Assessment of Autism Spectrum Disorder in Toddlers using Speech Features

Harshit Kumar Gupta
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi-110 016, INDIA
Email: harshit.knit@gmail.com

Dr. Santanu Chaudhury
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi-110 016, INDIA
Email: santanuc@ee.iitd.ernet.in

Dr. Nandini Chatterjee Singh
Natiional Brain Research Centre
Manesar, Gurgaon-122 051, India
Email: nandini@nbrc.ac.in

***Abstract*—Machine learning methods with signal processing techniques are used widely for classification tasks. This paper presents a novel classification technique for categorizing the children into two groups, autistic children and typical children, using low-level acoustic features from their speech. First, wavelets are used for extracting features form speech signal. Wavelets provide both spectral and temporal resolution of speech signal unlike Fourier transform which provide only spectral resolution. Various parameters were derived from these extracted wavelet features. Then support vector machines were trained on these parameters for classification. Comparison between classification accuracy of two wavelet architectures namely discrete wavelet transform and discrete wavelet packet analysis are shown. Experimental results show that discrete wavelet transform used with support vector machine outperforms discrete wavelet packet analysis.**

Index Terms- autism spectrum disorder(ASD), discrete wavelet transform(DWT), discrete wavelet packet analysis(DWPA), classification, support vector machine(SVM)

I. INTRODUCTION

The basic idea behind this research was to use speech features to diagnose this neurological disease. For this diagnosis, extraction of features from speech is currently an emerging field. One such neurological disease is Autism Spectrum Disorder in which children lack communication and interaction ability with society. These children exhibit repetitions in their activity, behavior and functioning. Children with ASD have following characteristics: delayed patterns of speech, missing or abnormal communication gestures, speech patterns diverging from normal, different speech qualities and reduced verbal communication. The purpose of this research is to recognize this Autism Spectrum Disorder among the children by utilizing the features extracted from their speech.The features helps to detect the neurological disease in its early stage.

It is very important to understand the behavior of autistic children by acoustically monitoring them. It is an open problem in signal processing and machine learning to reliably identify autistic children by speech sample collected in the acoustic environment. Some of the major challenges include simultaneously vocalizing multiple voices and background noise.

Currently the diagnosis is done only by autism specialist doctors. Sometimes diagnosis takes too much time or due to the ignorant nature of parents, it becomes too late to help children to recover from autism spectrum disorder. This paper tries to define a method so that the diagnosis can be done without human intervention. The only thing required will be the speech signal of the subject.If we could identify articulatory signature in its early stage among children, then treatment of those autistic children can be done at early age so that they can become normal.

II. RELATED WORK

A. Study of Modulation Spectrum of Speech

In this research they analyzed natural sound ensemble by calculating the probability of amplitude envelope of sound. They obtain a Modulation Spectrum acoustic by calculating 2-D Fourier transform of the auto-correlation matrix of sound stimulus in its spectrogram. It was shown that sounds cannot have a sudden change in spectral and temporal modulation simultaneously. Speech signals are generally low passed signal and contains most of their spectral modulation power in low temporal modulation.

First Spectrogram (time-frequency representation) is calculated from speech by applying Fourier transform on each frame of signal. Then spectrogram is decomposed as a weighted sum of sine or cos grating of varying period, amplitude and phase. Each grating of spectrogram corresponds to particular sound called ripple sound acoustic. These ripple sounds are sinusoidal function in time and frequency which can be represented as follows:

$$S_i(t,f) = A_i \cos(2\pi\omega_{t,i}t + 2\pi\omega_{f,i} + \phi_i) \tag{1}$$

Spectrogram of sound can be written as sum of ripple components as follows:

$$S(t,f) = A_0 + \sum_i S_i(t,f) \tag{2}$$

A_i denote the relative amplitude of modulation depth (A_0).Here ω_f denote temporal modulation frequency in Hertz , ω_f denote spectral modulation frequency and ϕ represents initial phase of ripple. Number of frequency units spanned over one second is calculated by $\frac{\omega_t}{\omega_f}$. This term is also known as ripple velocity. Power density of each ripple component is estimated and plotted in 2-dimensional graph.

