Autism Speech Analysis Using Acoustic Features

Abhijit Mohanta
Indian Institute of Information Technology
Sri City, Chittoor, Andhra Pradesh, India
abhijit.mohanta@iiits.in

Vinay Kumar Mittal
Professor, K L University
Vijayawada, Andhra Pradesh, India
drvinaykrmittal@gmail.com

Abstract

Autism speech has distinct acoustic patterns, different from normal speech. Analyzing acoustic features derived from the speech of children affected with *autism spectrum disorder (ASD)* can help its early detection. In this study, a comparative analysis of the discriminating acoustic characteristics is carried out between *ASD affected* and *normal children* speech, from speech production point of view. Datasets of English speech of children affected with *ASD* and *normal* children were recorded. Changes in the speech production characteristics are examined using the *excitation source features* F0 and strength of excitation (SoE), the *vocal tract filter features* formants (F1 to F5) and dominant frequencies (FD1, FD2), and the *combined source-filter features* signal energy and zero-crossing rate. Changes in the acoustic features are compared in the five vowels regions of the English language. Significant changes in few acoustic features are observed for *ASD affected* speech as compared to *normal* speech. The differences between the *mean* values of the formants and dominant frequencies, for ASD affected and normal children, are highest for vowel /i/. It indicates that ASD affected children have possibly more difficulty in speaking the words with vowel /i/. This study can be helpful towards developing systems for automatic detection of ASD.

keywords: acoustic analyses of autism, autism spectrum disorder, ASD, dominant frequencies, formants

1 Introduction

ASD is a pervasive developmental disorder, defined clinically by observing the abnormalities in three areas: communication, social reciprocity, and hyperfocus or reduced behavioral flexibility (Kjelgaard and Tager-Flusberg, 2001; Diehl et al., 2009; McCann and Peppé, 2003). Study shows, at least 50% of the total population of ASD tends to show atypical acoustic patterns in their speech, and it persists throughout the improvement of other language aspects (DePape et al., 2012; Baltaxe and Simmons, 1985; Fusaroli et al., 2017). In fact, the exact characteristics of autism and its underlying mechanisms are also unclear (Kanner et al., 1943; Bonneh et al., 2011). According to study, 1 in 150 individuals with autism was reported in 2002, which became 1 in 68 in 2014 (Kumar et al., 2018; Autism and Investigators, 2014). It is reported that there are tens of millions of individuals with ASD worldwide, and it is affecting approximately 1.5% of our total population (Santos et al., 2013; Parish-Morris et al., 2016).

Communication impairments, abnormal voice quality and disturbances of prosody are some of the most important aspects among individuals with ASD who speak (Paul et al., 2005; Bonneh et al., 2011). Individuals with ASD speak with distinctive acoustic patterns in their speech, and as a result they face social interaction deficits (Fusaroli et al., 2017). The reason behind the language impairment in autism is the result of primary linguistic disorder with a focus on pragmatic impairments (Baltaxe, 1977). Besides, the speech signal of the children with ASD is reported as improperly modulated, wooden, and dull (Baltaxe and Simmons, 1985). In fact, in many cases, a significant spoken language delay and repetitive language can also be encountered (Mower et al., 2011). In general, normal children start establishing their vocabularies at the age of two years, whereas the children with ASD may not be able to do the same (Tager-Flusberg et al., 2005; Short and Schopler, 1988).

Previous studies mostly based on either speech prosody or unusual suprasegmental features of speech production of children with ASD (Bonneh et al., 2011). Like, in Shriberg et al. (2001), authors had reported the segmental and suprasegmental speech features of individuals with high-

D M Sharma, P Bhattacharyya and R Sangal. Proc. of the 16th Intl. Conference on Natural Language Processing, pages 85–94
Hyderabad, India, December 2019.

functioning autism (HFA). Also, some studies used objective measures to quantify speech related issues in autism (Bonneh et al., 2011). Some of the most significant analyses based on pitch features of individuals with ASD were reported in Brisson et al. (2014), Quigley et al. (2016), etc., where in each study authors had reported different result from others. For instance, in Brisson et al. (2014), authors had reported higher pitch value for ASD children as compared with normal children. On the other hand, in Quigley et al. (2016), authors had reported lower pitch value for ASD children as compared with normal children. Besides, in the case of the intensity based analyses, some of the studies indicated no significant differences between ASD and normal children (Quigley et al., 2016; Hubbard and Trauner, 2007). Likewise, based on duration (syllable duration, utterance duration, etc.), voice patterns, speech rate, etc., researchers had done some significant analyses on individuals with ASD (Santos et al., 2013; Kakihara et al., 2015; Bone et al., 2013). But, none of the previous studies had done only on English vowels, especially pronounced by non-native Indian English speakers with ASD. Also, many robust speech features like dominant frequencies (FD1, FD2), strength of excitation (SoE), etc., had not been considered in previous studies. Therefore, in this study, we have considered all these mentioned points.

