Linguistic and Acoustic Features for Automatic Identification of Autism Spectrum Disorders in Children's Narrative

Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, Satoshi Nakamura
Graduate School of Information Science, Nara Institute of Science and Technology
{hiroki-tan, ssakti, neubig, tomoki, s-nakamura}@is.naist.jp

Abstract

Autism spectrum disorders are developmental disorders characterised as deficits in social and communication skills, and they affect both verbal and non-verbal communication. Previous works measured differences in children with and without autism spectrum disorders in terms of linguistic and acoustic features, although they do not mention automatic identification using integration of these features. In this paper, we perform an exploratory study of several language and speech features of both single utterances and full narratives. We find that there are characteristic differences between children with autism spectrum disorders and typical development with respect to word categories, prosody, and voice quality, and that these differences can be used in automatic classifiers. We also examine the differences between American and Japanese children and find significant differences with regards to pauses before new turns and linguistic cues.

1 Introduction

Autism spectrum disorders (ASD) are developmental disorders, first described by Kanner and Asperger in 1943 and 1944 respectively (Kanner, 1943; Asperger, 1944). The American Psychiatric Association defines the two characteristics of ASD as: 1) persistent deficits in social communication and social interaction across multiple contexts, and 2) restricted, repetitive patterns of behavior, interests, or activities (American Psychiatric Association, 2013). In particular, the former deficits in social communication are viewed as the most central characteristic of ASD. Thus, quantifying the degree of social communication skills is a necessary component of understanding the nature of ASD, creating systems for automatic ASD screening, and early intervention methods such as social skills training and applied behaviour analysis (Wallace et al., 1980; Lovaas et al., 1973).

There are a number of studies finding differences between people with ASD and people with typical development (TD). In terms of deficits in social communication, there have been reports describing atypical usage of gestures (Ashley and Inge-Marie, 2010), frequency of eye-contact and laughter (Geraldine et al., 1990), prosody (McCann and Peppe, 2003; Rhea et al., 2005), voice quality (Asgari et al., 2013), delay responses (Heeman et al., 2010), and unexpected words (Rouhizadeh et al., 2013). In this paper, we particularly focus on the cues of ASD that appear in children's language and speech

In the case of language, Newton et al. (2009) analyze blogs of people with ASD and TD, and found that people with ASD have larger variation of usage of words describing social processes, although there are no significant differences in other word categories. In the case of speech, people with ASD tend to have prosody that differs from that of their peers (Kanner, 1943), although McCann and Peppe (2003) note that prosody in ASD is an under-researched area and that where research has been undertaken, findings often conflict. Since then, there have been various studies analyzing and modeling prosody in people with ASD (Daniel et al., 2012; Kiss et al., 2013; Santen et al., 2013; Van et al., 2010). For example, Kiss et al. (2012) find several significant differences in the pitch characteristics of ASD, and report that automatic classification utilizing these features achieves accuracy well above chance level. To our knowledge, there is no previous work integrating both language and speech features to identify differences between people with ASD and TD. However, it has been noted that differences in person-

ality traits including introversion/extroversion can be identified using these features (Mairesse et al., 2007).

In this paper, we perform a comprehensive analysis of language and speech features mentioned in previous works, as well as novel features specific to this work. In addition, while previous works analyzed differences between people with ASD and TD, we additionally investigate whether it is possible to automatically distinguish between children with ASD or TD using both language and speech features and a number of classification methods. We focus on narratives, where the children serving as our subjects tell a memorable story to their parent (Davis et al., 2004). Here, the use of narrative allows us to consider not only single-sentence features, but also features considering interaction aspects between the child and parent such as pauses before new turns and overall narrative-specific features such as words per minute and usage of unexpected words. Given this setting, we perform a pilot study examining differences between children with ASD and TD, the possibilities of automatic classification between ASD and TD, and the differences between American and Japanese children.