An alternate way to calculate the modulation spectrum is to decompose a signal into frequency bands using Gaussian filters. This decomposition results into narrow-band signals. Then an auto-correlation matrix is calculated by cross-correlation of narrow-band signal with all other narrow-band signals including itself. Then all cross-correlation function is averaged over same band to calculate auto-correlation matrix. Then the 2-D Fourier transform of this auto-correlation matrix is plotted which is known as modulation spectrum.

B. Study of Pitch Patterns in Children

This study focuses on finding patterns in pitch of children with ASD. In autistic children speech features are found in unusual prosody [**?**]. Prosody means only the style of speaking. Prosody can be measured with the help of following measures:

1) Pitch (F0) Measures : Measure average, max and min pitch.
2) Pitch Range : Variation between max and min pitch.
3) Pitch Excursion : Non native speaker of any language uses pitch of the accented syllable, while native speaker use pitch, frequency and amplitude. If we analyze the pitch contour, then it has more than 13 semitones/Sec. We calculate pitch excursion by following a formula:

$$PitchExcursion = \frac{39.863 \log \frac{F_{max}}{F_{min}}}{duration} \qquad (3)$$

If we see the fundamental frequency contour of each speech chunks, then there is a lot variation in F0 frequency over time. During analysis they found that there is a large gap between the values of Mean Pitch, Mean Pitch Range and Mean Pitch Excursion in different group of children (ASD, DD and TD). This study shows that pitch patterns of children in age 4 to 10 years with ASD are different from normal children of the same age group.

C. Study of Speech Motor Function in Children

There have been studies which suggest that we can predict ASD in children with the help of some signature other than their communication and interaction ability. Analysis of speech shows us difference in verbal abilities of children with ASD even at a very early stage. Development of speech motor function motor tries to use articulatory features. A measure of articulatory feature is based on acoustic differences in vocalizations. Vocal pattern belongs to two classes of articulatory features which include slow varying or fast varying amplitude.

Spectral analysis techniques are used on speech samples of children to make judgment whether articulatory features are good speech motor function. Two techniques were used for speech sampling, one in which child and one of parent interact and second in which a group of children participate in an activity. A novel technique called Speech Modulation Spectrum was developed to analyze articulatory features. At a particular time if we analyze variation in energy on a range of frequency, then that is called spectral modulations. Similarly at a particular frequency if we analyze variation in energy on a range of time then that is called temporal modulations.

Speech Modulation Spectrum is generated by taking the 2-D Fourier transform of the spectrogram. This modulation spectrum contains bands of three different types of information. These articulatory features arti are as follows.

1) Syllabic Rhythm (SR) : Region between 2 to 10 Hertz.
2) Formant Transition (FT) : Region between 25 to 40 Hertz.
3) Place of Articulation (POA) : Region between 50 to 100 Hertz.

Modulation Spectrum is a plot of Temporal modulation v's Spectral modulation in which it shows the power distribution with respect to SR, FT and POA. The area inside contour of these regions defines articulatory features. At last correlation between articulatory features and behavioral features is analyzed to find patterns.

III. CLASSIFICATION SYSTEM

Detail of preprocessing and feature extraction techniques are described as follows.

A. Preprocessing

The speech samples of children were sampled at rate of 44.1 KHz with 16-bit resolution. In this experiment we have speech samples of different durations. So we padded the signal with zero to make their length equal. We divide continuous 1-D speech signals into small frames of N samples, then we take next frame by shifting the window M samples, this causes adjacent samples to overlap by M-N samples.Choice of M and N is an application specific maintaining constraint $M < N$. We choose frame size sufficiently small to capture significant information.

We also apply window function to each frame to prevent spectral leakage at start and end of frame. We have used hamming window function which optimizes maximum side lobe of signal.

B. Wavelet Transform

Wavelet are functions which provide an orthonormal basis for function in L^2 space like sine and cosine are an orthonormal basis in the Fourier transform. Wavelet provides us multi-resolution analysis.

All wavelets are derived from mother wavelet, so wavelet with scale s and time τ is equal to mother wavelet normalized by the square root of scale s and shifted in time by τ and change in scale s.

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s}) \qquad (4)$$

where s represents scale and τ represents time.

