Optimal Curve Fitting of Speech Signal for Disabled Children

Anandthirtha. B. GUDI[1], and H. C. Nagaraj[2]

[1] *Department of Electronics and Communication Engineering,*

Sri Bhagawan Mahaveer Jain College of Engineering,

Bangalore-562112, Karnataka, India

Phone: 0091 80 27577200, Fax: 0091 80 27577199

gudi_anand@rediffmail.com

[2] *Department of Electronics and Communication Engineering,*

Nitte Meenakshi Institute of Technology,

Yelahanka, Bangalore-560064, Karnataka, India.

Abstract

In this work, the amplitude profile of sampled speech data were fitted by employing sum of sine functions with a confidence level more than 90%. Furthermore, amplitude correlation technique is applied between original speech signal samples of normal and pathological subjects and correlation technique is also applied between the curve fit constant values for normal and pathological subjects. Results obtained in both the techniques were compared to determine the varying degrees of speech disability severity.

Keywords

Correlation, Curve fitting, Discrete time signal, Linear filter, Speech disability.

1. Introduction

Speech is the most basic means communication between humans. In speech communication an idea is first formed in the mind of speaker, and then it is transformed in to words, phrases and sentences according to the grammatical rules of the language. Effective determination of quantitative speech disability remains one of the challenges in medical profession.

However millions of people are suffering from disabilities associated with speech. According to an UNICEF estimate as many as 150 to 250 million children are disabled world wide [1]. Speech disability forms one of the widely encountered among the various disabilities among children [2-8]. Speech synthesis among the disabled children is the first step for making speech corrective tool kit. Most speech reproduction models of vowels, semi-vowels, nasals consist of linear filter excited by a periodic signal p(t) having the following form

$$p(t) = \sum_{k=0}^{\infty} \left\{ \begin{array}{l} p_k \cos 2\pi \kappa\alpha(t+\theta) \\ + \hat{p}_k \sin 2\pi\kappa\alpha(t+\theta) \end{array} \right\} \qquad (1)$$

Where α is the fundamental voice frequency, p_k and $\hat{p}_k$ are constants, t is a random variable uniformly distributed between 0 and $1/\alpha$. This model has been used in all the speech synthesizers. Speech reproduction is necessary for applying corrective techniques [4]. Various techniques were researched by several researchers over the years like articulation density index [4], hidden Markov model [9-12], Hidden Bernoulli Model [13], zero crossing referencing [14], and LPC techniques [15-16] for speech signal processing.

In this paper, the speech uttered by children of age group of 5 to 8years was recorded and digitized. The digitized signal was further processed in a high level computational platform of MATLAB. . In Section 2, the procedure of speech recording and digitization were explained; Section 3 deals signal amplitude correlation technique, Section 4 deals with curve fitting procedure for speech signals, in section 5 discussion has been made regarding the procedure followed in determining the optimal window block size which results in better curve fit and Section 6 the results obtained from amplitude correlation technique as well as curve fitting techniques were explained. Finally, conclusions were drawn in Section 7.

2. Experimental Speech Recording

Acoustic signals or speech signals are continuous signals in time domain. To convert such signals into discrete time signals, Analog to Digital converters (A to D Converters) are used. The current state of the art recording devices such as IPOD (Intelligent Portable Occular Device) are equipped with advanced A to D converters largely due to the technological advancements in VLSI. Built-in A to D converters in IPOD system enables the creation of discrete- time signal at discrete intervals of time, when speech signal passed through such system.

The study was conducted on children of native speakers of Kannada language (Indian Dravidian Language which was recognized as classical language recently [20]). Acoustic speech signals of word “Namma” uttered by normal child (normal subject) and pathological subjects (age between 5 to 8 years) are recorded using IPOD recorder in digital form.

The data was collected at J.S.S. ‘Sahana’ integrated and special school located at Bangalore, India. Prior to the data collection, the teachers were requested to train and acclimatize the disabled children aged between 5 and 8 years facing speech disability problem, to read few sentences written in kannnada language. The experiment was conducted in a pleasant atmosphere, the children are made to read the sentences and the said sentences were recorded. Similarly the normal samples were obtained at ‘vinayaka’ school, Bangalore by following same procedure with the normal children in the same age group. About 30 samples were taken in all for normal cases. The averages of these samples were taken and the child’s voice which nearly coincides with the average is considered as reference data mentioned as ‘normal’.