2 Data Description

As a target for our analysis, we first collected a data set of interactions between Japanese children and their parents. In collecting the data, we followed the procedure used in the creation of the USC Rachel corpus (Mower et al., 2011). The data consists of four sessions: doh (free play), jenga (a game), narrative, and natural conversation. The first child-parent interaction is free play with the parent. The child and parent are given play doh, Mr. Potato Head, and blocks. The second child-parent interaction is a jenga game. Jenga is a game in which the participants must remove blocks, one at a time, from a tower. The game ends when the tower falls. The third child-parent interaction is a narrative task. The child and parent are asked to explain stories in which they experienced a memorable emotion. The final child-parent interaction is a natural conversation without a task. These child-parent interactions are recorded and will enable comparison of the child's interaction style and communication with their parent. Each session continues for 10 minutes. During interaction, a pin microphone and video camera record the speech and video of the child and the parent.

In this paper, we use narrative data of four children with ASD (male: 3, female: 1) and two children with TD (male: 1, female: 1) as an exploratory study. The intelligence quotient (IQ) for all subjects is above 70, which is often used as a threshold for diagnosis of intellectual disability. Each subject's age and diagnosis as ASD/TD is provided in Table 1. In the narrative session, each child and parent speaks "a memorable story" for 5 minutes in turn, and the listener responds to the speaker's story by asking questions. After 5 minutes, the experimenter provides directions to change the turn.

Table 1: Subjects' age and diagnosis

Subject	A1	A2	A3	A4	T1	T2
Age	10	10	10	13	10	12
Diagnosis	ASD	ASD	ASD	ASD	TD	TD

In this paper, we analyze the child-speaking turn of the narrative session in which the parent responds to the child's utterances. All utterances are transcribed based on USC Rachel corpus manual (Mower et al., 2011) to facilitate comparison with this existing corpus. In the transcription manual, if the speaker pauses for more than one second, the speech is transcribed as separate utterances. In this paper, we examine two segment levels, the first treating each speech segment independently, and the second handling a whole narrative as the target. When handling each segment independently, we use a total of 116 utterances for both children with ASD and TD.

3 Single Utterance Level

In this section, we describe language and speech features and analysis of these characteristics towards automatic classification of utterances based on whether they were spoken by children with ASD or TD. We hypothesize that based on the features extracted from the speech signal we are capable to classify children with ASD and TD on a speech segment level, as well as on narrative level after temporally combining all the segment-based decisions.

3.1 Feature Extraction

We extract language and speech features based on those proposed by (Mairesse et al., 2007) and

(Hanson, 1995). Extracted features are summarized in Table 2. We also add one feature not covered in previous work counting the number of occurrences of laughter.

Table 2: Description of language and speech features.

Language	Features
General descriptor	Words per sentence (WPS) Words with more than 6 letters Occurrences of laughter
Sentence structure	Percentage of pronouns, conjunctions, negations, quantifiers, numbers
Psychological proc.	Percentage of words describing social, affect, cognitive, perceptual, and biological
Personal concerns	Percentage of words describing work, achievement, leisure, and home
Paralinguistic	Percentage of assent, disfluencies, and fillers
Speech	**Features**
Pitch	Statistics of sd and cov
Intensity	Statistics of sd and cov
Speech rate	Words per voiced second
Voice quality	Amplitude of a3 Difference of the h1 and the h2 Difference of the h1 and the a3

3.1.1 Language Features

We use the linguistic inquiry and word count (LIWC) (Pennebaker et al., 2007), which is a tool to categorize words, to extract language features. Because a Japanese version of LIWC is not available and there is no existing similar resource for Japanese, we implement the following procedures to automatically establish correspondences between LIWC categories and transcribed Japanese utterances. First, we use Mecab[1] for part-of-speech tagging in Japanese utterances, translate each word into English using the WWWJDIC[2] dictionary, and finally determine the LIWC category corresponding to the English word. Among the language features described in Table 2, we calculate sentence structures, psychological processes, and personal concerns using LIWC, and other features using Mecab. Here, we do not consider language-dependent features and subcategories of LIWC.

