Acoustic-Prosodic and Physiological Response to Stressful Interactions in Children with Autism Spectrum Disorder

Daniel Bone[1], *Julia Mertens*[2], *Emily Zane*[2], *Sungbok Lee*[1], *Shrikanth Narayanan*[1], *Ruth Grossman*[2,3]

[1]Signal Analysis and Interpretation Laboratory (SAIL), USC, Los Angeles, CA, USA
[2]Face Lab, Emerson College, Boston, MA, USA
[3]Department of Communication Sciences and Disorders, Emerson College, Boston, MA, USA

dbone@usc.edu, http://sail.usc.edu

Abstract

Social anxiety is a prevalent condition affecting individuals to varying degrees. Research on autism spectrum disorder (ASD), a group of neurodevelopmental disorders marked by impairments in social communication, has found that social anxiety occurs more frequently in this population. Our study aims to further understand the multimodal manifestation of social stress for adolescents with ASD versus neurotypically developing (TD) peers. We investigate this through objective measures of speech behavior and physiology (mean heart rate) acquired during three tasks: a low-stress conversation, a medium-stress interview, and a high-stress presentation. Measurable differences are found to exist for speech behavior and heart rate in relation to task-induced stress. Additionally, we find the acoustic measures are particularly effective for distinguishing between diagnostic groups. Individuals with ASD produced higher prosodic variability, agreeing with previous reports. Moreover, the most informative features captured an individual's vocal changes between low and high social-stress, suggesting an interaction between vocal production and social stressors in ASD.

Index Terms: stress, acoustic-prosody, physiology, autism spectrum disorder, interaction

1. Introduction

Stressors are pervasive in our daily lives, impacting our mood, our general sense of well-being, and even our health [1]. In fact, our ability to deal with and adapt to stress is associated with positive health outcomes. Anxiety disorders are the most prevalent disorder in the United States, affecting 18% of the population [2]. When stress causes anxiety, it leads to increased physiological arousal in the body [3], which we express in our verbal and non-verbal behavior. One such stressor is social anxiety, as with public speaking. Under stress, a person experiences unconscious sympathetic responses; e.g., the laryngeal folds may tighten, leading to a rise in vocal pitch [4]. But there is still much to learn about the ways in which individuals experience and express stress; one viable approach uses scalable objective measures of behavior, i.e., Behavioral Signal Processing [5].

A meta-analytic study reported that social anxiety occurs more frequently in individuals with autism spectrum disorder, or ASD [6], affecting 40% of the population. ASD is a highly heterogeneous, highly prevalent (1 in 68 [7]) neurodevelopmental disorder defined by impairments in social communication and reciprocity, as well as restricted, repetitive behavioral patterns and interests [8]. Given the prevalence of anxiety in ASD, researchers are striving to better understand when individuals become stressed and how they respond (e.g., from skin conductance responses [9]). Since it can be difficult for those with ASD to understand and communicate their emotions, acoustic analyses may provide an effective measurement of stress.

There has been limited work specifically focused on the acoustic correlates of stress, likely due to the challenges of collecting high-quality, naturalistic speech under stress [10]. Speech researchers have primarily focused on optimizing emotion classification within a database [11], whether the target is categorical or dimensional (i.e., arousal, valence, dominance). Yet, studies continue to find such models tuned for one database do not readily transfer to another [12], which is critical to the realization of speech-based behavioral health systems operating "in the wild". Approaches have included knowledge-inspired system design [13], unsupervised neural-network adaptation [14], and multimodal behavioral integration [15, 16]. A survey article by Juslin and Scherer reported several measures that reliably increase with arousal or stress: pitch and intensity mean and variability; the ratio of high-frequency energy; and speaking rate [17]. In this study, we extract corresponding features, but use functionals that were found to be more robust for tracking arousal such as median or interquartile ratio [13].