This paper analyzed the *autism speech*, i.e., the speech signal of the children with ASD, by differentiating them from the normal children. Differences are made in terms of the speech production features of the ASD and the normal children. Here, only English vowels, i.e., /a/, /e/, /i/, /o/, and /u/ are taken into consideration, because of their relatively longer duration in the case of children with ASD. Also, the production of vowels sounds by an individual is not a random process; hence it is important to find characteristics of the speech production mechanism of children with ASD during the pronunciation of vowels sounds. This study on analyzing the speech production characteristics of the children with ASD has high importance, because it may play a vital role in improving the communication impairments associated with ASD. In addition, current diagnostic criteria for ASD do not include any atypical vocalizations (Bonneh et al., 2011). Hence, this study can be utilized as a diagnostic marker to identify ASD.

Table 1: Dataset Details of the *ASD* and the *Normal* Children

Attributes	*Group*	*Statistics*	
		Male	*Female*
Total children	ASD	11	02
	Normal	11	09
Age (in years)	ASD	03 to 09	3.5
	Normal	03 to 09	3.5 to 09
Reading skill (English)	ASD	Beginner	Beginner
	Normal	Beginner	Beginner
Data Duration (in sec)	ASD	6850	2500
	Normal	6000	6000

This study consists of four major steps. Firstly, two speech signal datasets were collected, by recording the sound files of the ASD and the normal children. Secondly, unwanted signal parts were removed, and the speech signal files were arranged in two different databases for the ASD and the normal children. Thirdly, speech signal processing methods were applied on the collected datasets to extract the selected production features. Finally, results were made by differentiating between the ASD and the normal children in terms of their speech production features.

The rest of the paper is organized as follows. Details about the two collected datasets of the ASD and the normal children are discussed in Section 2. Next, the signal processing methods and features used for analyses are discussed in Section 3. Section 4 presents key results and observations on results. Then, Section 5 discusses the analyses of observed results in speech production point of view. Section 6 represents key contributions. Lastly, Section 7 presents conclusions, along with the scope of future work on this topic.

2 Speech Datasets of ASD and Normal Children

Two speech signal datasets in the English language were recorded for this study, where one dataset contains the speech samples of 13 children with ASD, and another dataset contains the speech samples of 20 normal children. Details of both the datasets are given in Table 1. In this study, the number of ASD and normal children is different. There are numerous previous studies like Parish-Morris et al. (2016), Nakai et al. (2014), etc., where researchers took a different number of ASD

and normal children. Besides, children with age less than 3 years were not considered in this study, because typically the diagnosis of ASD starts by the age of 3 years when a child begins to show delays in developmental milestones (Santos et al., 2013; McCann and Peppé, 2003). Another reason was that the current study only focused on verbal children. Besides, in the case of the children with ASD, it was made sure by a well-experienced doctor and a psychologist that the children considered were diagnosed with ASD. The children with ASD considered for the data collection met the DSM-IV diagnostic criteria (Wing et al., 2011; Lord et al., 1994). Furthermore, all the children with ASD considered here had distinctive acoustic patterns in their speech, during the entire period of data collection. However, the normal children did not have any such issues and were living a normal life.

Speech samples were recorded every week (once or twice), for a period of over 1 year. Recordings took place in a noise-free empty room, which did not have any object that could distract the children. Also, the neutral emotional state of the children was affirmed during all the data collection sessions. The ASD and the normal children were asked to name in English a set of 25 specifically selected daily life pictures, shown to them along with each picture's name in English on a laptop. The pictures consisted of animals, vegetables, flowers, and English numbers. All the children were asked to pronounce only the object's name as a word, presented to them in the form of a picture. The children's first response was confronted by asking them to pronounce the picture's name. Then, we kept changing the pictures one by one, while the children named the object shown as a picture. Each child was asked to name the same set of pictures over each of the recording sessions. Five different pictures were selected for each of five English vowels, and the names of all the pictures were either in consonant-vowel-consonant (CVC) or consonant-vowel-vowel-consonant (CVVC) word format. The total utterances of 25 words by each child (5 vowels $\times$ 5 words) were recorded in each of the two such sessions, in a day.

Roland R-26 digital audio recorder was used with 48 KHz sampling rate to record the speech samples. The distance of 25 cm was maintained between the recorder and the speaker's mouth.

Our collected datasets have immense importance because of several reasons. Firstly, all the children considered here were non-native Indian English speakers. Whereas, in previous studies like Oller et al. (2010), Asgari et al. (2013), Marchi et al. (2015), Kakihara et al. (2015), etc., authors had not considered non-native Indian English speakers(children) with ASD. Secondly, in previous studies datasets were mostly collected from social interaction (Santos et al., 2013), constrained production (Bone et al., 2013) and spontaneous production (Fusaroli et al., 2017). But, here the datasets were recorded differently, as described earlier in this section.