3.1.2 Speech Features

For speech feature extraction, we use the Snack sound toolkit[3]. Here, we consider fundamental frequency, power, and voice quality, which are effective features according to previous works (McCann and Peppe, 2003; Hanson, 1995). We do not extract mean values of fundamental frequency and power because those features are strongly related to individuality. Thus, we extract statistics of standard deviation (fsd, psd) and coefficient of variation (fcov, pcov) for fundamental frequency and power. We calculate speech rate, which is a feature dividing the number of words by the number of voiced seconds. Voice quality is also computed using: the amplitude of the third formant (a3), the difference between the first harmonic and the second harmonic (h1h2), and the difference between the first harmonic and the third formant (h1a3) (Hanson, 1995).

3.1.3 Projection Normalization

For normalization, we simply project all feature values to a range of [0, 1], where 0 corresponds to the smallest observed value and 1 to the largest observed value across all utterances. For utterance i, we define the value of the jth feature as v_{ij} and define $p_{ij} = \frac{v_{ij} - min_j}{max_j - min_j}$, where p_{ij} is the feature value after normalisation.

3.2 Characteristics of Language and Speech Features

In this section, we report the result of a t-test, principal component analysis, factor analysis, and decision tree using the normalised features. We use R[4] for statistical analysis.

Table 3 shows whether utterances of children with ASD or TD have a greater mean on the corresponding feature. The results indicate that the children with ASD more frequently use words with more than 6 letters (e.g. complicated words), assent (e.g. "uh-huh," or "un" in Japanese), and fillers (e.g. "umm," or "eh" in Japanese) significantly more than the children with TD. In contrast, the children with TD more frequently use the words words categorized as social (e.g. friend), affect (e.g. enjoy), and cognitive (e.g. understand) significantly more than the children with ASD. In addition, there are differences in terms of fundamental frequency variations and voice quality (e.g.

Table 3: Difference of mean values between ASD and TD based on language and speech features from children's utterances. Each table cell notes which of the two classes has the greater mean on the corresponding feature (*: p < 0.01, **: p < 0.005).

WPS	6 let.	laughter	adverb	pronoun	conjunctions	negations	quantifiers	numbers	social
-	ASD*	-	-	-	-	-	-	-	TD**
affect	**cognitive**	**perceptual**	**biological**	**relativity**	**work**	**achievement**	**leisure**	**home**	**assent**
TD**	TD*	-	-	-	-	-	-	-	ASD**
nonfluent	**fillers**	**fsd**	**fcov**	**psd**	**pcov**	**speech rate**	**a3**	**h1h2**	**h1a3**
-	ASD*	TD**	TD*	-	-	-	-	-	ASD**

h1a3). In particular, we observe that the children with ASD tend to use monotonous intonation as reported in (Kanner, 1943). We do not confirm a significant differences in other features.

Next, we use principal component analysis and factor analysis to find features that have a large contribution based on large variance values. As a result of principal component analysis, features about fundamental frequency, power, and h1a3 have large variance in the first component, and the feature counting perceptual words also has large value in the second component. To analyze a different aspect of principal component analysis with rotated axes, we use factor analysis with the varimax rotation method. Figure 1 shows the result of factor analysis indicating that features regarding fundamental frequency and power have large variance. In addition, other features such as speech rate, a3, and h1a3 also have large variance. Here, we can see that for features such as statistics of fundamental frequency (fsd and fcov) and power (psd and pcov), the correlation coefficient between these features are over 80% (p < 0.01). For correlated features, we use only standard deviation in the following sections.

We also analyze important features to distinguish between children with ASD and TD by using a decision tree. Figure 2 shows the result of a decision tree with 10 leaves indicating that speech features fill almost all of the leaves (e.g. fsd is a most useful feature to distinguish between ASD and TD). In terms of the language features, we confirm that WPS and perceptual words are important for classification.