The present work builds upon several of our previous studies which sought acoustic correlates of the "atypical prosody" so commonly observed in autism spectrum disorders [18, 19, 20, 21, 22]. Our experiments [18, 19, 20] in a sample of 29 children from the USC CARE Corpus [23] found children with increasing ASD severity spoke less, spoke slower, responded later, had more variable prosody, and had more atypical voice quality. Since atypicality is not universal in ASD, we have also investigated human perception of atypicality or "awkwardness". We found that human agreement can be rather low for very specific dimensions of prosody, but that speech rate and rhythm cues were highly predictive of overall perceived "awkardness" in the read speech task [21]. In a large-scale study, we found that prosodic variability was significantly higher for individuals with ASD compared to peers with non-ASD developmental disorders [22], aligning with previous findings in smaller databases [24, 25]. Additionally, we presented novel features that measured reduced coordination between pitch and intensity or duration, quantifying a previous qualitative perception [26].

Previous work has primarily focused on a single modality; in this novel study, we investigate the multimodal presentation of stress in individuals with ASD and their neurotypically developing peers as they participate in a series of progressively stressful interactions. As a measure of latent physiology, we consider mean heart rate, which generally correlates with acute increases in stress [27]. We also explore a set of acoustic-prosodic features that are expected to be modulated by changes in affect [13] as well as ASD symptoms [22], observing global tendencies as well as changes that occur within a person between different tasks. Through this study, we aim to enhance our understanding of signal-derived measures of stress, which are crucial to development of clinical engineering systems.

2. Methodology

In this section, we discuss: three interactions of varying stress in our study; data collection and participant demographics; acoustic-prosodic and physiological features related to stress and autism; and data analysis and machine learning models.

2.1. Social Interactions of Variable Stress

Subjects participated in three types of social interactions expected to be progressively more stressful: a low stress one-on-one conversation, a medium stress one-on-one interview, and a high stress presentation to an audience. In the first task, the subjects watched YouTube clips, and then discussed the clips with a researcher (this conversation typically lasted under one minute). Because the interaction was casual (participants were not aware they were being recorded) and because the topic was impersonal, we assume that individuals experienced low levels of stress during these chats.

In the second scenario, the research assistant interviewed the subject about their hobbies, family, and school for twenty minutes (we analyze two minutes). Since questions were personal, the context was more formal, and the subject was aware they were being recorded, we expected the subjects to feel an increased level of pressure compared to the casual conversations.

The third interaction, an oral presentation, is hypothesized to be the most stressful. Subjects were to develop the ending of a story within five minutes, and then were asked to present in front of a seated audience of three adult judges, video-edited to appear as a live Skype call. Further intensifying any social anxiety, the subjects were told that their performance would be judged against their peers' performances (see [9, 28]).

2.2. Data Collection and Participants

Experimental data consists of video-recorded interactions from the three stressful scenarios for all subjects. Data are from 17 children with autism spectrum disorder (ASD) and 24 subjects with neurotypical development (TD). Participant demographics are presented in Table 1, including: ADOS diagnosis, age, and number of audio samples for each of the three tasks.

Data were collected at a single site as part of an IRB-approved study. Efforts were made to ensure video and audio quality consistency between subjects and tasks. All recordings took place in the same room. Camera microphone distance was not constant across sessions, but the distance is not known to be systematically different between groups. Still, we did not feel confident in using voice quality measures, which were previously shown to be characteristic of ASD speech [19], with the present far-field recordings; instead, we focus on prosodic measures that may be more robust to any recording variability.

Table 1: *Demographic information of all subjects presented as mean (stdv.). Differences between ASD and TD subject's age and gender are non-significant (p>0.05).*

	N	**Age in yr.**	**Female**	**Acquired Task Audio**		
				Low	Medium	High
ASD	17	13.7 (2.2)	19%	12	14	12
TD	24	13.4 (2.3)	39%	14	16	23
total	41	13.5 (2.2)	31%	26	30	34

Presence/absence of the subjects' speech was manually annotated. Audio for several sessions was not available due to recording difficulties or corrupted files. The number of session for which audio features were extracted is displayed by task in Table 1. Similar data loss occurred with heart-rate recordings. This loss primarily affects the joint audio-HR analyses, for which missing HR data reduces data size by 19%.

2.3. Acoustic-Prosodic and Physiological Features

We computed five classes of features: segmental pitch cues; segmental spectral cues; speaking rate; coordination between prosodic modalities (a novel feature type from [22]); and heart rate. Details of the feature extraction are provided below.

