


Comparisons of Speech Parameterisation Techniques for Classification of Intellectual Disability Using Machine Learning

Gaurav Aggarwal, Manipal University Jaipur, Jaipur, India

 <https://orcid.org/0000-0003-2261-1657>

Latika Singh, Ansal University, Gurgaon, India

ABSTRACT

Classification of intellectually disabled children through manual assessment of speech at an early age is inconsistent, subjective, time-consuming and prone to error. This study attempts to classify the children with intellectual disabilities using two speech feature extraction techniques: Linear Predictive Coding (LPC) based cepstral parameters, and Mel-frequency cepstral coefficients (MFCC). Four different classification models: k-nearest neighbour (k-NN), support vector machine (SVM), linear discriminant analysis (LDA) and radial basis function neural network (RBFNN) are employed for classification purposes. 48 speech samples of each group are taken for analysis, from subjects with a similar age and socio-economic background. The effect of the different frame length with the number of filterbanks in the MFCC and different frame length with the order in the LPC is also examined for better accuracy. The experimental outcomes show that the projected technique can be used to help speech pathologists in estimating intellectual disability at early ages.

KEYWORDS

Classification, Intellectual Disability (ID), Linear Predictive Coding, Mel-Frequency Cepstral Coefficients, Typically Developed (TD)

1. INTRODUCTION

In social communication, speech performs an important role to express feelings, emotions and thoughts. Communication impairments affect the process of initial cognitive development which continues to further stages. Neurodevelopment disorders, specifically, Intellectual Disability, Autism, Stuttering and Down Syndrome directly affect the speech and language development. Intellectually disabled children are highly prone to developing some speech, vocal or language disability, which has adverse effect on language development.

According to (World Health Organization, 1980), children with intellectual disability usually have impairments in language and speech which further classifies in use and comprehension of the language, deficits of the linguistic function, learning disabilities and impairments of communication. ID children also face difficulties in speech production, voice functions and speech content. The disorder

DOI: 10.4018/IJCINI.2020040102

This article, originally published under IGI Global's copyright on April 1, 2020 will proceed with publication as an Open Access article starting on February 1, 2021 in the gold Open Access journal, International Journal of Cognitive Informatics and Natural Intelligence (converted to gold Open Access January 1, 2021), and will be distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

causes restrictions in effective verbal communication which affect social involvement, educational attainment or professional performance, independently or in a group.

Communication is a vital component of quality of life, and speech considered as the primary mode of human communication out of various communication medium. In intellectually disabled (ID) children, the speech communication frequently disturbed by impairments in voice production and weakened hearing (Roberts, 2007), resulting in lack of communication and impairing both adaptive behaviour and intellectual functioning.

A authenticate clinical outcome shows the fact that a typically developing child achieve a remarkably good lexical and verbal system from their environment (De Villiers JG and De Villiers PA, 1974; Slobin, 1973; Pinker, 1994; Culter and Klein, 2005) but an ID child can't do the same at a similar age. Speech impairment is among the frequently addressed issue with the children with intellectual disability (Kumin, 2006). First, 3-5 years of a child's lifespan are highly crucial time for speech and language development. Typically developing children can learn the basic structure of language till this age (Tager-Flusburg, 1998) while children with ID can only acquire limited language components and understating of some common words. Researches on speech and language development are also affected due to less number of the speech pathologist and analyst for analysing recorded speech samples (Oller, 1980; Sheinkopf et al., 2000; Wetherby et al., 2004).

Intellectually disabled children participate in fewer activities in and outside school than their control peer groups (Abells, Burbidge & Minnes, 2008). Involvement in outside school activities is essential for all children as it enhances physical and mental growth, social awareness, and psychological security and creates community relationships (Murphy & Carbone, 2008). Children heaving intellectual disability may have lower communication, motor and social skills and have reduced cognitive functioning than their typically developed peers (Pratt & Greydanus, 2007; Westendrop, Houwen, Hartman, & Visscher, 2011). Due to less participation in outside school activities, children with intellectual disability may lead to social isolation (Rimmer, Rowland, & Yamaki, 2007) and inactive behaviour (Frey & Chow, 2006) which is generally reported in the literature.

Speech intelligibility is measured as how clear a person speaks so that his or her voice is understandable to the listener (Leddy, 1999). Lesser speech intelligibility often leads to lack of interest by others, frustration and misunderstanding. Social interaction helps in acquiring speech and language capabilities. Speech intelligibility is generally considered as voice pattern recognition in social communication (Smith, 1985). Although it is a combined effort of both speaker and listener who are participating in the recording process (Tjaden, 1995). Acquaintance with the speaker is especially helpful in case of impaired speech (Kent, 1993). Factor affecting speech intelligibility are the context of communication (e.g., verbal transmission of the voice signal, description of visual cues and related support for the signal to be transmitted) and the spoken material (e.g., length of the message and its linguistic assembly).

The children with Intellectual Disability (with mild, moderate and severe population) deficits in language development (Cardoso-Martins & Mervis, 1985; Mervis & Bertrand, 1990) in their speech. These kids also show many articulatory limitations (Stoel-Gammon, 2001) and exhibit deficits in syntax and grammar development as compared to typically developed (Singer, Bellugi, Bates, Jones & Rossen, 1997). Many neuro-anatomically and neuroimaging research studies refer to linguistic and phonological characteristics of speech of typically developing children, adults and children with Intellectual disability (Fowler, 1990). All the speech-related studies till now consider spectral acoustic features to investigate speech development (Assmann, P. F., Nearey, T. M., & Bharadwaj, 2013; Ballard, Kirrie J., et al., 2012).

In the literature, different feature extraction and classification algorithms have been implemented for classifying children with Intellectual Disability and its sub-groups (mild, moderate, severe and profound). Researches extracted MFCC and LPCC features for the classification in stuttering, Down syndrome and other language disabilities. In this research, different parameter's values of MFCC and LPC features were taken to calculate accuracy from the speech to classify children with ID from TD.

Four different classifiers, k-nearest neighbour (k-NN), support vector machine (SVM), radial basis function network (RBFNN) and linear discriminant analysis (LDA) were employed, for classifying the speech of ID children from the controls. A 10-fold cross validation technique was used to validate the reliability of the classification models. In future, we will also concentrate on the classification of the intellectual disability as mild, moderate and profound at the early age using speech.

2. METHODOLOGY

In this work, MFCC and LPC based feature extraction algorithms were used. Four classifiers, k-NN, SVM, RBFNN and LDA, were applied to assess the effectiveness of the speech features in the classification of intellectually disabled. The effectiveness was also computed on different values of the two parameters of both MFCC and LPC. The reliability of the all the classifiers were computed using the same database. The block diagram of the procedure has shown in Figure 1.

2.1. Speech Dataset

The dataset was created from an institute named SIRTAR (State Institute of rehabilitation, training, and research) at Rohtak – a government of Haryana, India over a duration of 8 months. The dataset comprised of recordings for monologues, picture naming, reading and repeating sentences. There were 96 different recording. The dataset consisted of 48 samples from each ID (mild and moderate) and TD group respectively for analysis. The distribution of the participants has shown in Table 1.