Table 1. The parameters used for transformation into .wav file

Rate [kbps]	512
Amplitude type [bit]	16
Sampling Frequency [Hz]	32000
Channel	Mono
Modulation	Pulse Coded

However the data after conversion should be made easy to use, manipulate, low loss and compatible to windows platform. The recorded digitized signal is further subjected to transformation into wav file, the most commonly used digitized sound format with the help of downloadable software known as “GOLDWAVE” [21]. The transformation procedure used in GOLDWAVE software results in .wav file format.

3. Amplitude Correlation Technique

The parameters used while importing the IPOD recorded speech into GOLDWAVE are shown in Table 1. The compiled data are marked as 'normal', 'subject1', 'subject2' and 'subject3' corresponding to normal case, and pathological cases with increase in severity.

The acoustic speech signals for common word uttered by normal child (normal subject) and pathological subjects are properly captured with the appropriate dimension consisting of uniform numbers of digital samples. Absolute value of such digital sample sequence is obtained in each individual normal and pathological case. Then correlation coefficient can be calculated between the normal subject samples sequence absolute values and each individual pathological subject sample sequence absolute values. Such correlation coefficient gives a measure like how much the individual pathological subjects speech signals are deviating from the normal subject speech signals.

4. Curve Fitting Procedure

Recorded acoustic speech signals for common Kannada word uttered by normal child (normal subject) and pathological subjects are having the same length of digitized sample sequence of 'n' numbers. MATLAB curve fitting tool (cftool) is used to achieve the better curve fitting for the above said samples sets.

To have better curve fitting with confidence level of more than 90% and with a minimum value of RMSE (root mean squared error), the sample sequence set is divided in to few number of windows. Each window must consist of some suitable number of samples.

The mathematical function (sum of sine functions) given in equation (2) is used to get proper curve fitting with the above said window samples.

$$\begin{aligned}&a_1 \mathrm{Sin}(b_1 * X_{(n)} + c_1) + a_2 \mathrm{Sin}(b_2 * X_{(n)} + c_2) + \dots\dots\dots\\ &\dots\dots\dots\dots + a_8 \mathrm{Sin}(b8 * X_{(n)} + c_8)\end{aligned} \tag{2}$$

a_1 to a_8, b_1 to b_8 and c_1 to c_8 are constants

X(n) represent the sample sequence, where n=1, 2,3,……

The above discussed "sum of sine function" can be rewritten to suite for different windows of normal subject and pathological subject as like bellow

$$\begin{aligned}&a_1 \mathrm{Sin}(b_1 * Xr_{(n)} + c_1) + a_2 \mathrm{Sin}(b_2 * Xr_{(n)} + c_2) + \dots\dots\dots\\ &\dots\dots\dots\dots + a_8 \mathrm{Sin}(b8 * Xr_{(n)} + c_8)\end{aligned} \tag{3}$$

$Xr_{(n)}$ represents the speech signal samples sequence set of normal subject which has been taken as reference.

$$\begin{aligned}&a_1 \mathrm{Sin}(b_1 * Xps_{(n)} + c_1) + a_2 \mathrm{Sin}(b_2 * Xps_{(n)} + c_2) + \dots\dots\dots\\ &\dots\dots\dots\dots + a_8 \mathrm{Sin}(b8 * Xps_{(n)} + c_8)\end{aligned} \tag{4}$$

$Xps_{(n)}$ represents the samples of pathological subject

At first the curve fitting technique is applied to the normal subject speech sample sequence window wise and those window related constants are obtained.

i.e., a_1 to a_8, b_1 to b_8 and c_1 to c_8 are constants

Such window related constant values obtained for different windows are looks to be different. All windows related constants values are arranged in a single column by keeping all window related constants in a sequence.

Next the curve fitting technique is applied to the pathological subject1 speech signal samples sequence windows and above said procedure must be followed and curve fitting constants are obtained in a single column.

Above procedure must be applied for the other pathological speech signal samples sequence also. Correlation coefficient or correlation index can be calculated between normal child curve fitting constant values and individual pathological subject curve fitting constant values to know how much deviation is there between normal subject speech signal curve fit and those of individual pathological subject speech signal curve fit.

