Analyzing Autism Speech of Children in English Vowels Regions by Analysis of Changes in Production Features

Abhijit Mohanta[1] **and Vinay Kumar Mittal**[2]
[1]Indian Institute of Information Technology, Sri City, Chittoor, Andhra Pradesh, India
[2]CEO, Ritwik Software Technologies Pvt. Ltd, Hyderabad, Telangana, India
abhijit.mohanta@iiits.in[1] and DrVinayKrMittal@gmail.com[2]

Abstract

Children with autism spectrum disorder (ASD) have difficulty in producing the speech that is different from the speech of normal children. Most children with ASD have difficulty in proper communication of information, their thoughts or their emotional state. Only a few studies have been carried out towards acoustic analysis of ASD speech. Objective of this study is to characterize the speech signal of the children affected with autism, by examining changes in the acoustic features. An autism speech dataset has been collected from the children affected with autism, for over a year. Autism speech is examined mainly in English vowels regions due to their relatively longer duration, and quasi-periodic nature of the vocal folds during pronunciation of the vowel sounds. Changes in the characteristics of autism speech are analyzed by examining changes in the production features. The excitation source characteristics are examined using the feature F0, and the vocal tract filter, i.e., system characteristics, by using dominant frequencies features. The combined characteristics of the source-system interaction are examined using features SoE, ZCR and signal energy. Changes are examined in five English vowels regions. Distinct patterns of changes observed in the autism speech of male and female children are discussed.

keywords: ASD, ZFF, F0, FD1, FD2

1 Introduction

Autism spectrum disorder (ASD) is a pervasive developmental disorder, defined clinically by observing abnormalities in three areas: communication, social reciprocity, and reduced or hyperfocus behavioral flexibility (Chaspari et al., 2014; Kjelgaard and Tager-Flusberg, 2001). It is knows as a spectrum disorder because of its heterogeneity of symptomatology (Bone et al., 2012). According to (Black et al., 2011; Mower et al., 2011a), 1 in 110 children are with ASD. Disturbances of prosody, communication impairments and abnormalities involving speech impairments are some of the most common aspects among many individuals with ASD (Fusaroli et al., 2017). Also, individuals with autism carry some specific biomarkers connected with the disorder by birth, and these biomarkers develop in such a way that can be clearly identified by 36 months of life or later (McCann and Peppé, 2003). However, there are no fixed rules to define autism, also the biological and genetic reasons behind the ASD are still unknown (Herbert et al., 2005). Still, only a few researches have been done on *autism speech*, i.e., the speech of ASD affected children. The purpose of this research is to analyze speech signal of the children with autism in English vowels regions, by examining changes in the production features.

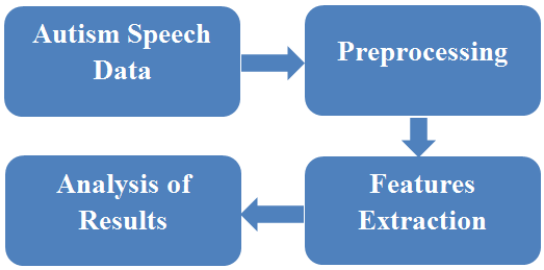

Figure 1: Block diagram of the proposed plan.

Sometimes the individuals with autism are associated with the difficulties in expressing their emotions as well as understanding others' emotional states from speech, facial expression, etc (Marchi et al., 2012). The reason behind language

Proc. of ICON-2018, Patiala, India. December 2018 ©*2018 NLPAI*, pages 145–149

Table 1: Datset details of the children with ASD.

Attributes	Statistics
Total number of the children	13 (11 male, 2 female)
Age (in years)	3.5 to 16
Native language	Tamil: 12 and Punjabi: 1
Reading skill (English)	Beginner: 2 Intermediate: 4 Fluent: 7
Total files	187
Duration of data	9350 sec

impairment in autism is the outcome of primary linguistic disorder with a focus on pragmatic impairments (Rutter, 1970; Baltaxe, 1977). In fact, in many cases, a significant spoken language delay and repetitive language could be encountered in the children with ASD (Mower et al., 2011b). In general, the children with typical development start establishing their vocabularies at the age of 24 months, but the children with autism could be unable to do the same (Short and Schopler, 1988). Also, compared to the individuals with typical speech, individuals with high-functioning autism (HFA) have a large variation in pitch, and some of them have the absence of terminal pitch contour in their speech (Shriberg et al., 2001; Diehl et al., 2009).