Figure 1. Block diagram of speech pattern recognition

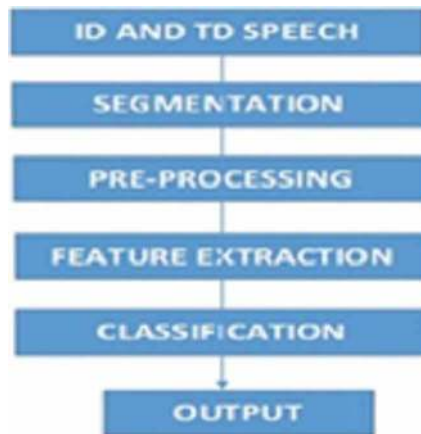


Table 1. Age and gender distribution of participants involved in the dataset

| Type | | Mean Age | Gender | | No. of Participants |
|------|---------------|----------|--------|-------|---------------------|
| | | | Boys | Girls | |
| ID | ID – Mild | 13.2 | 24 | 4 | 28 |
| | ID – Moderate | 14.6 | 14 | 6 | 20 |
| | ID – Profound | 0 | 0 | 0 | 0 |
| TD | TD | 14.2 | 36 | 12 | 48 |

The content of the speech recordings were vowels, pictures of animals, fruits and vegetables and imitation of two letters word of both Hindi and English languages. The participants for the TD group were chosen from The NorthCap University. None of the TD participants was suffered from any neurodevelopmental disorder, socio-communication and speech disabilities. All the participants signed the consent form before the participation in the research. This research is jointly approved by the ethical committee of SIRTAR, Rohtak and The NorthCap University, Gurgaon, India.

2.2. Procedure

Speech samples were recorded on the sampling rate of 22.5 kHz with 16-bit PCM (Pulse-code modulation). Participant's voice was recorded with a head fitted unidirectional Sony microphone (ICD-UX533F) with dynamic gain control. An instructor from the same school was present in the room while recording, to make the participant comfortable. Repetition was there to avoid the hesitation in the participant's voice. The recording procedure consisted of three tasks:

1. **Sustained phonation task:** Participants were asked to utter Hindi and English vowels/alphabet and to count for about 20 seconds;
2. **Imitation task:** Participants were invited to imitate the researcher voice which was mainly consisted of two to three words sentences of Hindi and English. This task was carried out to investigate speed and articulatory correctness;
3. **Picture naming task:** Picture naming task was comprised of the pictures of standard colors, vegetables, birds and animals. The participants were asked to recognize and pronounce the name of the picture showed. This task records various dimensions of voice quality, speaking rate and voice intensity with duration of pauses, duration of spoken syllable and sentence and voice onset time.

If any participant could not manage with the phonation interval, he/she was asked to stop pronouncing. Speech recordings were followed by enlightenment to the speaker about the recording procedure to avoid the psychological stimuli in the speaker's voice.

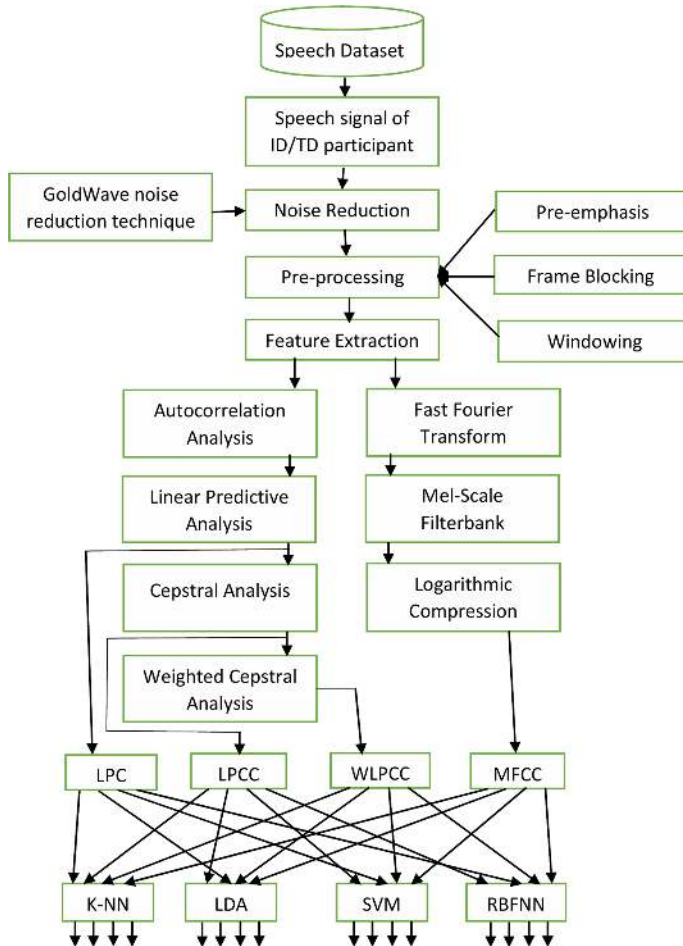
2.3. Speech Signal Pre-Processing

Speech samples were initially recorded at the sampling rate of 44.1 kHz. For further analysis, the recorded signal was then downsampled to 16 kHz because most prominent speech features are lies within 8 kHz range (Huang, Acero & Hon, 2001). Before feature extraction, the speech signal was pre-emphasized, it is necessary to even the spectral energy envelope by increasing the importance of high-frequency components and for eliminating the DC part in the speech signal. The downsampled speech signals were pre-emphasized with a first-order bandpass filter. The z-transform of the first-order bandpass filter is shown in Equation 1:

$$H(z) = 1 - \alpha * z^{-1} \quad (1)$$

For fixed-point implementation the value of $\alpha = 15/16 = 0.9$ is used (M. Hariharan, L. Sin Chee & S. Yaacob, 2012). Hamming window was used to divide a pre-emphasized signal into short frame segment as described in Figure 3. The frame length was varied from 10-50 ms (Aggarwal, G., & Singh, L. 2018). Acoustic features were extracted from each frame and used in the classification. The detailed system architecture is described in Figure 2.

Figure 2. Detailed workflow of the proposed system



3. SPEECH PARAMETERIZATION

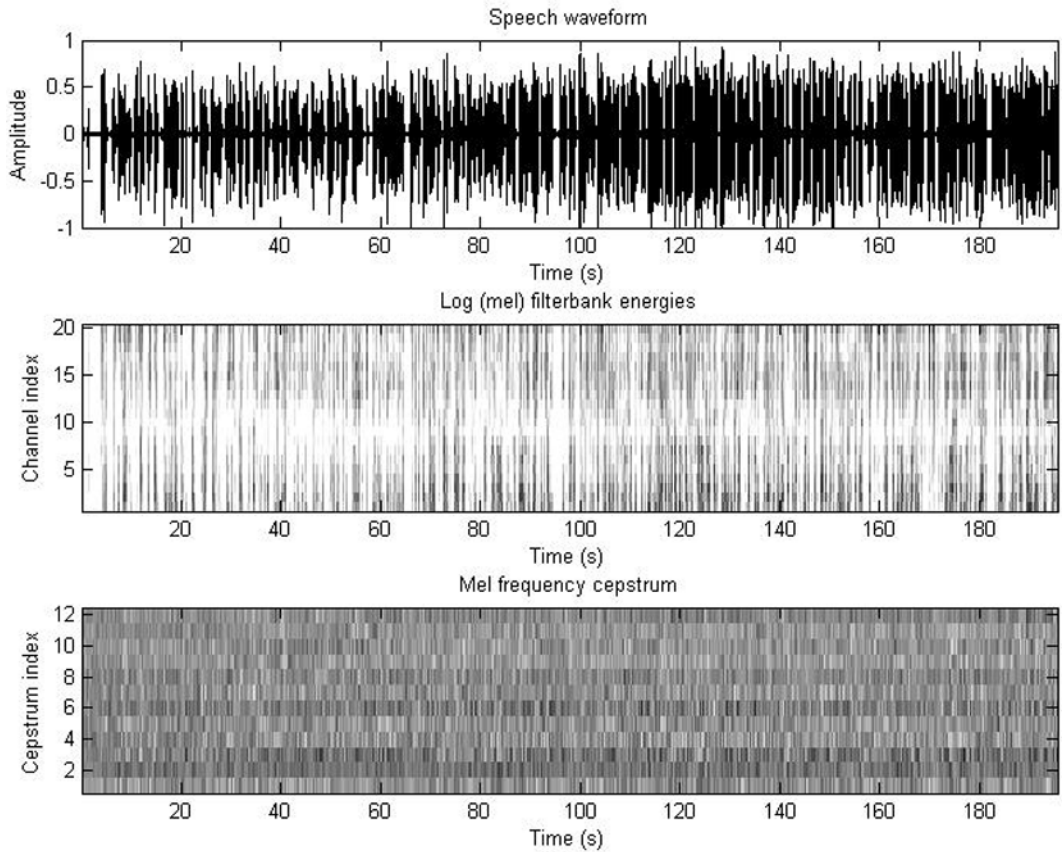
3.1. Mel-Frequency Cepstral Coefficients (MFCC)