2.3.1. Speaking Rate

Because transcripts were not available for these data—with which we could perform forced alignment—we needed to determine syllabic boundaries directly from the audio signal. Speaking rate estimation from prosodic and spectral signals has been of some interest to the speech processing community [29, 30, 31], but accurate estimation of syllabic boundaries remains challenging. We implemented a version of a pitch- and intensity-based method that has reported competitive performance [31]. Visual inspection suggested this syllabic segmentation was adequate. We computed two features using syllable boundaries: median speaking rate (syl/s) and syllable duration inter-quartile ratio (s), or IQR.

2.3.2. Segmental Prosodic Cues: Syllabic Contours

Pitch, volume, and the percentage of high-frequency energy are all expected to increase with anxiety, stress, and arousal [17]. Further, segmental intonation that captures speaker idiosyncrasies in micro-prosodic production have been used to characterize the speech of indivdiuals with ASD [19, 21, 22]. As such, we compute nine segmental prosodic features from pitch and intensity extracted via Praat [32], as well as median HF500 (the ratio of energy above 500Hz to that below) computed via the vocal arousal score toolkit (VC-AS) [13].

In particular, we extracted syllable-level second-order polynomial parametrization of pitch and intensity, then calculated session-level medians and inter-quartile ratios of slope (four features). The overall median and IQR of both log-pitch and intensity are also calculated (four features). Aside from median log-pitch, all pitch analysis is performed in the OME (Octave MEdian) scale [33], a log-pitch transformation as in Eq. 1 through which speaker's tend to have the same pitch range, i.e., one octave.

$$\text{OME} = \log(f0_{Hz}) - \log(\text{median}(f0_{Hz})) \quad (1)$$

Since a speaker's range has been observed to reliably be one OME around center in neutral speech, all speakers should have a comparable range regardless of median pitch (unlike for Hz).

2.3.3. Prosodic Coordination Features

In previous work, we found that subjects with ASD showed reduced coordination of pitch with other prosodic markers [22]. Aside from ASD diagnosis, stress may affect this prosodic coordination. Following the same approach [22], we quantified the simultaneous movements of pitch, duration, and intensity across syllables. These three feature streams are concatenated per session, and then the Spearman's rank-correlation coefficient is calculated pairwise, producing three features.

2.3.4. Physiological measure: heart rate

A person's heart rate generally hastens under acute stress, and has been specifically shown to increase in stressful speech interactions [27]. We compute mean heart rate per session, while excluding sensor artifacts. Although heart-rate variability (HRV) is commonly employed as a robust measure of complexity differentially affected by acute versus chronic stress [27], computation requires a minimum sampling period of five minutes [34], whereas each session lasts between 30 seconds and three minutes. Unlike HRV, mean HR is robust to the sampling period.

Table 2: *Correlations of features with ADOS severity and best-estimate diagnosis. * indicates $p<0.05$; n.s. is non-significant.*

Category	Feature	Task Stress Level		ASD Diagnosis	
		Trend with task stress	*Sp. ρ*	*Trend with diagnosis*	*Sp. ρ*
Pitch cues	log-f0 median	higher	0.49^*	*n.s.*	-0.15
	log-f0 IQR	*n.s.*	-0.12	higher	0.33^*
	log-f0 slope median	*n.s.*	0.20	lower	-0.27^*
	log-f0 slope IQR	*n.s.*	0.13	*n.s.*	0.14
Spectral cues	intensity median	lower	-0.47^*	higher	0.29^*
	intensity IQR	lower	-0.40^*	higher	0.36^*
	intensity slope median	higher	0.56^*	lower	-0.22^*
	intensity slope IQR	*n.s.*	-0.20	*n.s.*	0.18
	HF500 median	lower	-0.28^*	higher	0.32^*
Speaking Rate	syllable rate median	higher	0.23^*	*n.s.*	0.10
	syllable duration IQR	lower	-0.29^*	higher	0.27^*
Prosodic Coordination	corr. f0 & dur.	less	-0.29^*	*n.s.*	0.05
	corr. f0 & intensity	less	-0.25^*	*n.s.*	0.14
	corr. dur. & intensity	less	-0.21^*	*n.s.*	0.01
Physiology	heart rate median	*n.s.*	0.20	*n.s.*	0.22

2.4. Statistical Analysis and Machine Learning

We conducted both statistical correlation analyses and classification experiments (with support vector machine via Liblinear software [35]). Parameters are tuned using two-level nested cross-validation (CV), and averaged statistics of ten runs of leave-one-subject-out CV are reported. Spearman's rank-correlation coefficient and unweighted average recall (UAR, the mean of per-class recall) are selected as evaluation metrics. Note that in cases for which only two classes exists, the p-value for Pearson's rank-correlation coefficient is equivalent to that from ANOVA; following, the same is true with Spearman's rank-correlation coefficient, apart from the initial rank-based feature transformation.