Formula of wavelet transform can be written as follows:

$$\gamma(s,\tau) = \int f(t)\psi^*_{s,\tau}(t)dt \qquad (5)$$

where $\psi^*_{s,\tau}$ represent complex conjugate of mother wavelet.

Formula of inverse wavelet transform can be written as follows:

$$f(t) = \int\int \gamma(s,\tau)\psi_{s,\tau}(t)d\tau ds \qquad (6)$$

After wavelet Transform we find two types of coefficients:

1) Approximation Coefficients : These coefficients represent high scale, low frequency component of the signal. We calculate these coefficients by applying low decomposition filter (LDF) on the signal and then applying down-sampling by a factor of 2.
2) Detail Coefficients : These coefficients represent low scale, high frequency component of the signal. We calculate these coefficients by applying high decomposition filter (HDF) on the signal and then applying down-sampling by a factor of 2.

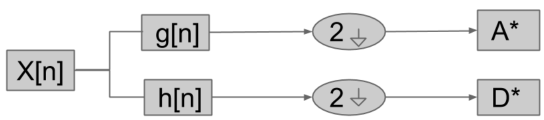

Fig. 1. Decomposition of signal

C. Discrete Wavelet Transform

Discrete Wavelet Transform utilizes low frequency component (approximation coefficient). In this transform we iteratively perform low pass filtering and high pass filtering at each level on approximation coefficient only. This multiple level decomposition is also called Wavelet decomposition tree.

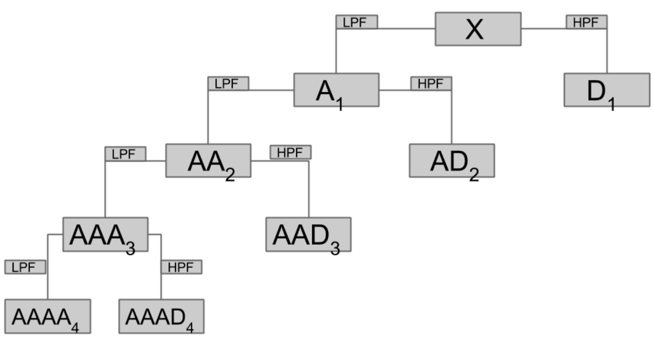

Fig. 2. Discrete Wavelet Transform

D. Discrete Wavelet Packet Analysis

Discrete Wavelet Packet Analysis utilizes both low frequency component (approximation coefficient) and high frequency component (detail coefficient). In this transform we iteratively perform low pass filtering and high pass filtering at each level. This multiple level decomposition is called Optimal Subband Tree Structuring.

E. Support Vector Machine (SVM)

This classifier constructs a hyperplane or a set of hyperplanes in high or infinite dimensional space. This classifier minimizes empirical classification error and maximize geometric margin. This is also known as a maximum margin classifier. It can also implement non-linear classifiers by applying kernel trick to the maximum margin classifier.

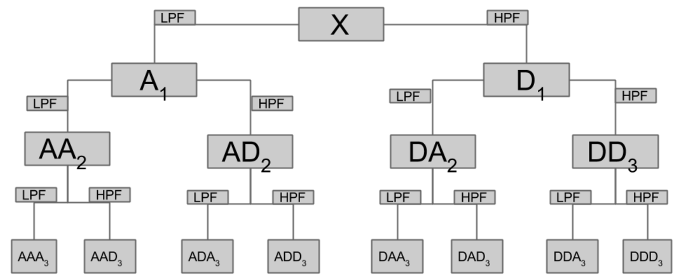

Fig. 3. Discrete Wavelet Packet Analysis

IV. Experiments

A. Datasets

We performed classification on dataset of 67 speech samples having 40 speech samples of autistic children and 27 speech samples of typical children. We have used Leave One Out cross validation to test the classification accuracy of models as our dataset is very small.

B. Feature Construction from DWT or DWPA coefficients

We apply DWT or DWPA on speech frames to get approximation and detail coefficients on several levels of decomposition. Then we take the sum and the variance of the coefficients of frames at each level. Then we take difference between these sum and variance of successive frames.