5. Results and Discussion

In this current study, the Kannada word "Namma" has been uttered by normal subject, pathological subject1, pathological subject2 and pathological subject3.The number of digitized speech signal samples obtained for this speech signal "Namma" are very large in number consisting of a total 21121 numbers of samples.

To determine the optimal window block size which result in better curve fit using the above said sum of sine function, three different cases of varying window block size were considered as mentioned below.

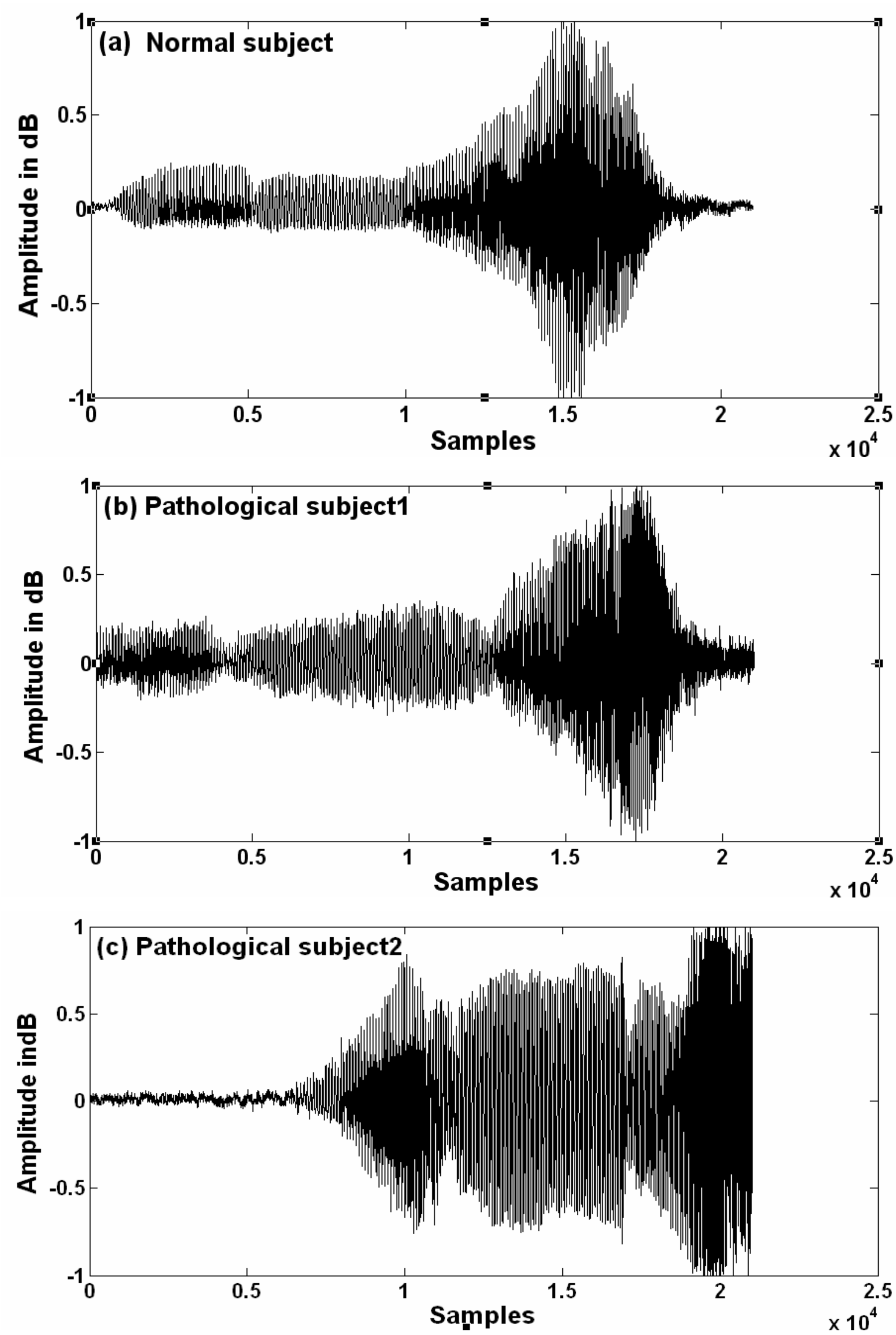
Case 1: The total numbers of samples are divided in to 43 window blocks. For window block 1 to window 42, the window block comprises 501 samples and for window block 43 is 79 samples. After curve fit processing for each of such window has 24 constants (a_1 to a_8, b_1 to b_8, and c_1 to c_8). In total, 1032 numbers of curve fit constant values can be listed from 43 window blocks and those values are arranged in a sequence in a single column.