3.3 Classification

In this section, we examine the possibility of automatic identification of whether an utterance belongs to a speaker with ASD or TD. Based on the previous analysis, we prepare the following
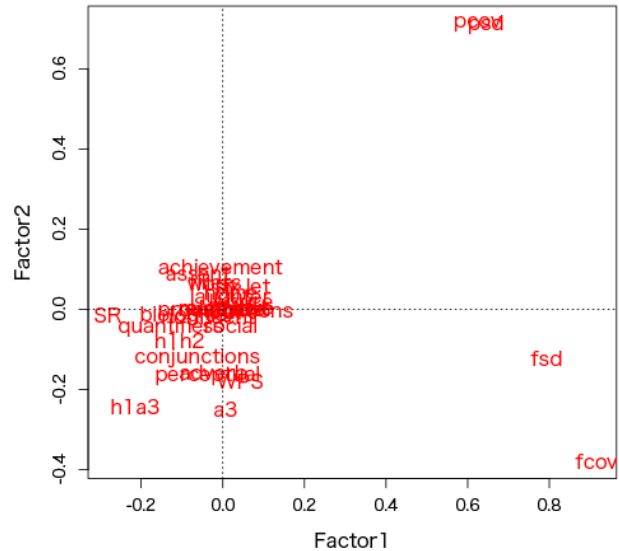

Figure 1: Factor analysis with varimax rotation method. First and second factors are indicated.

feature sets: 1) language features (Language), 2) speech features (Speech), 3) all features (All), 4) important features according to the t-test, principal component analysis, factor analysis, and decision tree (Selected), 5) important features according to the t-test that are not highly correlated (T-Uncor). The feature set of T-Uncor is as follows: 6 let., social, affect, cognitive, fillers, assent, fed, and h1a3. We also show the chance rate, which is a baseline of 50% because the number of utterances in each group is the same, and measure accuracy with 10-fold cross-validation and leave-one-speaker-out cross-validation using naive Bayes (NB) and support vector machines with a linear kernel (SVM). In the case of leave-one-speaker-out cross-validation, we use T-Uncor because the number of utterances without one speaker is too small to train using high dimensional feature sets.

Table 4 shows the result indicating that accu-

racies with almost all feature sets and classifiers are over 65%. The SVM with Selected achieves the best performance for the task of 10-fold cross-validation, and The SVM with T-Uncor achieves 66.7% for the task of leave-one-speaker-out. The accuracy for the task of leave-one-speaker-out on each speaker A1 to T2 is as follows: 78%, 60%, 53%, 51%, 82%, and 78%.

Table 4: Accuracy using Naive Bayes and SVM classifiers. The p-value of the t-test is measured compared to baseline (chance rate) (†: $p < 0.1$, *: $p < 0.01$)

Feature set	Accuracy [%]		
	Baseline	NB	SVM
Language		62.2†	70.3*
Speech		57.6	67.6*
All	50.0	65.0†	68.8*
Selected		67.4*	71.9*
T-Uncor		67.8†	68.1†
Per-Speaker	50.0	65.5†	66.7†

4 Narrative Level

In this section, we focus on the features of entire narratives, which allows us to examine other features of child-parent interaction for a better understanding of ASD and classification in children with ASD and TD. Each following subsection describes the procedure of feature extraction and analysis of characteristics at the narrative level. We consider pauses before new turns and unexpected words, which are mentioned in previous works, as well as words per minute.

4.1 Pauses Before New Turns

Heeman et al., (2010) reported that children with ASD tend to delay responses to their parent more than children with TD in natural conversation. In this paper, we examine whether a similar result is found in interactive narrative. We denote values of pauses before new turns as time between the end of the parent's utterance and the start of the child's utterance. We do not consider overlap of utterances. We test goodness of fit of pauses to a gamma and an exponential distribution based on (Theodora et al., 2013), because the later is a special case of gamma with a unity shape parameter, using the Kolmogorov-Smirnov test.

Figure 3 shows a fitting of pauses to gamma or exponential distributions, and we select a bet-
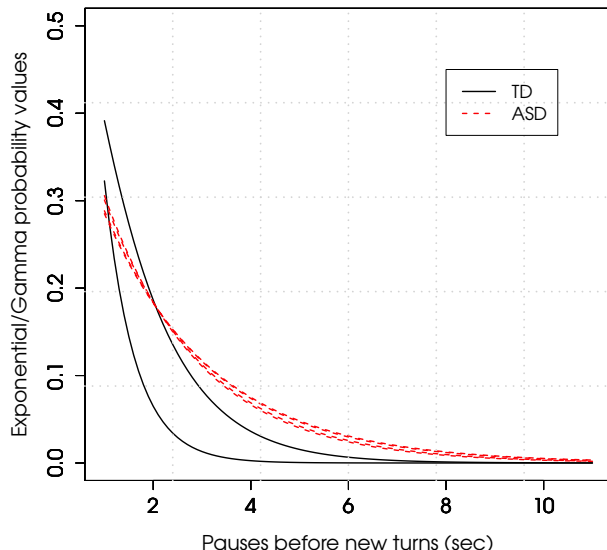