This paper discusses the analyzation of autism speech in the vowel regions of five English vowels. Speech dataset collection is one of the most challenging tasks in order to do the research on speech signal of the children with autism. Although, for this research purpose an English speech signal dataset has been collected by recording speech samples of the children with ASD. Only the vowels regions of English language have been considered in this study because of their relatively longer duration and the presence of sustained speech signal in vowel regions. The excitation source characteristics are examined using the feature instantaneous fundamental frequency (F0), and the system characteristics are examined using the first two dominant frequencies (FD1 and FD2). In addition, the combined characteristics are examined using the features strength of excitation (SoE), zero crossing rate (ZCR) and signal energy (E). All these features have been extracted only form the vowels parts of the speech signal.

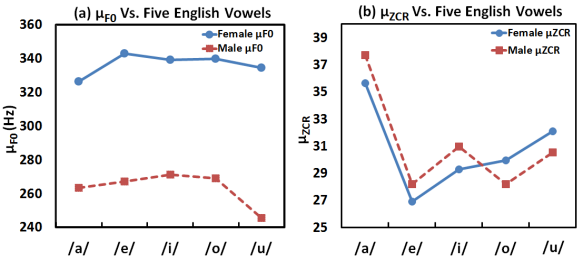

Figure 2: Average F0 and ZCR for male and female children with autism: (a) μ_{F0}, (b) μ_{ZCR}.

This study consists of four major steps, graphically represented in Figure 1. Firstly, a speech signal dataset was collected, by recording the sound files of the children with ASD. Secondly, in preprocessing step, unwanted signal parts were removed, and the speech signal files were arranged in a database. Thirdly, speech signal processing methods were applied on the collected dataset to extract the production features. Lastly, results were made by observing changes in the extracted features.

The organization of this paper is as follows. Details about the database collection of the children with ASD are discussed in section 2. Next, the signal processing methods and features used for the purpose of analysis are discussed in section 3. Section 4 presents the key results, observations and discussion on results. Lastly, section 5 presents the conclusions along with the scope of future work on this topic.

2 Speech Signal Dataset of ASD Children

A speech signal dataset of the children with ASD was recorded for this research purpose. Dataset details are tabulated in Table 1. Here, in this study, the children with age less than 41 months were not considered, because according to a study ASD start in the first 36 months of life (McCann and Peppé, 2003). Dataset was recorded in English. All the speakers who were selected for data collection had a knowledge of English, and also had some speaking related problems. It was made sure by a certified doctor that the children considered for the data collection were diagnosed with ASD. Also, before data collection it was made sure that they met the DSM–IV diagnostic criteria (Lord et al., 1994) and other diagnostic criteria for autism.

Data was collected once or twice in a week for a year period of time in a noise-free empty room.

Table 2: Mean (μ) values of the acoustic features of the male children with autism: (a) acoustic features and (b)-(f) mean values in five English vowels regions.

(a) Features	(b) /a/	(c) /e/	(d) /i/	(e) /o/	(f) /u/
F0 (Hz)	263.3	267.1	271.3	269.1	254.7
E×1000	36.53	31.41	43.18	41.25	54.29
SoE×100	34.9	43.4	44	39.1	35.9
ZCR×1000	37.7	28.18	30.98	28.21	30.55
FD1 (Hz)	1041.7	900.5	1043.4	863	951.8
FD2 (Hz)	3294.6	3234.2	3281.5	3291.4	3316.1

Table 3: Mean (μ) values of the acoustic features of the female children with autism: (a) acoustic features and (b)-(f) mean values in five English vowels regions.