Case 2: The total numbers of samples are divided in to 22 window blocks. For window block 1 to window 21, the window block comprises 1001 samples and for window block 22 is 100 samples. After curve fit processing for each of such window has 24 constants (a_1 to a_8, b_1 to b_8, and c_1 to c_8). In total, 528 numbers of curve fit constant values can be listed from 22 window blocks and those values are arranged in a sequence in a single column.

Case 3: The total numbers of samples are divided in to 11 window blocks. For window block 1 to window 10, the window block comprises 2001 samples and for window block 11 is 1111 samples. After curve fit processing for each of such window has 24 constants (a_1 to a_8, b_1 to b_8, and c_1 to c_8). In total, 264 numbers of curve fit constant values can be listed from 11 window blocks and those values are arranged in a sequence in a single column.

It is observed from the experimental verification that in case1 the curve fitting computation complexity is more and reproduction of sound with the curve fit samples matches with original sound signal samples. In case2 the curve fitting computation complexity is reduced and still reproduction of sound with the curve fit samples matches with original sound signal samples. In case3the curve fitting computation complexity is least and reproduction of sound with the curve fit samples does not match with original sound signal samples and appears to be noisy. Hence in this study for all practical purposes case2 is considered as most suitable and optimal for curve fit processing.

The speech signal samples plot of the word 'Namma' uttered by normal subject, pathological subject1, pathological subject2 and pathological subject3. are shown in figure1.


(a) Normal subject
Amplitude in dB
Samples
x 10^4
(b) Pathological subject1
Amplitude in dB
Samples
x 10^4
(c) Pathological subject2
Amplitude indB
Samples
x 10^4

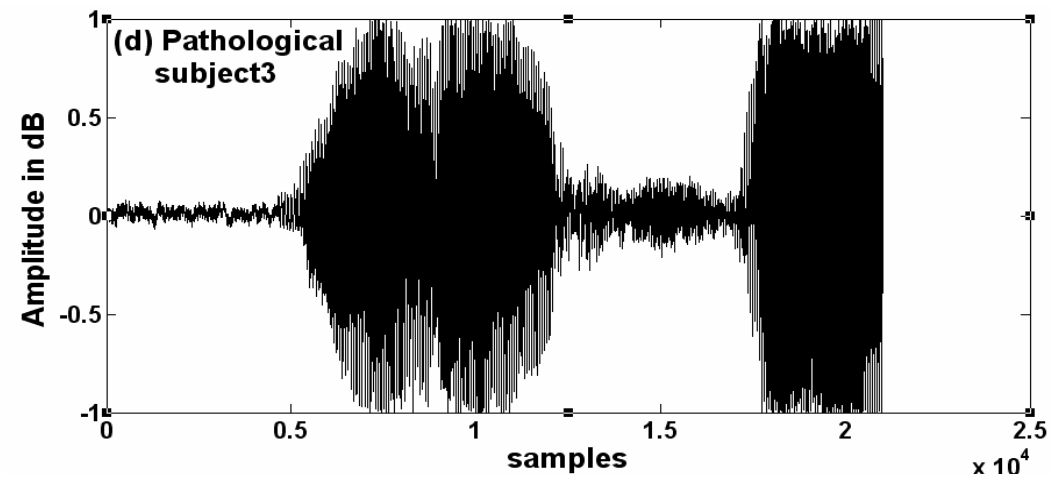


Figure1. Plot of the speech signal samples for the word 'Namma' (a) Normal subject (b) Pathological subject1 (c) Pathological subject2 and (d) Pathological subject3