Figure 3: Gamma/Exponential pause distributions with parameters computed using Maximum Likelihood Estimation (MLE) for children with ASD and TD.

ter fitted distribution. All subjects significantly fit ($p > 0.6$). As shown in Figure 3, we confirm that children with ASD tend to delay responses to their parent compared with children with TD. To reflect this information in our following experiments in automatic identification of ASD in narrative, we extract the expectation value of the exponential distribution

Heeman et al., (2010) also reported the relationship of the parent's previous utterance's type (question or non-question) and the child's pauses. We examine the relationship between the parent's previous question's type and pauses before new turns. For each of the children's utterances, we label the parent's utterance that directly precedes as either "open question," "closed question," or "non-question", and we calculate pause latency. Closed-questions are those which can be answered by a simple "yes" or "no," while open-questions are those which require more thought and more than a simple one-word answer. As shown in Table 5, children with ASD tend to delay responses to their parent to a greater extent than children with TD. We found no difference between open and closed questions, although a difference between questions and non-questions is observed. These results are consistent with those of previous work (Heeman et al., 2010) in terms of differences between questions and non-questions.
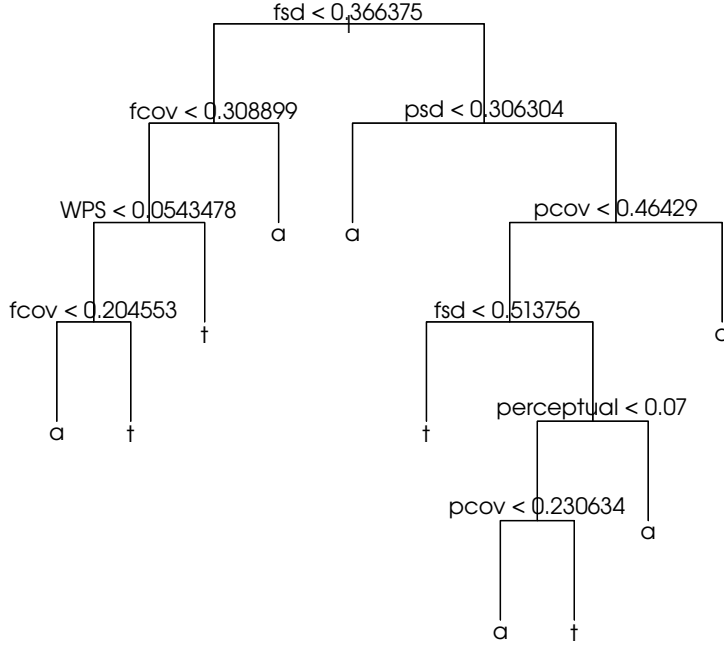
Figure 2: Decision tree with 10 leaves (a: ASD, t: TD).

Table 5: Relationship of pauses before new turns and parents' question types. The mean value and standard deviation are shown.

Question type	TD	ASD
Closed-question	0.47 (0.46)	1.61 (1.87)
Open-question	0.43 (0.34)	1.76 (1.51)
Non-question	0.95 (1.18)	2.60 (3.64)

Table 6: Mean value of words per minute.

Subj.	Averaged WPM
A1	18.25
A2	86.75
A3	23.75
A4	115.5
T1	99.25
T2	103.5

4.2 Words Per Minute

We analyze words per minute (WPM) in children with ASD and TD to clarify the relationship between ASD and frequency of speech. We use a total of 5 minutes of data in each narrative, and thus the total number of words are divided by 5 to calculate WPM. Table 6 shows the result. The data in this table indicates that some children with ASD have a significantly lower speaking rate than others with TD, but it is not necessarily the case that ASD will result in a low speaking rate such as the case of Asperger's syndrome (Asperger, 1944).