MFCCs have recognized as the most successful feature extraction methods in classification of disordered speech (Ravikumar, Rajagopal & Nagaraj, 2009). It is the approximate representation of the human vocal tract system [Davis and Mermelstein 1980] and precisely describes the shape of the auditory system by the envelope of a short time power spectrum.

MFCC is estimated by applying the Fast Fourier Transform (FFT) to the windowed signal which transforms the speech signal into short frames. Power spectrum density is computed for each subframe. Then apply Mel filterbank to the power spectra using a triangular bandpass filter called Mel-scale filter (Dhanalakshmi et al. 2009), add the energies of each filterbank and take the logarithm of all energies. The representation of linear frequency mapping to Mel frequency is shown in the Equation 2:

$$\text{Mel}(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (2)$$

Figure 3a. MFCC representation: TD child



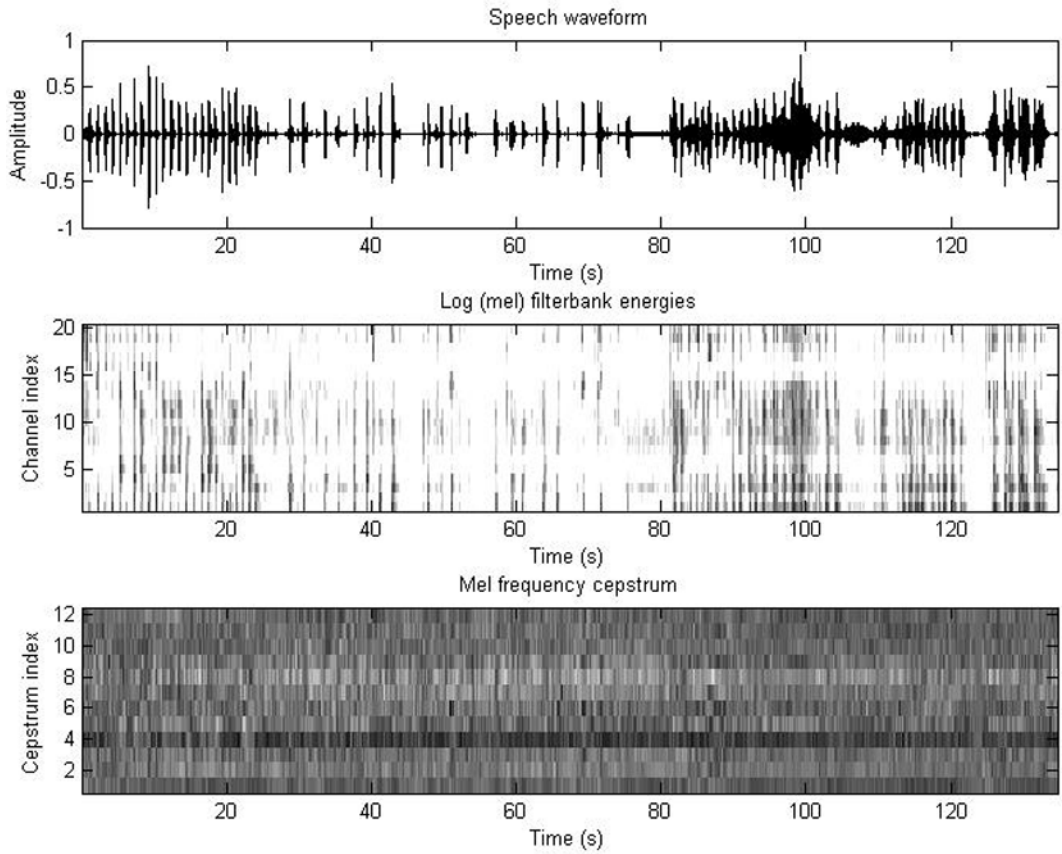
In the end, estimate the discrete cosine transform (DCT) of logarithmic energies converting log Mel-spectrum into time. It results in Mel-Frequency Cepstral Coefficients (MFCC) (Jothilakshmi et al., 2009). In speech related researches, speech samples can be demonstrated as the convolution of the source excitation and vocal tract filter. The deconvolution can be achieved using cepstral analysis of the two components.

Many speech recognition studies reject the 0th coefficients of MFCC due to its irregularity. Regardless of the fact, in this study, the 0th coefficient can be assessed as the average energies of every filter bank in the speech signal that is being evaluated. (Picone, 1993). In this study, the impact of different frame lengths (10-50 ms) were examined over the number of filterbanks (13, 15, 17 and 24). Parameters like frameshift = 20 ms, $\alpha=0.97$, number of filterbank channels = 20, cepstral sine lifter parameter = 22 and frequency limit = 300-3700 Hz were used for calculations. Figure 3 (a) and (b) represents the MFCC features of an ID child and a TD child of similar age.

3.2. Linear Prediction Analysis

In this study, three salient features of speech were extracted, Linear Predictive Coding (LPC), Linear Predictive Cepstral Coefficients (LPCC) and Weighted Linear Predictive Cepstral Coefficients (WLPCC). In linear predictive analysis, the next sample is estimated through the linear combination of past p samples where p is the order of prediction. If $s(n)$ is the current sample then it is calculated using past p samples as Equation 3:

Figure 3b. MFCC representation: ID child



$$\hat{s}(n) = \sum_{m=1}^p a_m s(n-m) \tag{3}$$

The difference between actual and the calculated sample is called prediction error $e(n)$. It is defined as Equation 4:

$$e(n) = s(n) - \hat{s}(n) \quad e(n) = s(n) - \sum_{m=1}^p a_m s(n-m) \tag{4}$$

where, a_m stands for linear predictive coefficients, Autocorrelation is applied on each of the windowed signal using Equation 5:

$$\gamma(m) = \sum_{n=0}^{N-1-m} x(n)x(n+m), \quad m = 0, 1..p \tag{5}$$

solve both nch pre –

processing and parametrization accuracy of normal speech at early ages.

where p is the order of the LPC analysis. In literature, the value of p is fixed as 2, 8, 10 and 14 (Aggarwal & Singh, 2018; Hariharan, Chee & Yaacob, 2012; Ai, Hariharan, Yaacob, & Chee, 2012). LPC analysis is implemented using Durbin-Levinson recursive algorithm for converting autocorrelation coefficients into LPC coefficients. The final LPC coefficients is given by Equation 6:

$$a_j = a_j^p, 1 \leq j \leq p \quad (6)$$

The cepstral coefficients can be calculated using LPC coefficients using the following recursive method shown in Equation 7 and 8:

$$c_m = a_m + \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k}, 1 \leq m \leq p \quad (7)$$

$$c_m = \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k}, m > p \quad (8)$$

where, m stands for the order of the cepstral coefficients, LPCC features are obtained. Although, the higher order cepstral coefficients are prone to noise like variability and the lower order cepstral coefficients are sensitive to the overall spectral shape (Rabiner & Juang, 1993). To minimize the sensitivities, a standard methodology is to weigh the cepstral coefficients by a tapered window is applied.

