning: Wait time effects on the performance of children with language learning disabilities," *Exceptionality Education Canada*, vol. 3, pp. 71–92, 1993.
- [6] M. S. Goodwin, J. Groden, W. F. Velicer, L. P. Lipsitt, M. G. Baron, S. G. Hofmann, and G. Groden, "Cardiovascular arousal in individuals with autism," *Focus on autism and other developmental disabilities*, vol. 321, pp. 100–123, 2006.
- [7] H. D. Critchley, "Electrodermal responses: what happens in the brain," *Neuroscientist*, vol. 8, pp. 132–142, 2002.
- [8] R. W. Picard, "Future affective technology for autism and emotion communication," *Phil. Trans. R. Soc.*, vol. 364, pp. 3575–3584, 2009.
- [9] T. B. Brazelton, B. Koslowski, and M. Main, "The origins of reciprocity: the early mother-infant interaction," in *The effect of the infant on its caregiver*, pp. 49–76. New York: Wiley, 1974.
- [10] R. Feldman, "Parent-infant synchrony: Biological foundations and developmental outcomes," *Current directions in psychological science*, vol. 16, pp. 340–345, 2007.
- [11] C. M. Atance, D. M. Bernstein, and A. N. Meltzoff, "Thinking about false belief: Its not just what children say, but how long it takes them to say it," *Cognition*, vol. 116, pp. 297301, 2010.
- [12] J. L. Evans and K. Viele, "Response latency and verbal complexity: stochastic models of individual differences in children with specific language impairments," *Journal of Speech, Language and Hearing Research*, vol. 40, pp. 754–764, 1997.
- [13] P. A. Heeman, R. Lunsford, E. Selfridge, L. Black, and J. van Santen, "Autism and interactional aspects of dialogue," *Proceedings of the SIGDIAL 2010 Conference*, 2010.
- [14] G. E. Schwartz, "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.
- [15] T. Chaspari, C. C. Lee, and S. Narayanan, "Interplay between verbal response latency and physiology of children with autism during ECA interactions," *Proceedings of InterSpeech, Portland, OR*, 2012.
- [16] D. Bone, M. P. Black, C. C. Lee, M. E. Williams, P. Levitt, S. Lee, and S. Narayanan, "Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist," *Proceedings of InterSpeech, Portland, OR*, 2012.
- [17] R. Gupta, C. C. Lee, D. Bone, S. Lee, and S. Narayanan, "Acoustical analysis of engagement behavior in children," *Proceedings of Workshop on Child, Computer and Interaction (WOCCI)*, 2012.
- [18] C. C. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Computing vocal entrainment: A signal-derived PCA-based quantification scheme with application to affect analysis in married couple interactions," *In Press: Computer Speech and Language*, 2012.
- [19] M. Z. Poh, N. C. Swenson, and R. W. Picard, "A wearable sensor for unobtrusive, long-term assessment of electrodermal activity," *IEEE Trans. on Biomedical Engineering*, vol. 57, pp. 1243–1252, 2010.
- [20] J. W. Pennebaker, M. E. Francis, and Booth R. J., "Linguistic inquiry and word count (LIWC): LIWC2001," *Mahwah: Lawrence Erlbaum Associates*, 2001.
- [21] H. H. Clark and J. E. F. Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, pp. 73–111, 2002.
- [22] S. Koelstra, C. Muhl, Yazdani Lee, J. S., T. A., Ebrahimi, T. Pun, A. Nijholt, and A. Patras, "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*, vol. 3, pp. 18–31, 2011.
- [23] R. O. Duda, P. E. Hart, and D. G. Stork, "Multiple discriminant analysis," in *Pattern Classification*. John Wiley & Sons, second edition, 2000.
- [24] P. McCullagh and J. A. Nelder, "The components of a generalized linear model," in *Generalized Linear Models*. New York: Chapman & Hall, second edition, 1989.
- [25] A. J. Dobson, "Exponential family and generalized linear models," in *An Introduction to Generalized Linear Models*. New York: Chapman & Hall, second edition, 2002.
- [26] S. A. Schoen, L. J. Miller, B. Brett-Green, and S. L. Hepburn, "Psychophysiology of children with autism spectrum disorder," *Research in Autism Spectrum Disorders*, vol. 2, pp. 417–429, 2008.