(a) Features	(b) /a/	(c) /e/	(d) /i/	(e) /o/	(f) /u/
F0 (Hz)	326.4	343	339.3	339.9	334.6
E×1000	39.5	35.3	48.35	63.75	58.65
SoE×100	32	37.1	38.1	42.9	37.1
ZCR×1000	35.65	26.9	29.3	29.95	32.1
FD1 (Hz)	864.7	686.1	783.3	802.8	859.9
FD2 (Hz)	3185	3285.9	3268.8	3057.6	3112.3

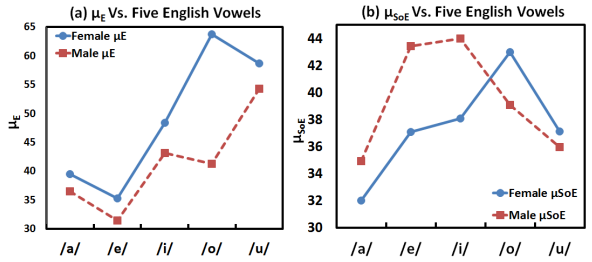

Figure 3: Average E and SoE for male and female children with autism: (a) μ_E, (b) μ_{SoE}.

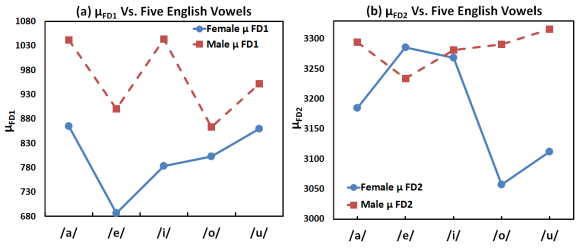

Figure 4: Average FD1 and FD2 for male and female children with autism: (a) μ_{FD1}, (b) μ_{FD2}.

The children were asked to pronounce a set of twenty five especially selected consonant-vowel-consonant (CVC) and consonant-vowel-vowel-consonant (CVVC) English words and numbers. These words and numbers were shown to them with pictures and text on a laptop. A study shows that children with autism have an early interest in letters and numbers, and that is a big reason behind showing them words and numbers (Volkmar et al., 1997; Tager-Flusberg et al., 2005).

Each child with ASD was asked to pronounce the same set of words, over the entire data collection period. There were five English words selected for each of the five English vowels. So, in a single day by each speaker, total utterances of 25 words were recorded for each of two such sections. Roland R-26 audio recorder was used for the recording purpose. In addition, data was recorded at a sampling rate of 48 KHz and in .wav format.

3 Signal Processing Methods and Features

The speech signal files of the children with ASD were analyzed by observing the changes in the source characteristics (F0), system characteristics (dominant frequencies), and combined characteristics (SoE, ZCR and signal energy). The F0 was derived using zero-frequency filtering (ZFF) method with the sampling frequency 10 KHz (Murty and Yegnanarayana, 2008; Yegnanarayana and Murty, 2009). The ZFF method involves computing the output of the cascade of two zero-frequency resonators (ZFRs) (Murty and Yegnanarayana, 2008; Yegnanarayana and Murty, 2009). The zero-frequency filter signal output of ZFR is given as:

$$y_1[n] = -\sum_{k=1}^{2} a_k y_1[n-k] + x[n] \qquad (1)$$

where, x[n] is pre-processed input signal, $a_1 = -2, and\ a_2 = 1$. This operation is repeated twice, for a cascade of ZFRs. The trend in this output is removed by subtracting the moving average corresponding to the 10 ms window at each sample. The resultant trend removed signal, called ZFF signal is given by:

$$y[n] = y_2[n] - \frac{1}{2N+1} \sum_{m=-N}^{N} y_2[n+m] \qquad (2)$$

where, 2N+1 is the window length in terms of sample number. The resultant signal is called the ZFF signal. Its positive giving zero crossings indicate the glottal closure instants (GCIs), which are used to estimate the F0 (Murty and Yegnanarayana, 2008). In addition, the slope of the ZFF signal around the GCIs gives a measure of the SoE (Murty and Yegnanarayana, 2008; Mittal and Yegnanarayana, 2015a).