Weighted Linear Predictive Cepstrum Coefficients can be determined by multiplying LPCC to the weighted formula, explained in Equation 9:

$$w_m = \left[1 + \frac{Q}{2} \sin\left(\frac{m\pi}{Q}\right) \right], 1 \leq m \leq Q \quad (9)$$

In the cepstral region, a weighted function can be represented as bandpass filter to de-emphasize c_m around $m=1$ to Q . Adding to it, the sensitivity of higher order cepstral coefficients onto overall spectral scope at $m=1$ and lower order cepstral coefficients to noise at $m=Q$ are reduced by the filter. WLPCCC, \widehat{c}_m were calculated by Equation 10:

$$\widehat{c}_m = c_m * w_m, 1 \leq m \leq Q \quad (10)$$

In earlier studies, WLPCCC has been implemented as a technique to extract speech features in different applications with different neuro-development disorders. In this work, three linear prediction techniques were applied for speech feature extraction to differentiate between the speech samples of ID and TD children. The classification was dependent on the value of order p . Order p from 2 to 14 have been applied with 8 kHz sampling frequency (Aggarwal, G., & Singh, L. 2018). The impact of order p (2, 8, 10 and 14) was analysed with different frame length (10, 20, 30, 40 and 50 ms). Non-

overlapping window with $\alpha = 0.97$, 256-point FFT was applied for calculations. LPC representations of age matched TD and ID children are described in Figure 4 (a) and (b). The results of the analysis will be discussed in detail in the Result section.

4. CLASSIFICATION

4.1. k-Nearest Neighbor

k-NN is a simple classification model used lazy learning. In speech pattern recognition, the k-NN is used to classify objects depends on the nearest training examples using distance formulas in the feature space. It is supervised learning model calculate the majority votes and classify the object in the class similar to its k nearest neighbour (k stands for a positive integer) Minimum distance between the training set and the test feature vector is calculated to decide the k-NN category. In this work, for each test sample, the minimum distance was calculated from the test speech sample to each of the training speech samples and found the k-NN category of the training set. Euclidean distance matrix $d_E(x,y)$ was used to compute the nearest neighbour between training and testing data, where x and y are the training and testing speech samples composed for N features, described in Equation 11:

$$\begin{aligned} x &= (x_1, x_2, \dots, x_N) \\ y &= (y_1, y_2, \dots, y_N) \end{aligned} \tag{11}$$

Figure 4a. LPC parameterization: TD child

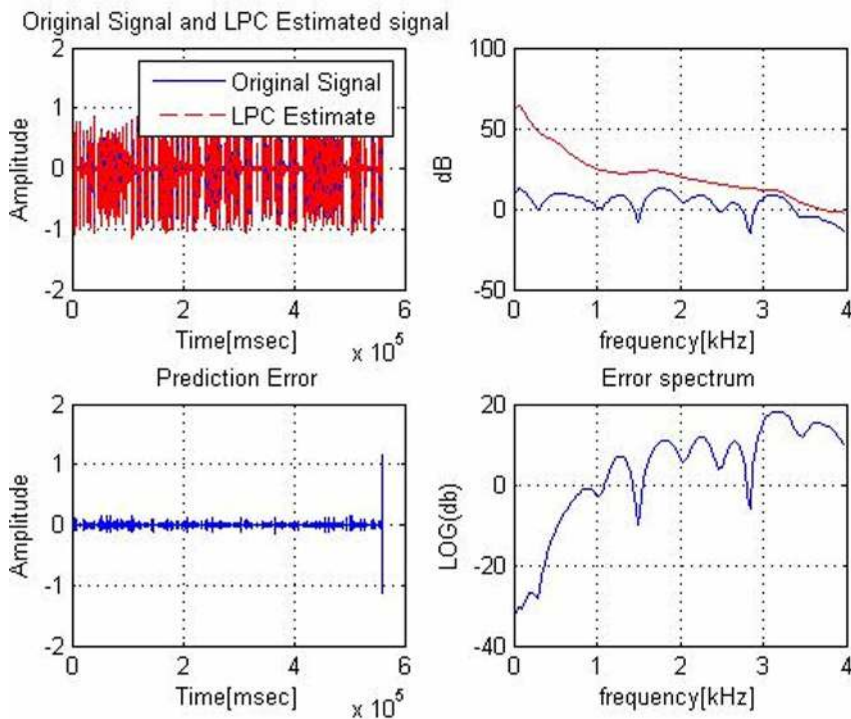
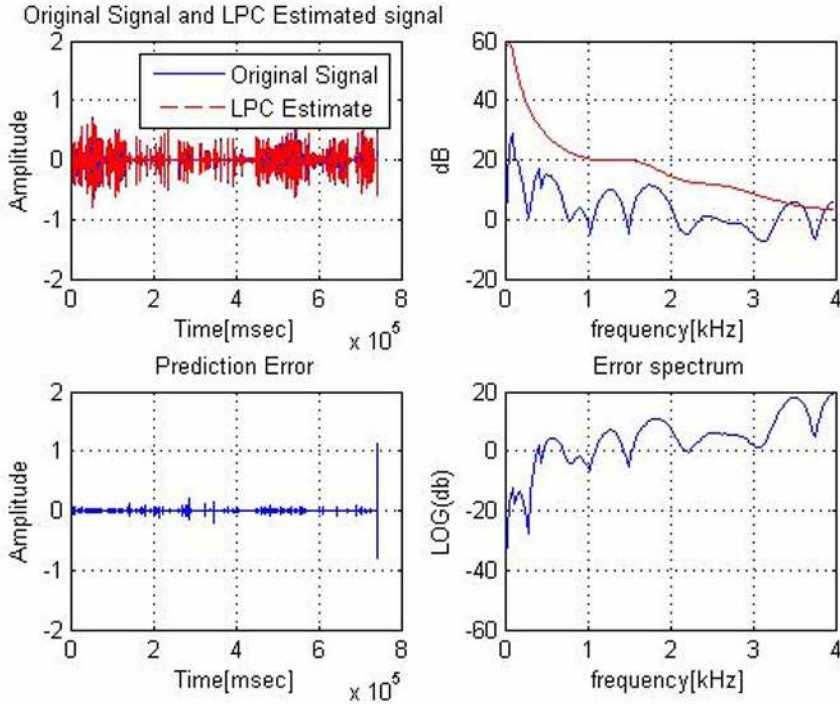


Figure 4b. LPC parameterization: ID child



Euclidean distance is defined in Equation 12:

$$d_E(x, y) = \sqrt{\sum_{i=1}^N (x_i^2 - y_i^2)} \quad (12)$$

Thus, the value of k depicts an essential role in the k -NN classifier. Usually, the k value should be given in advance, and the optimum k -value depends on the dataset. Generally, more significant the k value less will be the effect of noise on the classification but the classes boundaries become close to each other (Liu, Lee, & Lin, 2010).