3 Signal Processing Methods and Features

The production characteristics of speech signal of the ASD and the normal children are differentiated by examining changes in the source features, vocal tract system features and combined source-filter features. The source features F0 and strength of excitation (SoE), and the vocal tract filter features dominant frequencies (FD1, FD2) and first five formants (F1 to F5) are examined. The combined source-filter features signal energy (E) and zero-crossing rate (ZCR) are also examined. Here, for each speech feature, the mean (μ) or average values are computed. The mean values are computed for each English vowel by taking the average of all the calculated values of a particular speech feature, and this procedure is followed for each speaker. Besides, the μ_{SoE}, μ_E and μ_{ZCR} values are multiplied by 100, 1000, and 1000, respectively, for a better understanding.

3.1 Excitation Source Features

The excitation source feature F0 was derived using zero-frequency filtering (ZFF) method (Murty and Yegnanarayana, 2008; Yegnanarayana and Murty, 2009). The ZFF method involves computing the output of the cascade of two zero-frequency resonators (ZFRs). That is $y_1[n] = -\sum_{k=1}^{2} a_k y_1[n-k] + x[n]$ and $y_2[n] = -\sum_{k=1}^{2} a_k y_2[n-k] + y_1[n]$. Where, x[n] is pre-processed input signal, $a_1 = -2$ and $a_2 = 1$. This operation is repeated twice (denoted as $y_1[n]$ and $y_2[n]$) 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 the ZFF signal, given as $y[n] = y_2[n] -$

Table 2: Mean (μ) Values of the Source Features (*F0* and *SoE*), Combined Source-filter Features (*Energy E* and *Zero-crossing Rate ZCR*) and Vocal Tract Filter Features (*Formants Frequencies* and *Dominant Frequencies*) of the *Male* Children with *ASD* and *Normal* (Nm): (a) Acoustic Features and (b)-(f) Mean Values for Five English Vowels; *F1* to *F5* Indicate First Five *Formants Frequencies*, Respectively, and *FD1* and *FD2* are First and Second *Dominant Frequencies*, Respectively

(a) Features	*(b) /a/*		*(c) /e/*		*(d) /i/*		*(e) /o/*		*(f) /u/*	
	ASD	*Nm*	*ASD*	*Nm*	*ASD*	*Nm*	*ASD*	*Nm*	*ASD*	*Nm*
F0 (Hz)	263	258	267	260	271	262	269	256	246	236
SoE×100	34.9	32.1	43.4	46.3	44.0	47.7	39.1	35.7	35.9	31.1
E×1000	36.5	24.7	31.4	30.7	43.2	33.7	41.2	35.6	54.3	34.9
ZCR×1000	37.7	39.3	28.2	34.7	30.9	30.9	28.2	32.0	30.5	33.6
F1 (Hz)	720	453	554	452	589	424	657	557	662	498
F2 (Hz)	1628	1238	1665	1207	1658	1255	1310	1111	1466	1185
F3 (Hz)	2694	2486	2726	2551	2686	2566	2603	2504	2673	2446
F4 (Hz)	3712	3552	3715	3613	3675	3642	3561	3572	3603	3651
F5 (Hz)	4471	4455	4467	4435	4427	4425	4394	4320	4410	4331
FD1 (Hz)	1042	819	900	580	1043	519	863	824	952	731
FD2 (Hz)	3295	3470	3234	3171	3282	3125	3291	3375	3316	3368

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

The excitation feature, SoE was derived using the ZFF method. The slope of the ZFF signal around the glottal closure instants (GCIs) gives a measure of the SoE (Murty and Yegnanarayana, 2008; Murty et al., 2009; Mittal and Yegnanarayana, 2015b).

3.2 Vocal Tract Filter Features

The first five formants (F1 to F5) were derived by using linear prediction (LP) spectrum (Makhoul, 1975; Hermansky, 1990; Atal and Hanauer, 1971; Yegnanarayana, 1978). The sound files were resampled to 10 KHz and LP order as 10.

The first two dominant peak frequencies (FD1 and FD2) were derived from the acoustic signal using LP analysis (Makhoul, 1975; Hermansky, 1990). With the LP order 5, the LP spectrum will have a maximum of two peaks corresponding to two complex conjugate pole pairs (Mittal et al., 2014). The corresponding frequencies of these two peaks are known as the dominant frequencies, denoted as FD1 and FD2, respectively (Mittal and Yegnanarayana, 2015a). The dominant frequencies represent the frequency response with high spectral energies. These high spectral energies give an idea of the concentration of energy in the spectrum (Mittal and Yegnanarayana, 2015a).

3.3 Combined Features

The 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.

In the context of discrete-time signals, ZCR is defined as the number of times in any specific time interval/frame that the amplitude of the speech signal goes through a value of zero (Bachu et al., 2008). The definition of ZCR as given in (Bachu et al., 2008) is $Z_n = \sum_{m=-\infty}^{\infty} |sgn[x(m)] - sgn[x(m-1)]|\, w(n-m)$. Where, $sgn[x(n)] = \begin{cases} 1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases}$ and $w(n) = \begin{cases} \frac{1}{2N} for, 0 \leq n \leq N-1 \\ 0\, for, otherwise \end{cases}$.