4.3 Unexpected Words

Characteristics of ASD include deficits in social communication, and these deficits affect inappro-

priate usage of words (Rouhizadeh et al., 2013). We evaluate these unexpected words using two measures, term frequency-inverse document frequency (TF-IDF) and log odds ratio. We use the following formulation to calculate TF-IDF for each child's narrative i and each word in that narrative j, where c_{ij} is the count of word j in narrative i. f_j is the number of narratives from the full data of child narratives containing that word j, and D is the total number of narratives (Rouhizadeh et al., 2013).

$$tf - idf_{ij} = (1 + \log c_{ij}) \log \frac{D}{f_j}$$

The log odds ratio, another measure used in in-

formation retrieval and extraction tasks, is the ratio between the odds of a particular word, j, appearing in a child's narrative, i. Letting the probability of a word appearing in a narrative be p_1 and the probability of that word appearing in all other narratives be p_2, we can express the odds ratio as follows:

$$\text{odds ratio} = \frac{\text{odds}(p_1)}{\text{odds}(p_2)} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

A large TF-IDF and log odds score indicates that the word j is very specific to the narrative i, which in turn suggests that the word might be unexpected or inappropriate. In addition, because the overall amount of data included in the narratives is too small to robustly analyze these statistics for all words, we also check for the presence of each word in Japanese WordNet[5] and determine that if it exists in WordNet it is likely a common (expected) word. Table 7 shows the result of TF-IDF, log odds ratio, and their summation, and we confirm that there is no difference between children with ASD and TD. This result is different from that of previous work (Rouhizadeh et al., 2013). The children in that study were all telling the same story, and one possible explanation for this is due to the fact that in this work we do not use language-constricted data such as narrative retelling, and thus differences due to individuality are more prevalent.

Table 7: TF-IDF, log odds ratio, and their summation.

Subj.	TF-IDF	Log-odds	T+L
A1	0.50	1.01	1.52
A2	0.58	0.49	1.08
A3	0.66	1.23	1.89
A4	0.66	0.31	0.96
T1	0.74	0.49	1.23
T2	0.62	0.44	1.06

4.4 Classification

In this section, we examine the possibility of automatic classification of whether an interactive narrative belongs to children with ASD or TD. Because of the total number of subjects is small (n=4 for ASD, n=2 for TD), we perform classification

with a K-NN classifier with K=1 nearest neighbour. As features, we compute the features mentioned in Section 3.1, and use the average over all utterances as the features for the entire narrative. Finally, we use pauses before new turns (expectation value of the exponential distribution), WPM, TF-IDF, log odds ratio, 6 let., social, affect, cognitive, assent, fillers, fsd, h1a3, and calculate accuracy with leave-one-speaker-out cross-validation.

As a result, we achieved an accuracy of 100% in classification between ASD and TD on the full-narrative level, which shows that these features are effective to some extent to distinguish children with ASD and TD. However, with only a total of 6 children, our sample size is somewhat small, and thus experiments with a larger data set will be necessary to draw more firm conclusions.

5 Data Comparison

As all our preceding experiments have been performed on data for Japanese child-parent pairs, it is also of interest to compare these results with data of children and parents from other cultures. In particular, we refer to the USC Rachel corpus (Mower et al., 2011) (the subjects are nine children with ASD) for comparison. Using the USC Rachel corpus, there is a report mentioning the relationship of parent's and child's linguistic information and pauses before new turns (Theodora et al., 2013). In this paper, we follow this work using Japanese data. The USC Rachel corpus includes a session of child-parent interaction, and the same transcription standard is used. We extract pauses before new turns, and short and long pauses are differentiated based on the 70th percentile of latency values for each child individually. We investigate the relationship between the parent and child's language information based on features used in Section 3.1, and short and long pauses.

Table 8 and 9 show significantly greater mean values performed using bootstrap significance testing on the means of the two pause types. By observing the values in the table, we can see that the trends are similar for both American and Japanese children. However, in terms of WPS, there is a difference. The American ASD children have greater means for WPS in the case of long pauses, while Japanese children have greater means for WPS in the case of short pauses. We analyze these differences in detail.