Therefore, in this research, each experiment was performed with different values of k ranging from 1 to 10. For each k value, the experiment was iterated for 10 times with different training and testing sets. The classification results were represented in the form of classification accuracy with standard deviation.

4.2. Support Vector Machine

In this paper, SVM is used as a classification model and it is widely used to solve both linear and non-linear classification problems like density estimation and pattern recognition tasks. It uses a non-linear mapping to translate the training data into a higher dimension using a nonlinear transformation ϕ and then apply a linear separation.

For building a nonlinear SVM classifier, the intermediate product (X, Y) is substituted by a kernel function $K(X, Y)$:

$$f(x) = \text{sign} \left(\sum_{i=1}^L \alpha_i y_i K(x_i, x) + b \right) \quad (13)$$

SVM consists of two layers. In the learning phase, the first layer opts for the base $K(x_i, x)$ where $i=1, 2, \dots, N$, from the bases defined by the kernel. The second layer builds a linear function in the feature space. Constructing the optimal hyperplane in the analogous feature space is the same. Hyperplane with the large margin to be more accurate than the smaller margin for classifying feature data. The hyperplane with maximum margin is interpreted as the shortest distance is equal between the hyperplane and margins on both sides. Equation 14 defined the separating hyperplane:

$$W \cdot X + b = 0 \quad (14)$$

where W is the weight vector $W = (w_1, w_2, \dots, w_n)$, n is the number of attributes and b is the scalar (Suykens, Van Gestel, & De Brabanter, 2002).

In this work, the LS-SVMLab toolbox was applied to classify the speech of ID from TD children. To achieve a better classification accuracy, the two parameters, regularization parameters (γ , gam) and σ^2 (sig^2) which was the squared bandwidth of RBF kernel was chosen optimally.

4.3. Linear Discriminant Analysis

The linear discriminant analysis is a widely used supervised learning algorithm in the area of feature selection and classification. It gives precisely good results in various research fields such as image processing, signal processing, speech processing and pattern recognition and many more. In this work, LDA was applied to classify a two-class problem. The role of LDA is to locate a linear transformation (discriminant function) that gives maximum class separability in small dimensional space (Duda, Hart & Stork, 2000). The discriminant function was generated by transforming the speech feature vector to a projection feature vector. The discriminant function is shown in Equation 15:

$$f_i = \mu_i S_w^{-1} x_k^T - \frac{1}{2} \mu_i S_w^{-1} \mu_i^T + \ln(P_i) \quad (15)$$

where μ_i - mean features in group i and the value of i will either 1 or 2:

x_k - x represents the features of all the speech data, k represents one acoustic feature.

P_i - total samples of each group divided by total samples.

Firstly, the mean μ_3 of the complete dataset and mean μ_1 and μ_2 of each feature are calculated which has shown in Equation 16:

$$\mu_3 = P_1 \mu_1 + P_2 \mu_2 \quad (16)$$

where P_1 and P_2 are the prior probabilities of Class1 and Class2 respectively. Class discrimination is defined by implementing between-class S_b scatter and within-class scatter S_w and S_w is measured using Equation 17:

$$S_w = P_1 cov_1 + P_2 cov_2 \quad (17)$$

Thus, cov_1 and cov_2 should be in symmetry. Covariance can be calculated using the above equation and S_b is calculated using Equation 18 and 19:

$$cov_i = (x_i - \mu_i)(x_i - \mu_i)^T \quad (18)$$

$$S_b = \sum_i (\mu_i - \mu_3)(\mu_i - \mu_3)^T \quad (19)$$

The Linear Discriminant function was calculated using all feature data. Training and testing data were dragged into new dimensions. Classification algorithms were implemented in Matlab R2013b.

4.4. Radial Basis Function Neural Network

The radial basis function neural network (RBFNN) (Haykin, 2001) is primarily used for the time series sequence data with a two-layer feed-forward network architecture includes an input layer, hidden layer and an output layer as shown in Figure. Hidden layer is acted as the layer of processing unit. Radial basis function is applied at input and output layers. The input layer has n_i units for a n_i dimension input vector which is fully connected to n_h hidden layer units, which are further fully connected to n_c output layer units. In this study, Gaussian activation function was chosen for the hidden layers and was characterized by their mean vectors μ_i and covariance matrices $C_i, i = 1, 2, \dots, n_h$. For ease of calculations, it was presumed that the covariance matrices were of the form of $C_i = \sigma_i^2 I, i = 1, 2, \dots, n_h$. The activation function of the i th hidden layer for an x_j input vector, is given by Equation 20:

$$g(x_j) = \exp\left(\frac{-\|x_j - \mu_i\|^2}{2\sigma_i^2}\right) \quad (20)$$

The σ_i^2 and μ_i can be calculated using an appropriate clustering algorithm. In this paper, k-means clustering was used to calculate the centroids.

Here an assumption $\sigma_i^2 = \sigma^2$ was made, to avoid the activation function go beyond the limits, neither too peaked nor too flat. It is defined by Equation 21:

$$\sigma^2 = \frac{\gamma d^2}{2} \quad (21)$$

where d is the maximum distance calculated between two centroid points, γ is the empirical scale factor which helps in smoothening of the mapping function. So, the above Equation 22 is rewritten as:

$$g_i(x_j) = \exp\left(\frac{-\|x_j - \mu_i\|^2}{\gamma d^2}\right) \tag{22}$$

The hidden layer is fully connected with the n_c output layer through weights w_{ik} . The result of the k th output unit for an input unit x_j is defined by Equation 23 and the block diagram is represented in Figure 5:

$$y_k(x_j) = \sum_{n_h}^{i=0} w_{ik} g_i(x_j) \tag{23}$$

where $k = 1, 2, \dots, n_c$, $g_0(x_j) = 1$.

5. RESULTS

MFCC and LPC-based parameterization techniques were used to extract acoustic features from the speech signals. Different numbers of MFCCs filterbanks (13, 15, 17 and 24) and LPC order (2, 8, 10 and 14) were used to characterize the disordered speech of ID children. 10-fold cross-validation was applied to validate the reliability of the classifiers. Four classification models were applied for classifying the speech of ID from TD children. For better accuracy, parameters for all the classifiers were chosen suitably. LDA incorporated the linear discriminant function. In k-NN different ‘k’ values were used from 1 to 10. In SVM optimal values of regularization parameters were taken as 90 and 0.9 respectively. In RBFNN, time-dependent neural network-based activation function is used. The results of MFCCs and LPCs using all the four classifiers are displayed in Table 2 and 3. The average and the standard deviation of the classification accuracies of both ID and TD speech signals are also tabulated. Standard deviation clearly shows the consistency of the classifiers results. If the standard deviation is significant, the classification model is inconsistent, and it also shows that the learning attributes of the model disturb the performance of the classifiers.