4 Results and Observations

The obtained results indicate higher μ_{F0} values for the children with ASD as compared with the normal children, and this statement is true for all English vowels. Besides, according to the tongue position, female children with ASD have the highest μ_{F0} value for mid-vowel /e/ and have the low-

Table 3: Mean (μ) Values of the Source Features (*F0* and *SoE*), Combined Source-filter Features (*Energy E* and *Zero-crossing Rate ZCR*) and Vocal Tract Filter Features (*Formants Frequencies* and *Dominant Frequencies*) of the *Female* Children with *ASD* and *Normal* (Nm): (a) Acoustic Features and (b)-(f) Mean Values for Five English Vowels; *F1* to *F5* Indicate First Five *Formants Frequencies*, Respectively, and *FD1* and *FD2* are First and Second *Dominant Frequencies*, Respectively

(a) Features	*(b) /a/*		*(c) /e/*		*(d) /i/*		*(e) /o/*		*(f) /u/*	
	ASD	*Nm*	*ASD*	*Nm*	*ASD*	*Nm*	*ASD*	*Nm*	*ASD*	*Nm*
F0 (Hz)	326	314	343	321	339	330	340	310	335	313
SoE×100	32.0	30.5	37.1	41.9	38.1	50.9	42.9	33.0	37.1	35.3
E×1000	39.5	23.4	35.3	24.2	48.3	20.7	63.7	34.9	58.6	30.3
ZCR×1000	35.6	55.3	26.9	48.3	29.3	39.4	29.9	38.3	32.1	40.8
F1 (Hz)	711	457	572	517	633	438	670	646	693	546
F2 (Hz)	1554	1261	1636	1213	1630	1141	1322	1231	1438	1278
F3 (Hz)	2653	2484	2776	2523	2746	2476	2537	2487	2588	2532
F4 (Hz)	3720	3559	3778	3571	3782	3501	3558	3612	3629	3649
F5 (Hz)	4439	4411	4425	4404	4429	4411	4417	4345	4396	4348
FD1 (Hz)	865	827	686	783	681	560	803	784	860	810
FD2 (Hz)	3185	3436	3286	3111	3269	3129	3058	3432	3112	3175

est μ_{F0} value for low-vowel /a/ as compared with other English vowels. But, in the case of the normal female children, high-vowel /i/ gives the highest and mid-vowel /o/ gives the lowest μ_{F0} values as compared with other English vowels. However, in the case of the male children with ASD, such results have not been found. It is observed that male children with ASD follow a similar μ_{F0} trend with the normal male children for all English vowels. These results can be analyzed from Table 2 and 3.

Like μ_{F0}, in the case of μ_E also, the children with ASD have higher values for all the five English vowels as compared with the normal children. Also, for all the five English vowels, the female children with ASD have higher μ_E values as compared with the male children with ASD, but this is vice versa for the normal children. Besides, in the case of the children with ASD, the same vowel /e/ has the lowest μ_E values for both male and female children, whereas this is not the same for the normal male and female children. Likewise, in the case of the normal children, the same vowel /o/ has the highest μ_E values for both male and female children, whereas this is not true for the male and female children with ASD. These statements can be observed from μ_E values in Table 2 and 3.

Regarding μ_{SoE}, only front vowels /e/ and /i/ indicate lower values for the children with ASD as compared with the normal children. But, in the case of mid and rear vowels, i.e., /a/, /o/, and /u/, μ_{SoE} indicate higher values for the children with ASD than the normal children. Besides, in the case of both the normal male and female children, the same vowel /i/ has the highest μ_{SoE} values as compared with other English vowels. But, this statement is not true in the case of the children with ASD. Again, in the case of both the male and female children with ASD, the same vowel /a/ has the lowest μ_{SoE} values as compared with other English vowels, whereas this is not the case with the normal children. All these results can be observed from μ_{SoE} values, tabulated in Table 2 and Table 3.

The μ_{ZCR} have lower values for the children with ASD as compared with the normal children, and it is true for all English vowels. This observation is graphically represented in Figure 1(g) and 1(h). Also, in the case of the front and mid vowels, i.e., /a/, /e/, and /i/, the male children with ASD have higher μ_{ZCR} values as compared with the female. But, it is vice versa in the case of the normal children. Besides, in the case of both male and female children with ASD, the same vowel /e/ has the lowest μ_{ZCR} values as compared with other English vowels, whereas this is not the case with the normal children. These results can be observed from μ_{ZCR} values, given in Table 2 and 3.