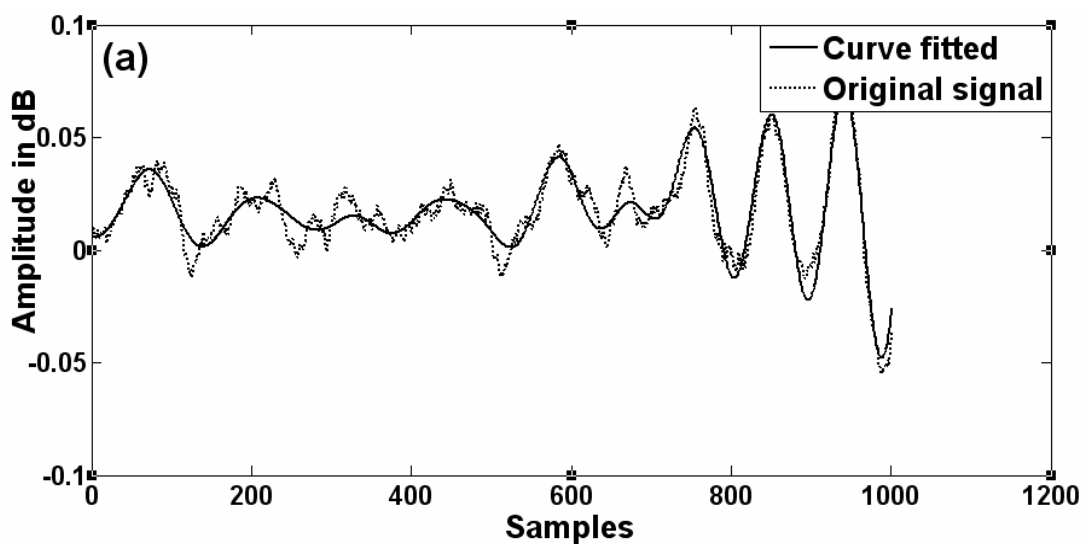
If the values of two variable deviate in the same direction i.e., if the increase in the values of one variable results, on an average, in a corresponding increase in the values of the other variable or if a decrease in the values of one variable results, on an average, in a corresponding decrease in the values of the other variable, correlation is said to be positive or direct.

On the other hand, correlation is said to be negative or inverse if the variables deviate in the opposite direction i.e., if the increase (decrease) in the values of one variable results, on the average, in a corresponding decrease (increase) in the values of the other variable.

The correlation coefficients values are calculated between the normal subject samples sequence absolute values and each individual pathological subject sample sequence absolute values and the results are tabulated in Table-2

Table 2 Amplitude Correlation Study

S No	Comparative study	Correlation coefficients
1	'normal' and 'subject1'	0.28
2	'normal' and 'subject2'	0.24
3	'normal' and 'subject3'	-0.188