It has concluded from Table 2, that MFCCs give more than 96% classification accuracy with 13 MFCCs filterbanks. The best accuracies are calculated for the frame length of 20 ms and 30 ms The

Figure 5. Block diagram of RBFNN

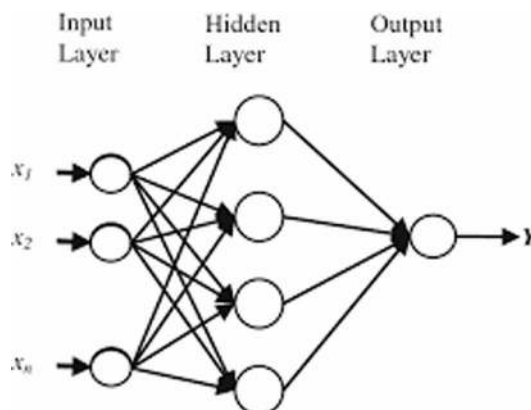


Table 2. Classification accuracies between ID and TD children of four classifiers using MFCC

| No. of Filters | Classifier | Frame Length (ms) | | | | |
|----------------|------------|-------------------|-------------|-------------|-------------|-------------|
| | | 10 | 20 | 30 | 40 | 50 |
| 13 | k-NN | 88.73± 2.14 | 92.23± 1.42 | 91.23± 1.10 | 93.06± 0.76 | 91.11± 1.03 |
| | LDA | 89.31± 1.12 | 89.91± 1.05 | 89.02± 0.98 | 89.91± 0.95 | 87.23± 1.25 |
| | SVM | 95.03± 0.72 | 95.65± 0.92 | 94.22± 0.70 | 94.85± 0.95 | 94.01± 0.70 |
| | RBFNN | 95.03± 0.68 | 96.25± 0.75 | 96.25± 0.62 | 95.12± 0.82 | 93.34± 1.02 |
| 15 | k-NN | 88.52± 1.55 | 91.98± 1.35 | 91.02± 0.77 | 91.88± 0.95 | 91.21± 1.78 |
| | LDA | 88.72± 0.88 | 89.91± 1.12 | 89.54± 0.62 | 90.22± 0.56 | 88.97± 0.65 |
| | SVM | 95.30± 0.65 | 96.18± 0.76 | 95.23± 0.23 | 96.28± 0.50 | 95.91± 0.56 |
| | RBFNN | 95.20± 0.76 | 96.87± 0.65 | 96.13± 0.31 | 96.54± 0.48 | 94.59± 0.56 |
| 17 | k-NN | 88.21± 2.12 | 91.72± 1.12 | 90.14± 0.76 | 92.14± 1.62 | 90.01± 1.86 |
| | LDA | 88.55± 0.95 | 91.12± 1.03 | 90.07± 0.59 | 91.17± 1.10 | 90.33± 0.42 |
| | SVM | 95.18± 0.88 | 95.83± 0.82 | 95.02± 0.76 | 95.08± 0.46 | 94.88± 0.85 |
| | RBFNN | 95.45± 0.98 | 96.55± 0.56 | 95.92± 0.82 | 96.48± 0.76 | 94.32± 0.74 |
| 24 | k-NN | 87.93± 2.58 | 88.21± 1.72 | 90.43± 1.32 | 90.32± 1.84 | 90.17± 1.77 |
| | LDA | 88.12± 1.00 | 88.45± 1.03 | 88.23± 0.76 | 88.78± 1.28 | 88.70± 0.88 |
| | SVM | 94.28± 1.12 | 95.43± 1.14 | 94.55± 0.96 | 93.97± 0.84 | 94.12± 0.98 |
| | RBFNN | 95.03± 1.08 | 95.55± 0.96 | 95.20± 0.84 | 95.33± 0.85 | 93.98± 0.92 |

RBFNN classification model gives higher accuracy than the other three classifiers. Classification results show that RBFNN is most reliable and suitable classifier for the disordered speech classification. From Table 3, it has noticed that the WLPCC features outperform the LPC and LPCC parameters with accuracy more than 96% using frame length, 20 ms and 30 ms. With our previous findings, the LPC orders were chosen (2, 8, 10, and 14). Order of the LPC decides the number of LPC coefficients. It has observed from Table 3 that, on low order value, the LPC spectrum take only the central resonance peak and gives an unfortunate result. When the order of the LPC is higher, it brings many spurious peaks in the LPC spectrum, which also gives poor results. So the appropriate order should be selected for the better classification accuracy. The outcomes for the current study are very encouraging, but it is challenging to compare these results with other researches since most studies are based on the gesture, gross motor skills and daily activities of a child with different databases and approaches.

The speech data is time dependent data, so the classifiers has chosen are good in classification of time variant data. From the literature, we have seen that many research groups, who are using speech as their primary data, suing k-NN, LDA and SVM for classification. RBFNN is the advanced version of ANN which can correctly classify the time dependent data.

6. CONCLUSION

In this paper, a comparison of two speech parametrization techniques is presented for the classification of ID and TD children. Four classifiers (k-NN, LDA, SVM AND RBFNN) are used. Both types of speech samples are undergoing feature extraction using LPC based parameterization and MFCCs method. In LPC based feature extraction, the order was varied with the frame length, and in MFCC, the number of filterbanks are varied with different frame length and analysed

Table 3. Classification accuracies between ID and TD children of four classifiers using LPC, LPCC and WLPCC parameterization methods