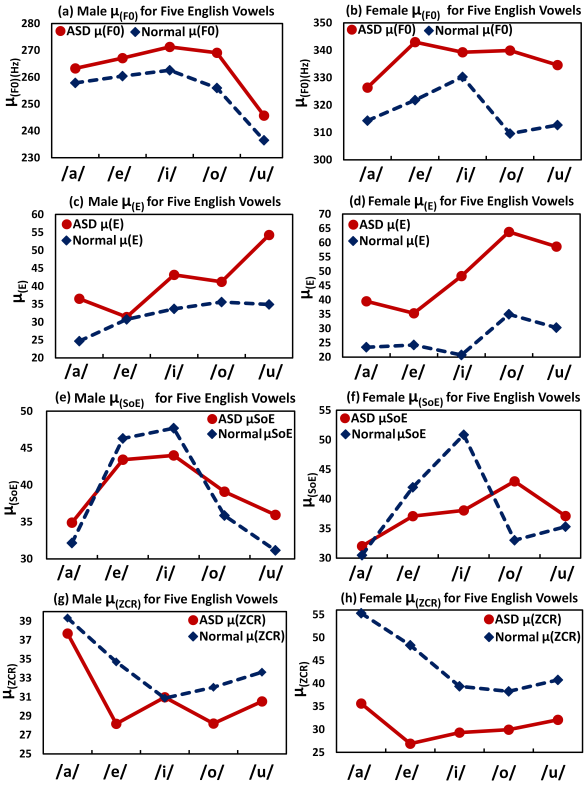

Figure 1: Differences in the Mean Values of F0, E, SoE, and ZCR between the *ASD Affected* and the *Normal* Children.

The children with ASD have significantly higher μ_{F1} values for all English vowels as compared with the normal children. Next, it is observed that the normal female children have higher μ_{F1} values for all the five English vowels as compared with the normal male children, whereas this statement is not true in the case of the children with ASD. According to the tongue position, in the case of both the male and female children with ASD, the μ_{F1} indicates the highest values for the low vowel /a/ as compared with the high and mid vowels. But, in the case of both the male and female normal children, the μ_{F1} indicates the highest values for the mid vowel /o/ as compared with the high and low vowels. The μ_{F1} results are tabulated in Table 2 and 3.

The μ_{F2} values are higher for all English vowels in the case of the children with ASD as compared with the normal children. Also, the μ_{F2} values for all the five English vowels of both the male and female children with ASD follow a similar trend, whereas there is no such trend observed in the case of the normal children. Besides, according to the tongue position, both the male and female children with ASD have the highest μ_{F2} values for the mid vowel /e/ as compared with the high and low vowels. But, in the case of the normal children, as compared with the mid and low vowels the high vowels /i/ and /u/ give the highest μ_{F2} values for both the male and female children, respectively. All these results can be analyzed from μ_{F2} values tabulated in Table 2 and 3.

Like μ_{F1} and μ_{F2}, the μ_{F3} values are also higher for all English vowels in the case of the children with ASD as compared with the normal children. According to the tongue position, in the case of both the male and female children with ASD, the μ_{F3} indicates the highest values for the mid vowel /e/ as compared with the high and low vowels. But, in the case of the normal children, the μ_{F3} indicates the highest values for the high vowels (/i/ and /u/) as compared with the mid and low vowels. The μ_{F3} values are tabulated in Table 2 and 3.

As compared with the normal children, the children with ASD have higher μ_{F4} values for the front and mid vowels only. Next, according to the tongue position, in the case of both the male and female children with ASD, the μ_{F4} gives the highest values for the mid vowel /o/ as compared with the high and low vowels. But, this is not the case for the normal children. The μ_{F4} results can be analyzed from the Figure 2(g) and 2(h), also from the μ_{F3} values, tabulated in Table 2 and 3.

The μ_{F5} indicates higher values for all the five English vowels in the case of the children with ASD as compared with the normal children, depicted in Figure 2(i) and 2(j). Also, both the male and female normal children have the lowest μ_{F5} values for the mid vowel /o/ as compared with the high and low vowels. But, this statement is not true in the case of the ASD children. The μ_{F5} values are tabulated in Table 2 and 3.

All the five English vowels have higher μ_{FD1} values for the children with ASD as compared with the normal children, depicted in Figure 3(a) and 3(b). According to the tongue position, both the male and female normal children have the lowest μ_{FD1} values for the high vowel /i/ as compared with the mid and low vowels. But, in the case of the ASD children, as compared with the high and low vowels the mid vowels /e/ and /o/ indicate the lowest μ_{FD1} values for both the female and male, respectively. The μ_{FD2} results can be analyzed from Table 2 and 3.