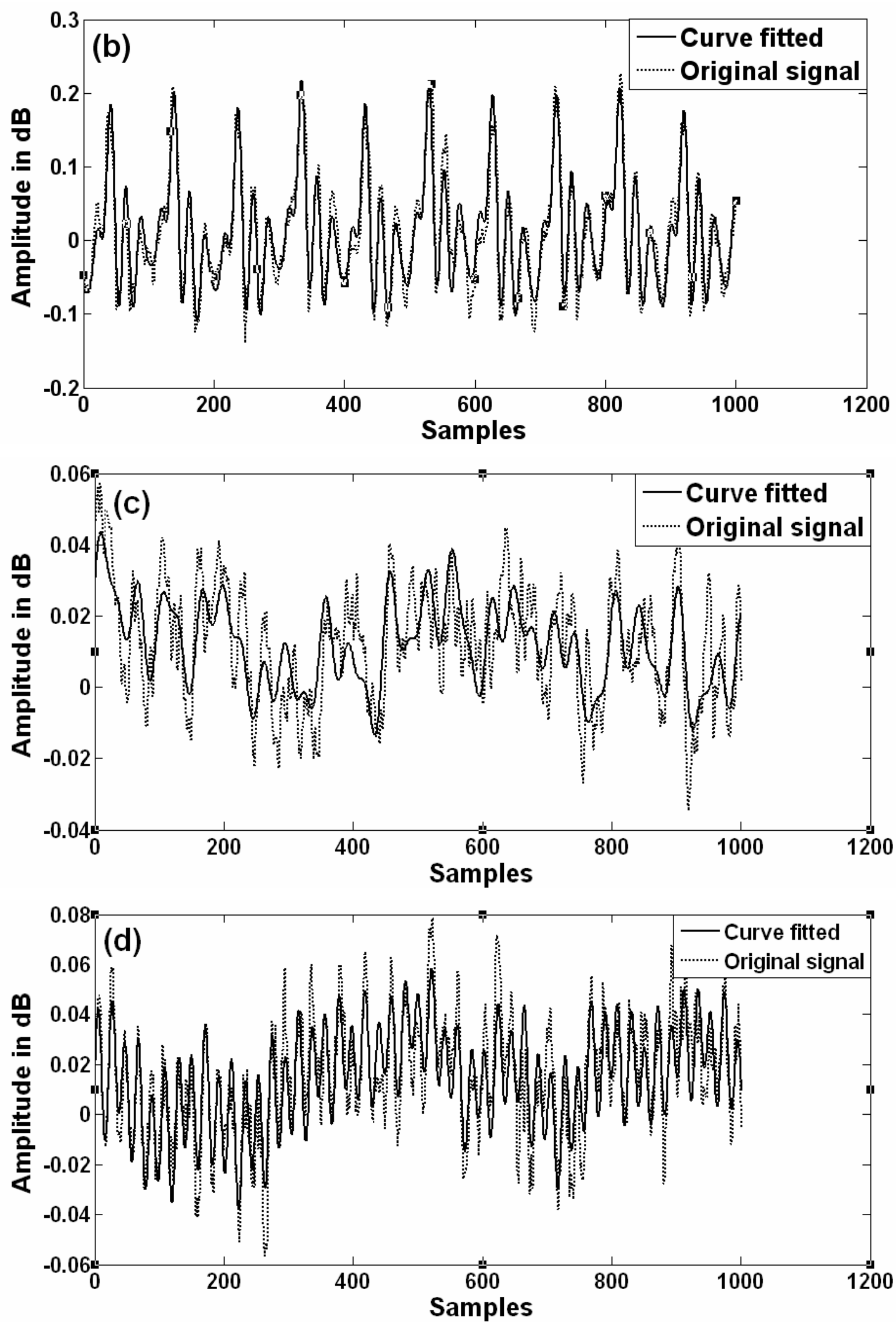


Figure 2 Cure fitting for window1 samples (a) Normal subject (b) Pathological subject1 (c) Pathological subject2 and (d) Pathological subject3 [The dotted line is the recorded speech signal and solid line is the fitted 0ne]

With reference to above cited case2, curve fitting obtained for the window block1 speech signal samples of normal, pathological subject1, pathological subject2 and pathological subject3 are shown in figure2.
Curve fit procedure mentioned in case2 is applied for normal and pathological subjects. The correlation coefficient between normal subject curve fit constants values and each individual pathological subjects curve fit constant values are calculated and results are tabulated in Table-3.

Table 3 Correlation Study

S No	Comparative study	Correlation coefficients
1	'normal' and 'subject1'	0.2067
2	'normal' and 'subject2'	0.0012
3	'normal' and 'subject3'	0.0005

6. Conclusions

For the collected sample space of four children the curve fitting technique is applied and curve fitting constant values are obtained. Then correlation coefficient between normal and pathological curve fitting constant values are calculated and it is observed that if correlation coefficient value starts deviating from its expected value 0.20, speech disability severity increases in pathological subjects in comparison with normal child.

7. Acknowledgement

The authors would like to thank the management of Sri Bhagawan Mahaveer Jain college of Engineering, Bangalore, Nitte Meenakshi Institute of Technology, Bangalore, J.S.S. 'Sahana' integrated and special school, Bangalore and Authorities and Department of Speech & Hearing, NIMHANS, Bangalore, for their Constant support and encouragement in undertaking the research work.

8. References

[1].http://www.disabilityworld.org/06-08-03/children/unicef.shtml

[2]. Lawrence Rabiner and Biling-Hwang Juang, "Fundamentals of Speech Recognition", second Impression, published by Dorling Kindersley (India) Pvt. Ltd., licensees of Pearson Education in South Asia.2007.