Table 8: In the case of USC Rachel corpus, bootstrap on difference of means between short (S) and long (L) pauses based on linguistic features from child's and parent's utterances (†: p < 0.1, *: p < 0.01). Each table cell notes which of the two types of pauses has greater mean on the corresponding feature.

Subj.	Child				Parent		
	WPS	conj.	affect	nonflu.	adverb	cogn.	percept.
S1	L*	L*	S*	-	L*	L*	L*
S2	L*	L*	S†	L*	L*	L*	L*
S3	L*	L†	-	S†	L*	L*	L*
S4	-	-	-	L*	L*	L*	L*
S5	L†	-	-	-	L*	L*	L*
S6	L*	-	S*	-	L*	L*	-
S7	L†	-	S†	-	L†	-	-
S8	L*	-	-	-	L*	L*	L*
S9	-	-	-	S†	L*	L*	L*

Table 9: Bootstrap for pause differences in the Japanese corpus.

Subj.	Child				Parent		
	WPS	conj.	affect	nonflu.	adverb	cogn.	percept.
A1	S*	-	-	-	S*	L*	-
A2	S†	-	S*	-	L*	L*	L*
A3	S†	-	-	-	L*	L*	L*
A4	S*	-	-	-	-	-	-

In the Japanese corpus, we observe that WPS is larger in the case of short pauses. As we noticed that the child often utters only a single word for responses that follow a long pause, we analyzed the content of these single word utterances. As shown in Figure 4, for example, A1 tends to use a word related to assent when latency is long, and A4 tends to use a word related to filler, assent or others when latency is long. Though there are individual differences, we confirm that the Japanese children with ASD examined in this study tend to delay their responses before uttering one word. These characteristics may be related to the parent's question types and the child's cognitive process, and thus we need to examine these possibilities in detail.

6 Conclusion

In this work, we focused on differentiation of children with ASD and TD in terms of social communication, particularly focusing on language and speech features. Using narrative data, we examined several features on both the single utterance
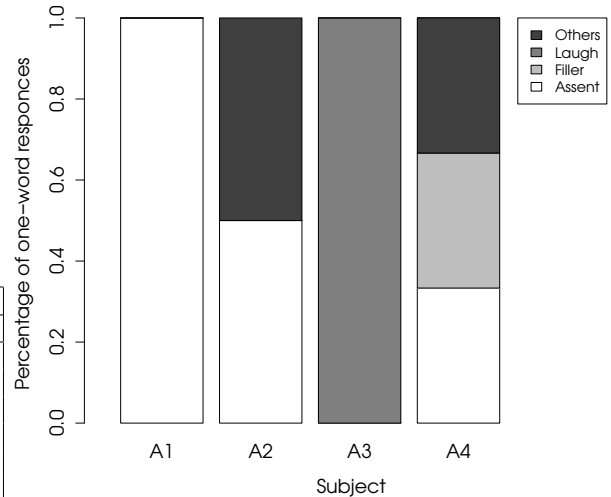

Figure 4: The language category of one-word responses in the case of a long pause.

level and the narrative level. We examined features mentioned in a number of previous works, as well as a few novel features. We confirmed about 70% accuracy in an evaluation over single utterances, and some narrative features also proved to have a correlation with ASD.

For future directions, we plan to perform larger scale experiments to examine the potential of these features for automated ASD screening. Given the results of this, we plan to move to applications including the development of dialogue systems for automatic ASD screening and social skills training.

Acknowledgments

References

American Psychiatric Association. 2013. *The Diagnostic and Statistical Manual of Mental Disorders: DSM 5.*

Asgari, Meysam, Alireza Bayestehtashk, and Izhak Shafran. 2013. Robust and Accurate Featuers for Detecting and Diagnosing Autism Spectrum Disorders. *Proceedings of Interspeech*, 191–194.

Asperger, H.. 1944. Die „Autistischen Psychopathen" im Kindesalter. *European Archives of Psychiatry and Clinical Neuroscience*, 117: 76–136.

Bone, D., Black, M. P., Lee, C. C., Williams, M. E., Levitt, P., Lee, S., and Narayanan, S.. 2012. Spontaneous-Speech Acoustic-Prosodic Features of Children with Autism and the Interacting Psychologist. *Proceedings of Interspeech*.