3. Results and Discussion

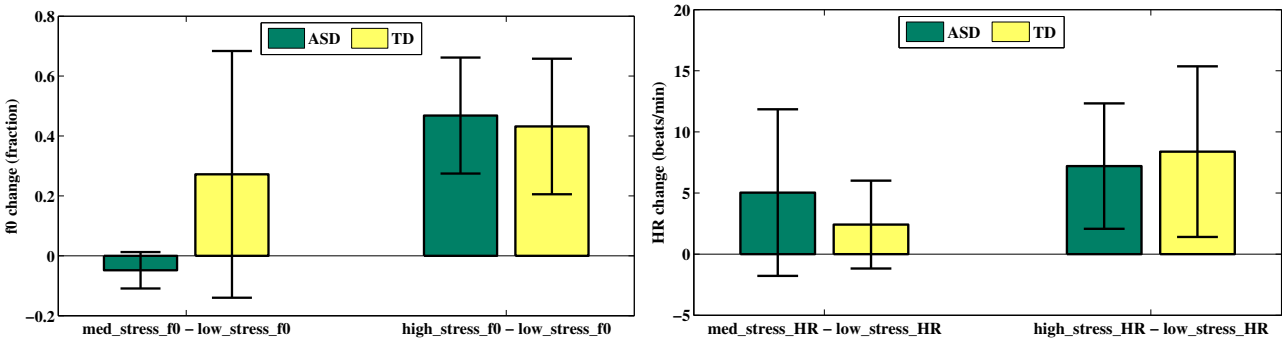
Relations between extracted behavioral features, task-induced stress, and autism spectrum disorder (ASD) diagnosis can inform large-scale behavioral analyses. In Section 3.1, the objective speech and heart rate cues are analyzed versus the hypothesized task-related stress level. Then, in Section 3.2, the cues are used to differentiate task-type and to predict ASD diagnosis.

3.1. Correlational Feature Analysis

Acoustic-prosodic and heart rate feature correlations with task-related stress level and ASD diagnosis are provided in Table 2. The low stress (casual conversation), medium stress (interview), and high stress (presentation) tasks are encoded with values of 0, 1, and 2, respectively, for purposes of analysis. Since the feature values are dependent on various sources of noise in data collection and feature extraction processes, we cannot confidently state that a construct is or is not informative of a target variable, only whether an extracted feature is in this experiment.

Segmental pitch cues capture short-term tendencies in usage of fundamental frequency. We expected median pitch to shift upward with increasing stress for a given speaker [17]; in fact, the only significant relation between pitch cues and task stress-level is that speakers tend to increase their pitch in more stressful tasks ($p<0.05$). Although one may expect pitch variability to have also increased, median pitch may be more robust, as it has been shown for a related percept, vocal arousal [13]. Regarding ASD, children with higher social-communicative deficits have previously shown more negative pitch curvature—which is possibly perceived as "flat" or "monotone" [18]—and displayed more prosodic variability [22]. For both cases, we confirm the previous findings; i.e., log-f0 variability is higher for ASD subjects, while log-f0 slope is lower. The relative (person-specific) changes in log-f0 between tasks are not significantly different between groups, while for both ASD and TD subjects there is a significant increase in log-f0 between low and high stress tasks (Figure 1a).

Segmental intensity cues may be similarly influenced by stress; however, sound level is also a function of distance and angle to the microphone. As such, we should interpret the intensity findings cautiously. Contrary to expectations, we find



(a) *Relative fundamental frequency changes between interactions.* (b) *Heart rate (HR) changes (beats/min) between interactions.*

Figure 1: *Mean relative changes (and standard-deviation) between stressful interactions: low (conversation), medium (interview), and high (presentation). For both groups and features, changes between low and high stress scenarios are statistically significant ($p<0.05$).*

that intensity median and IQR tend to decrease with increasing task stress, as does HF500 (which is a primary correlate of arousal [13]). Given the variability in audio conditions, it is difficult to ascertain if an incidental result of microphone distance is being measured, or if, in fact, subjects reduce their volume given increasing stress. We also observe that ASD subjects have higher and more variable vocal intensity.