| Frame Length | Order Classifier | 2 | | | 8 | | | 10 | | | 14 | | |
|--------------|------------------|------------|------------|------------|------------|------------|-------------|------------|------------|------------|------------|------------|------------|
| | | LPC | LPCC | WLPCC | LPC | LPCC | WLPCC | LPC | LPCC | WLPCC | LPC | LPCC | WLPCC |
| 10 | k-NN | 83.78±1.52 | 87.54±1.43 | 85.76±1.78 | 90.76±1.77 | 88.13±2.12 | 89.73±1.04 | 88.73±1.67 | 87.27±2.13 | 90.69±1.78 | 85.77±1.64 | 89.88±1.76 | 90.19±1.22 |
| | LDA | 64.31±1.31 | 77.14±1.10 | 63.98±1.67 | 79.39±0.88 | 85.23±0.74 | 87.98±0.54 | 78.37±1.23 | 89.16±1.21 | 88.19±0.93 | 73.69±1.35 | 75.12±1.23 | 88.12±0.72 |
| | SVM | 84.99±0.37 | 89.87±0.71 | 85.54±0.56 | 91.34±0.67 | 92.56±0.33 | 92.76±0.71 | 91.83±0.54 | 91.67±0.87 | 92.66±0.59 | 89.88±0.98 | 94/26±0.87 | 93.66±0.65 |
| | RBFFNN | 86.11±0.41 | 90.67±0.69 | 86.61±0.54 | 90.87±0.96 | 93.76±0.39 | 93.66±0.46 | 92.56±0.48 | 92.37±0.38 | 93.88±0.76 | 91.03±0.77 | 95.11±0.48 | 94.76±0.53 |
| | k-NN | 83.76±1.23 | 87.76±1.87 | 87.13±1.21 | 87.35±0.84 | 88.02±1.82 | 90.67±1.41 | 87.33±1.23 | 88.02±2.74 | 90.23±1.25 | 83.21±1.98 | 88.84±1.08 | 88.47±0.73 |
| | LDA | 65.67±1.69 | 78/73±0.58 | 67.83±1.24 | 79.39±0.97 | 85.98±0.82 | 87.88±0.98 | 77.69±1.63 | 88.98±1.48 | 89.04±0.82 | 73.47±1.88 | 88.98±1.62 | 88.12±0.88 |
| 20 | SVM | 87.46±0.75 | 88.64±0.58 | 85.92±0.71 | 90.73±0.78 | 92.48±0.20 | 93.23±0.68 | 91.09±0.93 | 92.59±0.68 | 93.49±0.89 | 87.92±0.82 | 93.30±0.68 | 93.09±0.89 |
| | RBFFNN | 86.58±0.57 | 88.68±0.39 | 86.38±0.73 | 91.24±0.65 | 93.84±0.35 | 94.02±0.72 | 92.79±0.69 | 94.23±0.62 | 94.92±0.83 | 88.76±0.78 | 94.50±0.58 | 94.21±0.83 |
| | k-NN | 83.79±1.90 | 88.76±1.28 | 86.93±1.33 | 90.86±0.97 | 88.51±1.64 | 90.16±1.005 | 88.73±1.43 | 87.89±2.14 | 92.08±1.73 | 83.24±1.40 | 90.14±1.50 | 89.67±1.05 |
| | LDA | 65.18±1.23 | 78.11±0.59 | 65.66±1.10 | 81.37±0.69 | 86.97±1.15 | 87.53±1.14 | 80.27±1.12 | 87.73±0.83 | 89.67±1.18 | 75.34±1.43 | 90.56±1.35 | 89.48±0.53 |
| | SVM | 86.24±0.86 | 89.86±0.24 | 85.33±0.62 | 89.11±0.92 | 92.81±0.48 | 92.87±0.54 | 91.18±1.28 | 90.19±1.11 | 92.29±0.57 | 86.98±1.21 | 94.49±0.68 | 92.98±0.82 |
| | RBFFNN | 87.39±0.56 | 90.06±0.39 | 85.44±0.61 | 89.71±0.84 | 93.58±0.51 | 93.39±0.41 | 92.49±1.04 | 91.53±0.91 | 94.51±0.55 | 88.24±0.88 | 93.22±0.52 | 93.66±0.56 |
| 30 | k-NN | 83.51±0.61 | 88.74±1.57 | 87.31±1.46 | 89.96±1.48 | 86.68±1.28 | 89.34±1.27 | 88.72±1.59 | 87.14±2.41 | 90.88±1.76 | 84.31±2.66 | 88.91±1.29 | 89.24±0.82 |
| | LDA | 66.24±0.92 | 79.29±0.94 | 64.83±1.68 | 78.92±0.94 | 84.92±1.84 | 87.21±1.44 | 78.60±1.23 | 89.03±1.14 | 89.22±0.86 | 72.82±1.38 | 88.23±0.83 | 87.72±0.80 |
| | SVM | 85.34±0.73 | 90.05±0.66 | 87.14±0.76 | 90.16±1.12 | 92.13±0.66 | 91.88±0.38 | 90.21±0.49 | 91.93±0.88 | 92.88±0.82 | 87.82±1.20 | 92.88±0.72 | 92.78±0.82 |
| | RBFFNN | 87.67±0.42 | 91.38±0.49 | 87.21±0.38 | 91.34±1.04 | 93.42±0.46 | 93.32±0.32 | 90.63±0.39 | 92.96±0.45 | 93.55±0.76 | 88.03±0.98 | 93.01±0.77 | 93.00±0.64 |
| | k-NN | 82.16±2.23 | 88.61±0.85 | 86.58±1.48 | 89.83±0.78 | 87.38±1.27 | 90.02±1.62 | 86.67±1.16 | 85.80±2.34 | 91.35±2.95 | 81.53±2.03 | 89.23±0.83 | 88.79±0.81 |
| | LDA | 65.34±1.86 | 75.59±1.23 | 66.28±0.66 | 77.29±0.81 | 85.81±0.94 | 87.28±1.19 | 72.42±1.96 | 86.49±0.83 | 89.38±2.73 | 71.93±1.31 | 89.27±0.81 | 89.27±0.61 |
| 50 | SVM | 85.20±1.09 | 90.16±0.46 | 87.32±0.58 | 88.79±0.92 | 92.66±0.30 | 92.54±0.53 | 86.02±1.18 | 91.78±1.41 | 93.21±2.11 | 85.94±1.32 | 92.76±0.60 | 93.02±1.82 |
| | RBFFNN | 86.06±0.69 | 91.98±0.52 | 88.02±0.48 | 89.12±1.05 | 93.12±0.41 | 92.96±0.40 | 86.32±0.92 | 93.03±0.93 | 93.83±1.94 | 86.28±1.12 | 93.01±0.95 | 93.12±1.21 |

Maximum accuracies of more than 94% and 96% were investigated in LPCs and MFCCs respectively.

the effect on the classification models. The RBFNN outperform SVM followed by k-NN and LDA. RBFNN gives the highest average classification accuracy of above 94% using WLPCC and 96% using MFCC features. These outcomes indicate that the proposed technique could be used as a significant tool in intellectual disability assessment for the speech pathologist. In the future, more feature selection algorithms will be implemented to select the best features. Other classifiers may also be used to improve the classification outcomes between disordered and normal speech at early ages.