[3]. Ben Gold and Nelson Morgan, "Speech and Audio Signal Processing (Processing and Perception of Speech and Music)", John Wiley & Sons, Inc.2006.

[4]. J. B. Allen, "How do humans process and recognize speech", IEEE Transactions on Speech and Audio Processing vol. 2, No 4, October 1994, pp. 567-577.

[5]. Harry Levitt," Speech Processing Aids for the Deaf an Overview" IEEE Transactions on Audio and Electroacoustics, Vol. Au-21, No.3, June 1973, pp. 269- 273.

[6]. Ronald W. Schlffer, Member, IEEE, "A Surveyof Digital Speech Processing Techniques"IEEE Transaction on Audio and Electro acoustics Vol. Au-20, No. 1, March,1972, pp. 28-35.

[7]. B. Garcia, J. Vicente, I. Ruiz, A. Alonso, "Multiplatform Interface Adapted To Pathological Voices", IEEE Symposium on Signal Processing and Information Technology, 2005, pp. 912-917.

[8].CHERIF Adnbne-Botiafif Lamia-Mhamdi mounir, "Analysis of Pathological Voices by Speech Processing", IEEE, 2003, pp. 365-367.

[9]. Y. Ephraim, D. Malah, and B. H. Juang, "On the application of hidden Markov models for enhancing noisy speech", IEEE Transactions on Speech and Audio Processing vol. 37, October 1989, pp.. 1846-1856.

[10]. J. Picone "Continuous speech recognition using Hidden Markov Models", IEEE, ASAP Magazine, July 1990, pp. 26-41.

[11]. T. Kobayashi, J. Furuyama, K. Masumitsu, "PartlyHidden Markov Model and its application to speech recognition", IEEE, 1991, pp. 121-124.

[12]. N. Ratnayake, M. Savic, J. Sorensen, "Use of Semi- Markov Models for speaker-independent phoneme recognition", IEEE, 1992, pp. 565-568

[13]. J. Kabudian, M. M. Homayounpour, S. Mohammad Ahadi, "Time-Inhomogeneous hidden Bernoulli Model:An Alternative to Hidden Markov Model for Automatic Speech recognition", Proceedings of ICASSP, 2008, pp. 4101-4104.

[14]. M. Robert ITO, R. W. Donaldson, "Zero-crossing measurements for analysis and recognition of speech sounds", IEEE Transactions on Audio and Electroacoustics, Vol. 19, No.3, 1971, pp. 235-242.

[15]. B. S. Atal and Suzanee L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", the Journal of the Acoustical Society of America, April 1971, pp. 637-655.

[16]. Susanna Varho and Paavo Alku, "Regressive Linear Prediction with Triplets-An Effective All-Pole Modeling Technique for Speech Processing", IEEE, 1998, pp. IV- 194 – IV-197.

[17].R. Kumrisan and C.S.Ramalingam, "On separating voiced-speech into its components", IEEE, 1993, pp.1041-1046.

[18].Yuanning Liu, Senmiao Yuan and Xiaodong Zhu, "A Time-frequency Field Fingerprint Enhancement Technology and Three-order Spline Curve Fitting Matching Algorithm Research", IMTC 2003- Instrumentation and Measurement Technology Conference, Vail, Co, USA, 20-22 May 2003, pp.1067-1069.

[19].K.Steiglitz, G. Winham and J.Petzinger, "Pitch Extraction by Trigonometric Curve Fitting",IEEE Transactions on acoustics, speech and signal processing, June 1975,pp.321-323.

[20]. http://www.hindu.com (dated 1st Nov 2008)

[21]. http://www.goldwave.com/

Anandthirtha B. Gudi obtained Bachelor of Engineering from S.J.M.Institute of Technology, Chitradurga, Mysore University**.** Master of Engineering from U. V. C. E.Bangalore, Bangalore University**.** Professor in the Department of Electronics and Communication Engineering, Sri Bhagawan Mahaveer Jain College of Engineering, Bangalore. Karnataka, India.

Dr. H. C. Nagaraj obtained Bachelor of Engineering from Mysore University**,** Master of Engineering from P. S. G. College of Technology, Coimbatore. PhD from I.I.T., Madras. Principal, Nitte Meenakshi Institute of Technology, Bangalore, Karnataka, India**.**