The FD1 and FD2 were derived using linear prediction (LP) analysis (Makhoul, 1975). With the LP order 5, the LP spectrum will have a maximum of two peaks. The frequencies corresponding to these peaks are called the dominant frequencies, denoted as FD1 and FD2, respectively (Mittal and Yegnanarayana, 2015b).

The signal energy (E) (Rihaczek, 1968) was calculated using the frame size 30 ms and frame shift 10 ms. Signal energy of a discrete-time signal $x[n]$ can be computed as: $E_w = \sum_{n=-w/2}^{w/2} |x[n]|^2$ where, w is the window length.

4 Results and Observations

The μ_{F0}, μ_E, μ_{SoE}, μ_{ZCR}, μ_{FD1}, and μ_{FD2} values for male and female children with autism are represented in Table 2 and 3, respectively. From Figure 2(a), it is observed that for each of the five English vowels, the female children with autism have the higher vocal fold vibration rate than the male children with autism. Although, this statement is also true in case of the individual with typical speech. Also, in case of both the male and female children with autism, front vowels i.e., /e/ and /i/ give the highest μ_{F0} values as compared to other English vowels. Lastly, in case of the μ_{F0} values of the rear vowels i.e., /o/ and /u/, a similar pattern is observed for both the male and female children with autism.

The μ_{ZCR} values of the male and female speakers are represented graphically in Figure 2(b). In addition, μ_{ZCR} values are multiplied by 1000 for better understanding purpose. The μ_{ZCR} values of the front and mid vowels follow a similar trend for both the male and female speakers, which could be observed from Figure 2(b). Next, in case of front and mid vowels, the male speakers have the higher μ_{ZCR} values as compared to female, but in case of rear vowels the female speakers have the higher μ_{ZCR} values as compared to male speakers.

The μ_E values of male and female speakers are represented graphically in Figure 3(a). In addition, μ_E values are also multiplied by 1000 for better understanding purpose. Here, it is observed that the μ_E values of front and mid vowels for both the male and female speakers follow a similar trend. It could be observed from Figure 3(a).

All the μ_{SoE} values are multiplied by 100 for the purpose of better understanding. It is observed that in case of the front and mid vowels, μ_{SoE} values follow a similar trend for both the male and female speakers. Also, in case of the front and mid vowels, μ_{SoE} have the higher values for the male speakers as compared to female speakers. These statements could be observed from Figure 3(b).

The μ_{FD1} and μ_{FD2} values are represented in Figure 4(a) and 4(b), respectively. In case of all English vowels, μ_{FD1} values are higher for male children as compared to female children. Also, in case of female vowel, /e/ gives the lowest μ_{FD1} value and the highest μ_{FD2} value as compared to other English vowels. Next, in case of the female children, μ_{FD2} value of vowel /e/ is the highest as compared to other vowels. On the other hand, in case of the male children, μ_{FD2} value of vowel /e/ is the lowest as compared to other vowels.

5 Conclusions

The aim of this research is to analyze changes in the various speech production features in English vowel regions of children with ASD. An autism speech dataset is recorded for this research purpose. Changes are analyzed by observing the differences in the source characteristics (F0), system characteristics (FD1 and FD2), and combined characteristics (SoE, ZCR and E). In the conclusion, it could be stated that in case of the male and female children with ASD, front and mid vowels show the similar trend for F0, E, SoE and ZCR. But, in case of the rear vowels such trends are not present. These robust results could be used to differentiate between the children with autism and the typically developed individuals.

A small size of speech data for female children with ASD is a limitation of this study. More acoustic features could be considered for future studies.

Acknowledgments

The authors are thankful to the Doctrine Oriented Art of Symbiotic Treatment (DOAST) Integrated Therapy Centre for Autism, Anna Nagar, Chennai, India, for giving the opportunity to collect autism speech data.

References

Christiane AM Baltaxe. 1977. Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology*, 2(4):176–180.

Matthew P Black, Daniel Bone, Marian E Williams, Phillip Gorrindo, Pat Levitt, and Shrikanth Narayanan. 2011. The usc care corpus: Child-psychologist interactions of children with autism spectrum disorders. In *Twelfth Annual Conference of the International Speech Communication Association*.