If F is all approximation or detail coefficients after LPF and HPF and i represent frame number then feature is constructed from following:

1) mean(F_i)
2) std(F_i)
3) mean(F_i) - mean(F_{i-1})
4) std(F_i) - std(F_{i-1})
5) mean(F_{i+1}) - $2 * mean(F_i)$ + mean(F_{i+1})
6) std(F_{i+1}) - $2 * std(F_i)$ + std(F_{i+1})

C. Selection of Number of decomposition levels and Mother wavelet

We should choose mother wavelet which should be computationally efficient and provide better distinguishing characteristics. We perform experiments from Haar, Symlet, Daubechies wavelet using different levels from 1 to 8. Then the number of decomposition level and mother wavelet is decided by their performance. We experimented a lot to get the best number of level of decomposition and best mother wavelet. We found out Daubechies-8 (db8) with 5 levels of decomposition gives the best results with classifiers.

D. Selection of Parameters of Classification Model

Support vector Machine For the selection of parameters we build different models with different parameters and select the model which gives better results. Following are the different parameters which we tuned by exhaustive search.

1) kernel : linear,polynomial, radial basis function (RBF) kernel

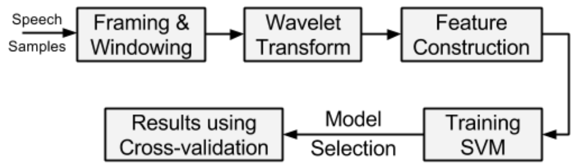

Fig. 4. Classification with Wavelet features

2) C : Penalty Parameter
3) gamma : Kernel coefficient for poly and RBF kernel
4) degree : Degree of polynomial kernel

V. Results

In this paper we have obtained a classification system for autistic and typical children. This system uses wavelet analysis to generate low-level acoustic feature which are used by SVM for training. 5-fold cross-validation score with DWT features is shown.

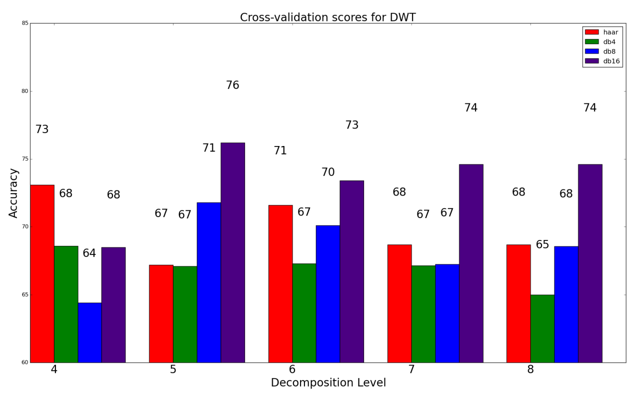

Fig. 5. Discrete Wavelet Transform

We have obtained best classification result using db-16 wavelet and doing 5 level of decomposition in DWT.

VI. Conclusion

The conclusion goes here.

Acknowledgment

The authors would like to thank National Brain Research Center, Manesar and UW Autism Center, University of Washington, Seattle for providing speech database of autistic and typical children for this experiment.

References

[1] Latika Singh and Nandini C. Singh, "The development of articulatory signatures in children" *Developmental Science*, pp 467-473, 2008.

[2] Nandini C. Singh and Frdric E. Theunissen, "Modulation spectra of natural sounds and ethological theories of auditory processing",*Journal of the Acoustical Society of America*, 114, 3394-3411, 2003

[3] Sharda M, Subhadra TP, Sahay S, Nagaraja C, Singh L, Mishra R, Sen A, Singhal N, Erickson D, Singh NC., "A novel method for assessing the development of speech motor function in toddlers with autism spectrum disorders", *Frontiers in Integrative Neuroscience*, doi:10.3389/fnint.2013.00017, 2013

[4] Sullivan K, Sharda M, Greenson J, Dawson G, Singh NC., "Sounds of melody–pitch patterns of speech in autism", *Neuroscience Letters*, doi:10.1016/j.neulet.2010.04.066, 2010

[5] Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E., "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, vol.12, pp.2825-2830, 2011

[6] Eric Jones and Travis Oliphant and Pearu Peterson and others, "SciPy: Open source scientific tools for Python", [Online]. Available:http://www.scipy.org/,2001