In the case of μ_{FD2}, only the front vowel /e/ and

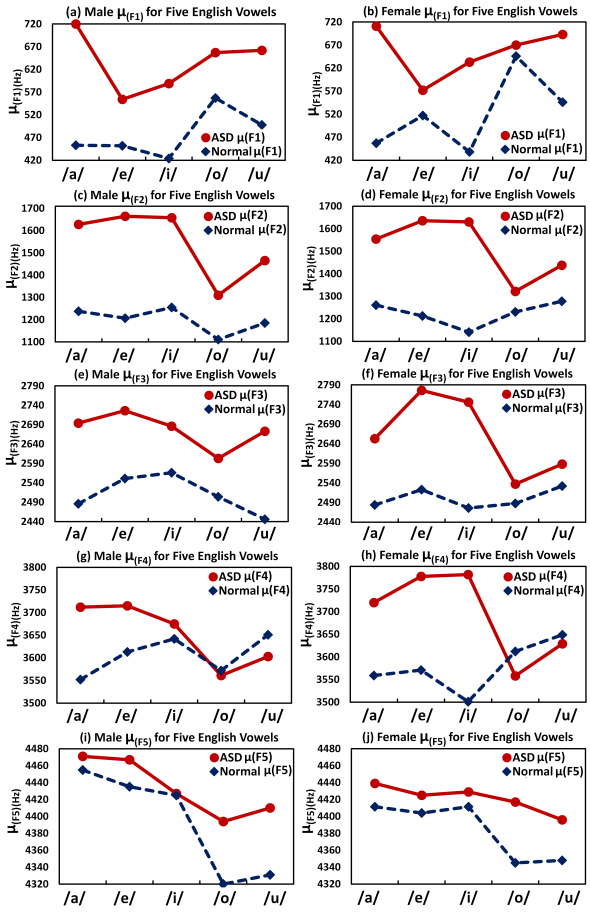

Figure 2: Differences in the Mean Values of *Formants Frequencies* (F1, F2, F3, F4, and F5) between the *ASD Affected* and the *Normal* Children.

/i/ have higher values for the children with ASD as compared with the normal children, graphically shown in Figure 3(c) and 3(d). In addition, according to the tongue position, both the male and female normal children have the highest μ_{FD2} values for the low vowel /a/ as compared with the mid and high vowels. On the other hand, as compared with other English vowels the high vowel /u/ has the highest μ_{FD2} value for the male ASD group and the mid vowel /e/ has the highest μ_{FD2} value for the female ASD group. The μ_{FD2} values are tabulated in Table 2 and 3 for the male and female children, respectively.

5 Analyses of Results

This section describes the observed results in speech production point of view. Firstly, the F0 which reveals the source characteristics of the speech production system, the result infers that in the case of all the five English vowels, the male and female children with ASD have a higher vocal fold vibration rate than the normal male and female children. This statement is true for all the five English vowels. Furthermore, in the case of female children with ASD, mid-vowel /e/ has the highest and low-vowel /a/ has the lowest vocal fold vibration rate as compared with other English vowels. On the other hand, in the case of the normal female children, high-vowel /i/ has the highest and mid-vowel /o/ has the lowest vocal fold vibration rate as compared with other English vowels. These observations can be analyzed from Figure 1(a) and 1(b).

In the case of E which gives the information about the combined source-system characteristics of the speech production system, the result implies that the children with ASD have louder speech and put more vocalization effort than the normal children. Also, in the case of all English vowels the female children with ASD put more vocalization effort than the male children with ASD, but this is vice versa in the case of the normal group. These results can be analyzed from μ_E values graphically depicted in Figure 1(c) and 1(d).

The observed SoE result infers that in the case of the front vowels the strength of impulse-like excitation is lower during the glottal activity (vibration of vocal folds) of the children with ASD as compared with the normal children. But, in the case of mid and rear vowels the strength of impulse-like excitation is higher for the ASD children than the normal children. This result can be analyzed from Figure 1(e) and 1(f).

The F1 result implies that in the case of all five English vowels, the children with ASD have a lesser oral constriction in the front half of the oral section of the vocal tract as compared with the normal children. Again, in terms of pharyngeal constriction, it can be stated that during the pronunciation of all the five English vowels the pharyngeal constriction is greater for the children with ASD as compared with the normal children. The F1 observed result also implies that both the male and female children with ASD have the greatest pharyngeal constriction for the low-vowel /a/ as compared with the mid and high vowels. But, the normal male and female children have the greatest pharyngeal constriction for the mid-vowel /o/ as compared with the high and low vowels. Furthermore, during the pronunciation of all English vowels the children with ASD increase their tongue higher than the normal children. Because the F1

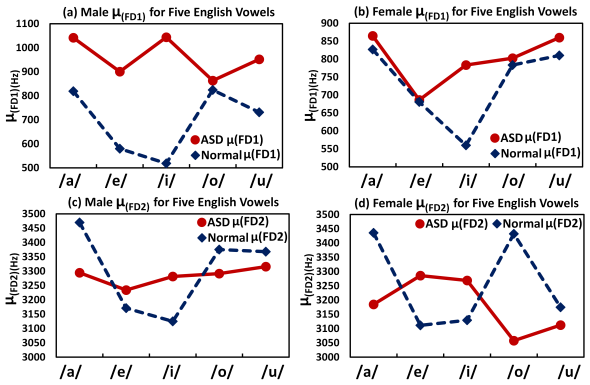

Figure 3: Differences in the Mean Values of *Dominant Frequencies* (FD1 and FD2) between the *ASD Affected* and the *Normal* Children.

value increases with increasing the tongue position higher. The F1 values for all English vowels are graphically depicted in Figure 2(a) and 2(b).