Daniel Bone, Matthew P Black, Chi-Chun Lee, Marian E Williams, Pat Levitt, Sungbok Lee, and Shrikanth Narayanan. 2012. Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist. In *Thirteenth Annual Conference of the International Speech Communication Association*.

Theodora Chaspari, Matthew Goodwin, Oliver Wilder-Smith, Amanda Gulsrud, Charlotte A Mucchetti, Connie Kasari, and Shrikanth Narayanan. 2014. A non-homogeneous poisson process model of skin conductance responses integrated with observed regulatory behaviors for autism intervention. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 1611–1615. IEEE.

Joshua J Diehl, Duane Watson, Loisa Bennetto, Joyce McDonough, and Christine Gunlogson. 2009. An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, 30(3):385–404.

Riccardo Fusaroli, Anna Lambrechts, Dan Bang, Dermot M Bowler, and Sebastian B Gaigg. 2017. Is voice a marker for autism spectrum disorder? a systematic review and meta-analysis. *Autism Research*, 10(3):384–407.

Martha R Herbert et al. 2005. Autism: a brain disorder or a disorder that affects the brain. *Clinical Neuropsychiatry*, 2(6):354–379.

Margaret M Kjelgaard and Helen Tager-Flusberg. 2001. An investigation of language impairment in autism: Implications for genetic subgroups. *Language and cognitive processes*, 16(2-3):287–308.

Catherine Lord, Michael Rutter, and Ann Le Couteur. 1994. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of autism and developmental disorders*, 24(5):659–685.

John Makhoul. 1975. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580.

Erik Marchi, Björn W Schuller, Anton Batliner, Shimrit Fridenzon, Shahar Tal, and Ofer Golan. 2012. Emotion in the speech of children with autism spectrum conditions: prosody and everything else. In *WOCCI*, pages 17–24.

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

Vinay Kumar Mittal and B Yegnanarayana. 2015a. Study of characteristics of aperiodicity in noh voices. *The Journal of the Acoustical Society of America*, 137(6):3411–3421.

Vinay Kumar Mittal and Bayya Yegnanarayana. 2015b. Analysis of production characteristics of laughter. *Computer Speech & Language*, 30(1):99–115.

Emily Mower, Matthew P Black, Elisa Flores, Marian Williams, and Shrikanth Narayanan. 2011a. Rachel: Design of an emotionally targeted interactive agent for children with autism. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*, pages 1–6. IEEE.

Emily Mower, Chi-Chun Lee, James Gibson, Theodora Chaspari, Marian E Williams, and Shrikanth Narayanan. 2011b. Analyzing the nature of eca interactions in children with autism. In *Twelfth Annual Conference of the International Speech Communication Association*.

K Sri Rama Murty and B Yegnanarayana. 2008. Epoch extraction from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8):1602–1613.

A Rihaczek. 1968. Signal energy distribution in time and frequency. *IEEE Transactions on information Theory*, 14(3):369–374.

Michael Rutter. 1970. Autistic children: infancy to adulthood. In *Seminars in psychiatry*, volume 2, page 435.

Andrew B Short and Eric Schopler. 1988. Factors relating to age of onset in autism. *Journal of autism and developmental disorders*, 18(2):207–216.

Lawrence D Shriberg, Rhea Paul, Jane L McSweeny, Ami Klin, Donald J Cohen, and Fred R Volkmar. 2001. Speech and prosody characteristics of adolescents and adults with high-functioning autism and asperger syndrome. *Journal of Speech, Language, and Hearing Research*, 44(5):1097–1115.

Helen Tager-Flusberg, Rhea Paul, Catherine Lord, F Volkmar, Rhea Paul, and Ami Klin. 2005. Language and communication in autism. *Handbook of autism and pervasive developmental disorders*, 1:335–364.

Fred R Volkmar, Kathy Koenig, and Matthew State. 1997. Childhood disintegrative disorder. *Handbook of Autism and Pervasive Developmental Disorders, Volume 1, Third Edition*, pages 70–87.

B Yegnanarayana and K Sri Rama Murty. 2009. Event-based instantaneous fundamental frequency estimation from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(4):614–624.