Speaking rate is another reported correlate of perceived vocal arousal [17]. In our data, subjects in higher stress situations speak faster, but also with lower durational variability; this indicates a more rigid, tense speech production. Also, children with ASD spoke with more durational variability—yet another indicator of increased variability associated with ASD.

Following our previous quantitative support [22] of a qualitative finding [26, 36], we suspected that ASD subjects with "atypical" prosody were sometimes modulating pitch incongruously with other modalities. We quantified this prosodic coordination as the pairwise correlation between three modalities: syllabic fundamental frequency, vocal intensity, and duration. In this study, we found no statistical difference between ASD and TD groups. However, we did find that subjects in higher stress tasks tended to have less coordination between prosodic modalities–a possible result of reduced motor control as a physiological response to stress.

Lastly, we investigate our single physiological measure, mean heart rate, which is anticipated to increase with task stress. We find that overall heart rate was not significantly higher in the higher stress tasks, and that there is no relation with diagnosis. But because heart rate varies from person to person (due to general health, respiration rate, etc.) we calculated relative (per-person) increases between tasks; in fact, we find that mean HR increased from low stress to high stress tasks (Figure 1b).

3.2. Prediction Experiments

Machine learning allows for building systems that incorporate multivariate dependencies which are not obvious in statistical observation. In this section, we analyze the performance of different feature categories for predicting tasks of varying stress (Table 3) and for predicting ASD diagnosis (Table 4). In addition to session-level features, we introduce relative features (as in Figure 1), which measure intra-personal changes between tasks. We compute relative changes between five features (log-f0, intensity, HF500, speaking rate, and HR) for all three task comparisons (medium-low, high-medium, and high-low).

We initially examine the predictive power of acoustic and heart rate features across diagnostic groups for task-stress as shown in Table 3. We report both UAR and Spearman's rank-correlation coefficient, given that the task stress-labels are ordinal. The acoustic features are significantly predictive of ASD severity within both ASD and TD populations ($p<0.05$). This is an intuitive finding, given the theoretical underpinnings and empirical evidence for the relation between the acoustic features and stress/anxiety/arousal. Interestingly, heart rate level alone is only predictive of stress level for the ASD subjects. As stated previously, dependence of resting heart rate on external factors may overcome the influence of certain acute stressors; thus, the relative HR features are most appropriate and useful. Feature fusion generally leads to nominal increases in performance.

Table 3: *Classification of task stress level from acoustic-prosodic and HR features. Results are presented in terms of UAR (baseline=33%) and Spearman's rank-correlation coefficient.* **Bolded** *statistics are significant at the* $\alpha=0.05$ *level.*

	Features					
	Acoustic		*Heart Rate*		*Combined*	
Group						
ASD	**56%**	**0.59**	**54%**	**0.45**	**62%**	**0.70**
TD	**52%**	**0.57**	34%	0.10	**55%**	**0.60**
All	**70%**	**0.72**	41%	0.17	**67%**	**0.69**

Table 4: *Classification of ASD diagnosis from acoustic-prosodic and HR features in different stressful interactions. Results are presented in terms of UAR (baseline=50%).* **Bolded** *statistics are significant at the* $\alpha=0.05$ *level.* "Session" refers to an individual task, while "relative" refers to comparisons between low/medium, medium/high, and low/high, respectively.

	Features			
	Acoustic		*Heart rate*	*Combined*
Task	*Session*	*Relative*	*Session*	*All*
Low	**70**	64 (M-L)	54	**65**
Medium	**73**	**73 (H-M)**	56	**77**
High	**70**	**87 (H-L)**	57	**84**
All	**69**	**75 (All)**	61	**73**

Next we consider classification of ASD diagnosis with behavioral features as a function of stressful interaction type. We hypothesized that there may be differences in the ways in which individuals on the spectrum experience and express stress in comparison to their typically developing peers. However, it is unclear if such a difference exists. We do observe that for the combined feature set classification performance is higher for more stressful tasks (65%, 75%, and 84%, respectively); but this appears driven primarily by the relative-change acoustic features (i.e., 64%, 73%, and 87%, respectively). Thus, further investigation is required to ascertain the degree to which stress changes modulate vocal behavior in ASD. Our physiological measure, task-level mean heart rate, did not achieve significant prediction, nor did the relative heart rate features of Table 3 (not shown due to space constraints).