Chaspari, T., Gibson, D. B., Lee, C.-C., and Narayanan, S. S. 2013. Using physiology and language cues for modeling verbal response latencies of children with ASD. *Proceedings of ICASSP*, 3702–3706.

Davis, Megan, Kerstin Dautenhahn, CL Nehaniv, and SD Powell. 2004. Towards an Interactive System Facilitating Therapeutic Narrative Elicitation in Autism. *Proceedings of NILE*.

Dawson, Geraldine, Deborah Hill, Art Spencer, Larry Galpert, and Linda Watson.. 1990. Affective exchanges between young autistic children and their mothers. *Journal of Abnormal Child Psychology*, 18: 335–345.

de Marchena, A. and Inge-Marie E.. 2010. Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. *Autism Research*, 3: 311–322.

Hanson M. H.. 1995. Glottal characteristics of female speakers. Harvard University, Ph.D. dissertation.

Heeman, P. A., Lunsford, R., Selfridge, E., Black, L., and Van Santen, J.. 2010. Autism and interactional aspects of dialogue. *Proceedings of SIGDIAL*, 249–252.

Kanner, L.. 1943. Autistic disturbances of affective contact. *Nervous Child*, 2: 217–250.

Kiss, G. and van Santen, J. P. H.. 2013. Estimating Speaker-Specific Intonation Patterns Using the Linear Alignment Model. *Proceedings of Interspeech* 354–358.

Kiss, G., van Santen, J. P. H., Prud'hommeaux, E. T., and Black, L. M.. 2012. Quantitative Analysis of Pitch in Speech of Children with Neurodevelopmental Disorders. *Proceedings of Interspeech*.

Lovaas, O Ivar, Robert Koegel, James Q Simmons, and Judith Stevens Long. 1973. Some generalisation and follow-up measures on autistic children in behaviour therapy. *Journal of Applied Behavior Analysis*, 6: 131–166.

Mairesse, Francois, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. 2007. Using Linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30: 457–500.

McCann, J. and Sue, P.. 2003. Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders*, 38(4): 325–350.

Mower, E., Black, M. P., Flores, E., Williams, M., and Narayanan, S.. 2011. Rachel: Design of an emotionally targeted interactive agent for children with autism. *Proceedings of IEEE ICME*, 1–6.

Newton, A. T., Kramer, A. D. I., and McIntosh, D. N.. 2009. Autism online: a comparison of word usage in bloggers with and without autism spectrum disorders. *Proceedings of SIGCHI*, 463–466.

Paul, Rhea, Amy Augustyn, Ami Klin, and Fred R Volkmar. 2005. Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 35: 205–220.

Pennebaker, James W, Martha E Francis, and Roger J Booth. 2005. Linguistic inquiry and word count: LIWC [Computer software] Austin, TX: liwc. net.

Rouhizadeh Masoud, Prud'hommeaux Emily, Roark Brian, and van Santen Jan. 2013. Distributional semantic models for the evaluation of disordered language. *Proceedings of NAACL-HLT*, 709–714.

Santen, Jan PH, Richard W Sproat, and Alison Presmanes Hill. 2013. Quantifying repetitive speech in autism spectrum disorders and language impairment. *Autism Research*, 6: 372–383.

Sharda, Megha, T Padma Subhadra, Sanchita Sahay, Chetan Nagaraja, Latika Singh, Ramesh Mishra, Amit Sen, Nidhi Singhal, Donna Erickson, and Nandini C Singh. 2010. Sounds of melody—Pitch patterns of speech in autism. *Neuroscience letters*, 478: 42–45.

Van Santen, Jan PH, Emily T Prud'hommeaux, Lois M Black, and Margaret Mitchell. 2010. Computational prosodic markers for autism. *Autism*, 14: 215–236.

Wallace, Charles J, Connie J Nelson, Robert Paul Liberman, Robert A Aitchison, David Lukoff, John P Elder, and Chris Ferris. 1980. A review and critique of social skills training with schizophrenic patients. *Schizophrenia Bulletin*, 6:42–63.