REFERENCES

- Abells, D., Burbidge, J., & Minnes, P. (2008). Involvement of adolescents with intellectual disabilities in social and recreational activities. *Journal on Developmental Disabilities, 14*, 88–94.
- Aggarwal, G., & Singh, L. (2018). Classification of intellectual disability using LPC, LPCC, and WLPC parameterization techniques. *International Journal of Computers and Applications, 1*–10. doi:10.1080/1206212X.2018.1475330
- Ai, O. C., Hariharan, M., Yaacob, S., & Chee, L. S. (2012). Classification of speech dysfluencies with MFCC and LPCC features. *Expert Systems with Applications, 39*(2), 2157–2165. doi:10.1016/j.eswa.2011.07.065
- Assmann, P. F., Nearey, T. M., & Bharadwaj, S. V. (2013). Developmental patterns in children's speech: Patterns of spectral change in vowels. In *Vowel inherent spectral change* (pp. 199–230). Berlin: Springer. doi:10.1007/978-3-642-14209-3_9
- Ballard, K. J., Djaja, D., Arciuli, J., James, D. G., & van Doorn, J. (2012). Developmental trajectory for production of prosody: Lexical stress contrastivity in children ages 3 to 7 years and in adults. *Journal of Speech, Language, and Hearing Research, 55*(6), 1822–1835. doi:10.1044/1092-4388(2012/11-0257) PMID:22473835
- Cardoso-Martins, C., & Mervis, C. B. (1985). Maternal speech to prelinguistic children with Down syndrome. *American Journal of Mental Deficiency, 31*58199
- Cutler, A., Klein, W., & Levinson, S. C. (2005). *The cornerstones of twenty-first century psycholinguistics. In Twenty-First Century Psycholinguistics: Four Cornerstones* (pp. 1–20). Mahwah, NJ: Erlbaum.
- De Villiers, J. G., & De Villiers, P. A. (1974). Competence and performance in child language: Are children really competent to judge? *Journal of Child Language, 1*(1), 11–22. doi:10.1017/S0305000900000052
- Dhanalakshmi, P., Palanivel, S., & Ramalingam, V. (2009). Classification of audio signals using SVM and RBFNN. *Expert Systems with Applications, 36*(3 Part 2), 6069–6075. doi:10.1016/j.eswa.2008.06.126
- Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.
- Fowler, A. E. (1990). Language abilities in children with Down syndrome: Evidence for a specific syntactic delay. *Children with Down syndrome: A developmental perspective, 9*, 302-328.
- Hariharan, M., Chee, L. S., Ai, O. C., & Yaacob, S. (2012). Classification of speech dysfluencies using LPC based parameterization techniques. *Journal of Medical Systems, 36*(3), 1821–1830. doi:10.1007/s10916-010-9641-6 PMID:21249515
- Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.
- Huang, X., Acero, A., Hon, H. W., & Reddy, R. (2001). *Spoken language processing: A guide to theory, algorithm, and system development* (Vol. 1). Upper Saddle River, NJ: Prentice hall PTR.
- Jothilakshmi, S., Ramalingam, S., & Palanivel, S. (2009). Unsupervised speaker segmentation with residual phase and MFCC feature. *Expert Systems with Applications, 36*(6), 9799–9804. doi:10.1016/j.eswa.2009.02.040
- Suykens, J. A., Van Gestel, T., & De Brabanter, J. (2002). *Least squares support vector machines*. World Scientific.
- Kent, R. D., Kaiser, A. P., & Gray, D. B. (1993). Speech intelligibility and communicative competence in children. In *Enhancing children's communication: Research foundations for intervention* (pp. 223-239). Academic Press.
- Kumin, L. (2006). Speech intelligibility and childhood verbal apraxia in children with Down syndrome. *Down's Syndrome: Research and Practice, 10*(1), 10–22. doi:10.3104/reports.301 PMID:16869369
- Leddy, M. (1999). The biological bases of speech in people with Down syndrome. In *Improving the communication of people with Down syndrome* (pp. 61-80). Academic Press.
- Liu, C. L., Lee, C. H., & Lin, P. M. (2010). A fall detection system using k-nearest neighbor classifier. *Expert Systems with Applications, 37*(10), 7174–7181. doi:10.1016/j.eswa.2010.04.014
- Mervis, C. B. (1997). Developmental relations between cognition and language: Evidence from Williams syndrome. In *Research on communication and language disorders: Contributions to theories of language development* (pp. 75-100). Academic Press.

- Murphy, N., & Carbone, P. (2008). Promoting the participation of children with disabilities in sports, recreation and physical activities. *Pediatrics*, *34*, 245–256. PMID:18450913
- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In *Child Phonology, Production* (Vol. 1, pp 93–112) New York: Academic Press. doi:10.1016/B978-0-12-770601-6.50011-5
- Picone, J. W. (1993). Signal modelling techniques in speech recognition. *Proceedings of the IEEE*, *81*(9), 1215–1247. doi:10.1109/5.237532
- Pinker, S. (1994). *The Language Instinct*. New York: Harper Perennial. doi:10.1037/e412952005-009
- Pratt, H. D., & Greydanus, D. E. (2007). Intellectual disability (mental retardation) in children and adolescents. *Primary Care: Clinics in Office Practice*, *34*(2), 375–386. doi:10.1016/j.pop.2007.04.010 PMID:17666233
- Rabiner, L. R., & Juang, B. H. (1993). *Fundamentals of speech recognition* (Vol. 14). Englewood Cliffs: PTR Prentice Hall.
- Ravikumar, K. M., Rajagopal, R., & Nagaraj, H. C. (2009). An approach for objective assessment of stuttered speech using MFCC features. *ICGST International Journal on Digital Signal Processing, DSP*, *9*(1), 19–24.
- Rimmer, J. H., Rowland, J. L., & Yamaki, K. (2007). Obesity and secondary conditions in adolescents with disabilities: Addressing the needs of an underserved population. *The Journal of Adolescent Health*, *41*(3), 224–229. doi:10.1016/j.jadohealth.2007.05.005 PMID:17707291
- Roberts, J. E., Price, J., & Malkin, C. (2007). Language and communication development in Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*(1), 26–35. doi:10.1002/mrdd.20136 PMID:17326116
- Sheinkopf, S. J., Mundy, P., Oller, D. K., & Steffens, M. (2000). Vocal atypicalities of preverbal autistic children. *Journal of Autism and Developmental Disorders*, *30*(4), 345–354. doi:10.1023/A:1005531501155 PMID:11039860
- Singer Harris, N. G., Bellugi, U., Bates, E., Jones, W., & Rossen, M. (1997). Contrasting profiles of language development in children with Williams and Down syndromes. *Developmental Neuropsychology*, *13*(3), 345–370. doi:10.1080/87565649709540683
- Slobin, D. (1973). Cognitive prerequisites for the development. In C. Ferguson & D. Slohin (Eds.), *Studies in Child Language Development* (pp. 175–208). Academic Press.
- Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes*, *4*(3), 333–342. doi:10.1111/j.1467-971X.1985.tb00423.x
- Steven, B., Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transaction ASSP*, *28*(4), 357–366.
- Stoel-Gammon, C. (2001). Down syndrome phonology: Developmental patterns and intervention strategies. *Down's Syndrome: Research and Practice*, *7*(3), 93–100. doi:10.3104/reviews.118 PMID:11721538
- Tager-Flusberg, H., & Sullivan, K. (1998). Early language development in children with mental retardation. In J. Burack, R. Hodapp, & E. Zigler (Eds.), *Handbook of Mental Retardation and Development* (pp. 208–239). New York: Cambridge University Press.
- Tjaden, K. K., & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics*, *9*(2), 139–154. doi:10.3109/02699209508985329
- Westendorp, M., Houwen, S., Hartman, E., & Visscher, C. (2011). Are gross motor skills and sports participation related in children with intellectual disabilities? *Research in Developmental Disabilities*, *32*(3), 1147–1153. doi:10.1016/j.ridd.2011.01.009 PMID:21310587
- Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. [PubMed]. *Journal of Autism and Developmental Disorders*, *34*(5), 473–493. doi:10.1007/s10803-004-2544-y PMID:15628603
- World Health Organization. (1980). International classification of impairments, disabilities, and handicaps: a manual of classification relating to the consequences of disease, published in accordance with resolution WHA29.35 of the Twenty-ninth World Health Assembly.

Gaurav Aggarwal received his B.Tech degree in Instrumentation from University Science and Instrumentation Centre, Kurukshetra University, India in 2006. He received his M.Tech degree in Computer Science and Engineering from Department of Computer Science and Application from Kurukshetra University, India in 2008. He has completed his PhD from The NorthCap University, Gurgaon, India in the area of machine learning. He is working as Assistant Professor for department of Information Technology in Manipal University Jaipur, India. His area of interests includes signal processing, machine learning, and cognitive sciences. He has published several technical papers and reports in the above research areas at national and International platform.

Latika Singh is an Associate Professor and chair of the computer science engineering and information technology department at The NorthCap University since 2012. She received her Bachelor's degree from Kurukshetra University and master's from GJU. She received her PhD from National Brain Research Institute in computational sciences in 2007. Her research interests center in studying the speech disorders in children with neurodevelopmental disorders mainly through the application of digital signal processing, data-mining, statistics and modelling. She has published several research papers in peer-reviewed journals of Elsevier, Blackwell, etc., and conferences. She has been member of several technical programme committees of International conferences. She was convener of IEEE conference IACC2014 where several delegates from India and abroad participated. She has also worked on two research projects sponsored by the Govt of India.