The most informative features are certainly the relative acoustic features, wherein vocal changes between low and high stress tasks achieve a predictive performance of 87% UAR. Session-specific features achieve a lower performance, although one that is consistent across tasks. This highlights an important concept, that each person has their own baseline, and the way in which behavior deviates from that baseline is quite informative.

4. Conclusion

In this work, we examined vocal and physiological measures of stress during social interactions designed to induce varying levels of anxiety in individuals with autism spectrum disorder and typically developing peers. Certain findings corroborate previous reports regarding acoustic-prosodic markers of autistic speech, including increased prosodic variability (pitch, intensity, and speaking rate) and more negative pitch slope (a possible correlate of perceived "monotone" speech in ASD). Furthermore, measurable differences in behavioral features were demonstrated through classification experiments in which those features could identify the corresponding stressful task as well as diagnosis. It is compelling that intra-personal acoustic deviation between low and high stress tasks was quite informative of ASD diagnosis. Still, further investigation is needed to better understand the covariation of covert and overt behavioral cues, acute stress, and autism spectrum disorder.

5. Acknowledgments

This work was supported by funds from the National Science Foundation and the National Institute of Health. We thank the children and families who generously gave their time.

6. References

[1] N. Schneiderman, G. Ironson, and S. D. Siegel, "Stress and health: psychological, behavioral, and biological determinants," *Annu. Rev. Clin. Psychol.*, vol. 1, pp. 607–628, 2005.

[2] R. C. Kessler, W. T. Chiu, O. Demler, and E. E. Walters, "Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication," *Archives of general psychiatry*, vol. 62, no. 6, pp. 617–627, 2005.

[3] T. Steimer, "The biology of fear-and anxiety-related behaviors," *Dialogues in clinical neuroscience*, vol. 4, pp. 231–250, 2002.

[4] K. R. Scherer, "Vocal affect expression: a review and a model for future research." *Psychological bulletin*, vol. 99, no. 2, 1986.

[5] S. Narayanan and P. G. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE*, vol. PP, no. 99, pp. 1–31, 2013.

[6] F. J. van Steensel, S. M. Bögels, and S. Perrin, "Anxiety disorders in children and adolescents with autistic spectrum disorders: a meta-analysis," *Clinical child and family psychology review*, vol. 14, no. 3, p. 302, 2011.

[7] J. Baio, "Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, united states, 2010." *Morbidity and mortality weekly report. Surveillance summaries*, vol. 63, no. 2, p. 1, 2014.

[8] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders, (DSM-5®)*. American Psychiatric Pub, 2013.

[9] J. Mertens, E. Zane, K. Neumeyer, and R. Grossman, "How anxious do you think i am? relationship between state and trait anxiety in children with and without asd during social tasks," *Journal of Autism and Developmental Disorders*, pp. 1–12, 2017.

[10] J. H. Hansen, S. E. Bou-Ghazale, R. Sarikaya, and B. Pellom, "Getting started with susas: a speech under simulated and actual stress database." in *Eurospeech*, vol. 97, no. 4, 1997, pp. 1743–46.

[11] C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, "Emotion recognition using a hierarchical binary decision tree approach," *Speech Comm.*, vol. 53, no. 9, pp. 1162–1171, 2011.

[12] B. Schuller, B. Vlasenko, F. Eyben, M. Wollmer, A. Stuhlsatz, A. Wendemuth, and G. Rigoll, "Cross-corpus acoustic emotion recognition: Variances and strategies," *IEEE Transactions on Affective Computing*, vol. 1, no. 2, pp. 119–131, 2010.

[13] D. Bone, C.-C. Lee, and S. Narayanan, "Robust unsupervised arousal rating: A rule-based framework with knowledge-inspired vocal features," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 201–213, 2014.

[14] J. Deng, Z. Zhang, F. Eyben, and B. Schuller, "Autoencoder-based unsupervised domain adaptation for speech emotion recognition," *IEEE Signal Proc. Letters*, vol. 21, no. 9, pp. 1068–1072, 2014.