The F2 result implies that in the case of all English vowels the back tongue constriction is lesser and the front tongue constriction is greater for the children with ASD than the normal children. Furthermore, it can be stated from the observed result that both the male and female children with ASD have the least back tongue constriction and the greatest front tongue constriction for the mid vowel /e/ as compared with the high and low vowels. On the other hand, the normal male children have the least back tongue constriction and the greatest front tongue constriction for the high-vowel /i/ as compared with the mid and low vowels, and the normal female children have the least back tongue constriction and the greatest front tongue constriction for the high-vowel /u/ as compared with the mid and low vowels. This observation can be analyzed from Figure 2(c) and 2(d).

The F3 result implies that in the case of the children with ASD lip-rounding is lesser during the pronunciation of all English vowels. Hence, the constriction is least and as a result all English vowels give higher μ_{F3} frequency values for the children with ASD as compared with the normal children. The results are graphically depicted in Figure 2(e) and 2(f).

Also, the results of the first three formants (F1, F2 and F3) indicate that the length of the pharyngeal-oral tract is shorter in the case of the children with ASD as compared with the normal children. Because, the formants values of vowels are inversely proportional to the pharyngeal-oral tract, and here the children with ASD have higher μ_{F1}, μ_{F2} and μ_{F3} values for all English vowels as compared with the normal children. Also, in terms of the lip-rounding, the F1, F2, F3 and F5 results imply that the children with ASD have a lesser lip-rounding as compared with the normal group.

In the case of formants frequencies and dominant frequencies, the differences between the ASD and the normal children are highest for vowel /i/. It implies that ASD children have probably more difficulty in pronouncing the words with vowel /i/.

6 Key Contributions

The key contributions of this study are as follows:

- The ASD and the normal children's speech datasets are collected by recording the speech samples of non-native Indian English speakers.
- Only English vowels (/a/, /e/, /i/, /o/, and /o/) are considered in this study.
- Some of the robust speech features like SoE, F5, FD1, and FD1 are considered here, which were not considered in similar types of previous studies.
- The F0, E, F1, F2, F3, and F5 results clearly distinguish the ASD and the normal children. All these features have significantly higher mean values for all English vowels in the case of the ASD children as compared with the normal children.
- The results of the formants and dominant frequencies indicate that children with ASD have probably more difficulty in pronouncing the words with vowel /i/.

7 Conclusions

The aim of this study is to analyze differences in various speech production features of the children with ASD as compared with the normal children. Only English vowels sounds are used in this study. An autism speech dataset and a normal childrens speech dataset are recorded separately for this research purpose. Then, differences between the children with ASD and the normal children are analyzed by observing the source characteristics (F0 and SoE), system characteristics (dominant frequencies and formants), and combined characteristics (ZCR and E). It is observed that there are significant differences between the ASD and the

normal children, in terms of their speech production characteristics in English vowels regions. In the case of most of the speech production features, the ASD children have significantly higher values than the normal children. These acoustic characteristics of the children with ASD can be used as markers to identify ASD. But, we did not find any single speech feature that can be utilized as a diagnostic marker for ASD.

A small size of speech data for female ASD children is a limitation of this study. In future studies, we will try to find a single speech feature that can be utilized as an acoustic marker to identify ASD.

Acknowledgments

The authors are thankful to Dr. N. P. Karthikeyen and DOAST Integrated Therapy Centre for Autism, Chennai, India, for providing the opportunity for ASD children's voice recording. The authors are also thankful to Chinmaya Vidhyalaya school, Sri City, Andhra Pradesh, India, for providing the opportunity to record normal children's speech dataset.

References

Meysam Asgari, Alireza Bayestehtashk, and Izhak Shafran. 2013. Robust and accurate features for detecting and diagnosing autism spectrum disorders. In *Interspeech*, pages 191–194.

Bishnu S Atal and Suzanne L Hanauer. 1971. Speech analysis and synthesis by linear prediction of the speech wave. *The journal of the acoustical society of America*, 50(2B):637–655.

Autism and Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators. 2014. Prevalence of autism spectrum disorder among children aged 8 yearsautism and developmental disabilities monitoring network, 11 sites, united states, 2010. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 63(2):1–21.

RG Bachu, S Kopparthi, B Adapa, and BD Barkana. 2008. Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal. In *American Society for Engineering Education (ASEE) Zone ference Proceedings*, pages 1–7.

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

Christiane AM Baltaxe and James Q Simmons. 1985. Prosodic development in normal and autistic children. In *Communication problems in autism*, pages 95–125. Springer.