[15] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 2004, pp. 205–211.

[16] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic, "Avec 2016: Depression, mood, and emotion recognition workshop and challenge," in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2016, pp. 3–10.

[17] P. Juslin and K. Scherer, *The New Handbook of Methods in Nonverbal Behavior Research*. Oxford: Oxford University Press., 2005, ch. 3. Vocal Expression of Affect, pp. 65–135.

[18] 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." in *Proceedings of Interspeech*, 2012, pp. 1043–1046.

[19] ——, "The Psychologist as an Interlocutor in Autism Spectrum Disorder Assessment: Insights from a Study of Spontaneous Prosody," *Journal of Speech, Language, and Hearing Research*, vol. 57, pp. 1162–1177, 2014.

[20] D. Bone, C.-C. Lee, T. Chaspari, M. Black, M. Williams, S. Lee, P. Levitt, and S. Narayanan, "Acoustic-prosodic, turn-taking, and language cues in child-psychologist interactions for varying social demand," in *Proceedings of Interspeech*, 2013.

[21] D. Bone, M. P. Black, A. Ramakrishna, R. Grossman, and S. Narayanan, "Acoustic-prosodic correlates of 'awkward' prosody in story retellings from adolescents with autism," in *Proceedings of Interspeech*, 2015.

[22] D. Bone, S. Bishop, R. Gupta, S. Lee, and S. Narayanan, "Acoustic-prosodic and turn-taking features in interactions with children with neurodevelopmental disorders," *Proceedings of Interspeech*, pp. 1185–1189, 2016.

[23] M. P. Black, D. Bone, M. E. Williams, P. Gorrindo, P. Levitt, and S. S. Narayanan, "The USC CARE Corpus: Child-Psychologist Interactions of Children with Autism Spectrum Disorders," in *Proceedings of Interspeech*, 2011.

[24] J. J. Diehl, D. Watson, L. Bennetto, J. McDonough, and C. Gunlogson, "An Acoustic Analysis of Prosody in High-Functioning Autism," *Applied Psycholinguistics*, vol. 30, pp. 385–404, 2009.

[25] R. B. Grossman, L. R. Edelson, and H. Tager-Flusberg, "Emotional facial and vocal expressions during story retelling by children and adolescents with high-functioning autism," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 3, pp. 1035–1044, 2013.

[26] C. Baltaxe, J. Q. Simmons, and E. Zee, "Intonation patterns in normal, autistic and aphasic children," in *Proceedings of the Tenth International Congress of Phonetic Sciences*. Foris Publications Dordrecht, The Netherlands, 1984, pp. 713–718.

[27] C. Schubert, M. Lambertz, R. Nelesen, W. Bardwell, J.-B. Choi, and J. Dimsdale, "Effects of stress on heart rate complexity?a comparison between short-term and chronic stress," *Biological psychology*, vol. 80, no. 3, pp. 325–332, 2009.

[28] C. Kirschbaum, K.-M. Pirke, and D. H. Hellhammer, "The 'trier social stress test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993.

[29] N. Morgan and E. Fosler-Lussier, "Combining multiple estimators of speaking rate," in *ICASSP*, vol. 2. IEEE, 1998, pp. 729–732.

[30] D. Wang and S. S. Narayanan, "Robust speech rate estimation for spontaneous speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 2190–2201, 2007.

[31] N. H. De Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior research methods*, vol. 41, no. 2, pp. 385–390, 2009.

[32] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[33] C. De Looze and D. Hirst, "The ome (octave-median) scale: A natural scale for speech prosody," in *Proceedings of the 7th International Conference on Speech Prosody (SP7)*, 2014.

[34] T. F. of the European Society of Cardiology *et al.*, "Heart rate variability standards of measurement, physiological interpretation, and clinical use," *Eur heart J*, vol. 17, pp. 354–381, 1996.

[35] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[36] L. D. Shriberg, R. Paul, J. L. McSweeny, A. Klin, D. J. Cohen, and F. R. Volkmar, "Speech and Prosody Characteristics of Adolescents and Adults with High-Functioning Autism and Asperger Syndrome," *JSLHR*, vol. 44, pp. 1097–1115, 2001.