Daniel Bone, Theodora Chaspari, Kartik Audhkhasi, James Gibson, Andreas Tsiartas, Maarten Van Segbroeck, Ming Li, Sungbok Lee, and Shrikanth Narayanan. 2013. Classifying language-related developmental disorders from speech cues: the promise and the potential confounds. In *INTERSPEECH*, pages 182–186.

Yoram S Bonneh, Yoram Levanon, Omrit Dean-Pardo, Lan Lossos, and Yael Adini. 2011. Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in human neuroscience*, 4:237.

Julie Brisson, Karine Martel, Josette Serres, Sylvain Sirois, and Jean-Louis Adrien. 2014. Acoustic analysis of oral productions of infants later diagnosed with autism and their mother. *Infant mental health journal*, 35(3):285–295.

Anne-Marie R DePape, Aoju Chen, Geoffrey BC Hall, and Laurel J Trainor. 2012. Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Frontiers in psychology*, 3:72.

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.

Hynek Hermansky. 1990. Perceptual linear predictive (plp) analysis of speech. *the Journal of the Acoustical Society of America*, 87(4):1738–1752.

Kathleen Hubbard and Doris A Trauner. 2007. Intonation and emotion in autistic spectrum disorders. *Journal of psycholinguistic research*, 36(2):159–173.

Yasuhiro Kakihara, Tetsuya Takiguchi, Yasuo Ariki, Yasushi Nakai, Satoshi Takada, Y Kakihara, et al. 2015. Investigation of classification using pitch features for children with autism spectrum disorders and typically developing children. *Am. J. Sign. Process*, 5:1–5.

Leo Kanner et al. 1943. Autistic disturbances of affective contact. *Nervous child*, 2(3):217–250.

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.

Manoj Kumar, Pooja Chebolu, So Hyun Kim, Kassandra Martinez, Catherine Lord, and Shrikanth Narayanan. 2018. A knowledge driven structural segmentation approach for play-talk classification during autism assessment. In *Interspeech*.

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 Schuller, Simon Baron-Cohen, Ofer Golan, Sven Bölte, Prerna Arora, and Reinhold Häb-Umbach. 2015. Typicality and emotion in the voice of children with autism spectrum condition: Evidence across three languages. In *Sixteenth Annual Conference of the International Speech Communication Association*.

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

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

Vinay Kumar Mittal, B Yegnanarayana, and Peri Bhaskararao. 2014. Study of the effects of vocal tract constriction on glottal vibration. *The Journal of the Acoustical Society of America*, 136(4):1932–1941.

Emily Mower, Chi-Chun Lee, James Gibson, Theodora Chaspari, Marian E Williams, and Shrikanth Narayanan. 2011. 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.

K Sri Rama Murty, Bayya Yegnanarayana, and M Anand Joseph. 2009. Characterization of glottal activity from speech signals. *IEEE signal processing letters*, 16(6):469–472.

Yasushi Nakai, Ryoichi Takashima, Tetsuya Takiguchi, and Satoshi Takada. 2014. Speech intonation in children with autism spectrum disorder. *Brain and Development*, 36(6):516–522.

D Kimbrough Oller, P Niyogi, S Gray, Jeffrey A Richards, Jill Gilkerson, Daoyi Xu, Umit Yapanel, and Steven F Warren. 2010. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, 107(30):13354–13359.

Julia Parish-Morris, Mark Liberman, Neville Ryant, Christopher Cieri, Leila Bateman, Emily Ferguson, and Robert Schultz. 2016. Exploring autism spectrum disorders using hlt. In *Proceedings of the third workshop on computational lingusitics and clinical psychology*, pages 74–84.

Rhea Paul, Lawrence D Shriberg, Jane McSweeny, Domenic Cicchetti, Ami Klin, and Fred Volkmar. 2005. Brief report: Relations between prosodic performance and communication and socialization ratings in high functioning speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 35(6):861.

Jean Quigley, Sinéad McNally, and Sarah Lawson. 2016. Prosodic patterns in interaction of low-risk and at-risk-of-autism spectrum disorders infants and their mothers at 12 and 18 months. *Language Learning and Development*, 12(3):295–310.

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

Joao F Santos, Nirit Brosh, Tiago H Falk, Lonnie Zwaigenbaum, Susan E Bryson, Wendy Roberts, Isabel M Smith, Peter Szatmari, and Jessica A Brian. 2013. Very early detection of autism spectrum disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7567–7571. IEEE.

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.

Lorna Wing, Judith Gould, and Christopher Gillberg. 2011. Autism spectrum disorders in the dsm-v: better or worse than the dsm-iv? *Research in developmental disabilities*, 32(2):768–773.

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.

Bayya Yegnanarayana. 1978. Formant extraction from linear-prediction phase spectra. *The Journal of the Acoustical Society of America*, 63(5):1638–1640.