ISSN 1726-5479

SENSORS & TRANSDUCERS

vol. 107
8/09

Sensors and Transducers Applications

International Frequency Sensor Association Publishing

IFSA

Sensors & Transducers

Volume 107, Issue 8
August 2009

www.sensorsportal.com

ISSN 1726-5479

Sensors & Transducers Journal (ISSN 1726-5479) is a peer review international journal published monthly online by International Frequency Sensor Association (IFSA).
Available in electronic and on CD.

Sensors & Transducers Journal

Contents

Volume 107
Issue 8
August 2009

www.sensorsportal.com

ISSN 1726-5479

Research Articles

Authors are encouraged to submit article in MS Word (doc) and Acrobat (pdf) formats by e-mail: editor@sensorsportal.com
Please visit journal’s webpage with preparation instructions: http://www.sensorsportal.com/HTML/DIGEST/Submition.htm

Sensors & Transducers Journal, Vol. 107, Issue 8, August 2009, pp. 157-164

Sensors & Transducers

ISSN 1726-5479

http://www.sensorsportal.com

Speech Disability Threshold Determination by Graphical and DSP Techniques

[1]Anandthirtha. B. GUDI and [2]H. C. Nagaraj
[1]Department of Electronics and Communication Engineering,
Sri Bhagawan Mahaveer Jain College of Engineering (SBMJCE), Bangalore-562112, Karnataka, India
Tel.: 0091 80 27577200, fax: 0091 80 27577199
[2]Department of Electronics and Communication Engineering,
Nitte Meenakshi Institute of Technology (NMIT), Yelahanka, Bangalore-560064, Karnataka, India
E-mail: abgudi@jaingroup.info, gudi_anand@rediffmail.com

Received: 8 March 2009 /Accepted: 15 August 2009 /Published: 25 August 2009

Abstract: In this paper, an attempt is made to graphically distinguish the normal and pathological cases of speech disabled children. The digitized acoustic signal from children of age range 5 to 8 years are captured and subjected to Linear Predictive Coding (LPC), FFT and CEPSTRUM processes. The output data obtained from LPC, FFT and CEPSTRUM are analyzed for four cases viz, one normal and three pathological cases of varying degrees of severity. The usefulness of the proposed technique is further illustrated as the locations of poles are found to have shifted within the unit circle. The results are summarized in the form of three indices namely co-relation index of FFT signal, co-relation index of CEPSTRUM and the order of prediction in LPC for severity threshold determination. The results from FFT and CEPSTRUM are found to match with that of the LPC technique favorably. *Copyright © 2009 IFSA.*

Keywords: CEPSTRUM, Correlation, FFT, Linear predictive coding and speech disability

1. Introduction

Speech is the most basic means of adult human communication and also speech is one of the basic forms of communication used in any civilized society. 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 is one of the widely encountered disabilities among the children [2-8]. The purpose of this paper is to explore how digital signal processing techniques can be applied to solve problems related to speech communication i.e. to quantitatively measure severity of speech disability among disabled children (Pathological Subjects) in comparison with medically fit child, using high level computational language like MATLAB as a platform.

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 (HMM) [9-12], Hidden Bernoulli Model [13], zero crossing referencing [14], and Linear Predictive Coding (LPC) techniques [15-16] for speech signal processing.

In this paper, the speech uttered by children was recorded and digitized. The digitized signal was further processed in a high level computational platform of MATLAB. These are further explained in subsequent Sections. In Section 2, the procedure of speech recording and digitization was explained; Section 3 deals with LPC processing technique, and Section 4 the results obtained from LPC technique as well as other techniques such as FFT and CEPSTRUM were explained. Finally Conclusions were drawn in Section 5.

2. Methodology

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 Converter) are required. 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 [17]). 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 years to 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 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’.

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” [18]. The transformation procedure used in GOLDWAVE software results in .wav file format.

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.

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

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

3. Speech Signal Processing Using LPC

In simple terms, the technique of Linear Prediction Coding (LPC) operates by determining the optimum values of filter coefficients required to make the prediction error as small as possible. The LPC technique is widely used in Filter design and speech coding. The higher the order of the filter used the lower the error between the predicted signal and the actual signal. The order of the filter is denoted by the symbol '*p*' – linear prediction coding coefficient. The values of the various fitted filter coefficients determine the strength of the poles.

In the case of linear prediction, the intention is to determine an FIR filter that can optimally predict future samples of an autoregressive process based on a linear combination of past samples. The difference between the actual autoregressive signal and the predicted signal is called the prediction error. Ideally, this error is white noise. The representation of LPC technique used for speech signal analysis is illustrated in Fig. 1.

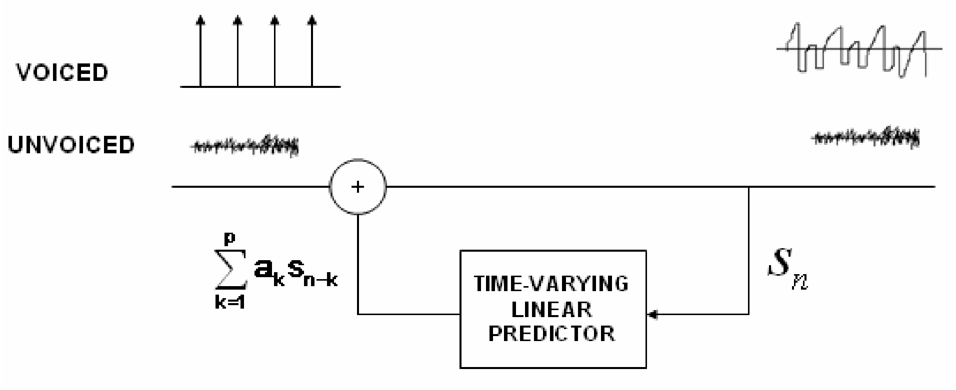

Fig. 1. Illustration of LPC technique used in Speech Signal Analysis.

Referring to Fig. 1, the Vocal – Cord excitation for voiced sounds is produced by a known generator with adjustable period and amplitude. The noise-like excitation of unvoiced sounds is produced by a white-noise source. The linear predictor *p*, a transversal filter with *p* delays of one samples interval each, forms a weighted sum of the past *p* Samples at the input of the predictor. The output of the linear filter at the nth sampling instant is given by,

$$S_n = \sum_{K=1}^{P} a_n S_{n-k} + S_{nP} \,, \tag{1}$$

where the "predictor coefficient" a_k account for the filtering action of the vocal tract, the radiation, and the glottal flow; and S_n represents the n^{th} sample of the excitation. The transfer function of the linear filter of Fig. 1 is given by

$$T(Z) = \frac{1}{\left(1 - \sum a_k Z^{-k}\right)} \tag{2}$$

The poles of T(Z) are the (reciprocal) zeros of the polynomial (in z^{-1}) in the denominator on the right side of equation (2). The linear filter thus has a total of *p* poles, which are either real or occur in conjugate pairs. The filter stability condition requires that the poles must reside inside the unit polar circle.

The number of coefficient *p* registered to present any speech segment adequately is determined by the number of resonance and anti resonance of the vocal tract in the in the frequency range of interest, the nature of the glottal volume flow function, and radiation. Normally two poles are usually adequate to represent the influence of the glottal flow and the radiation on the speech wave [15]. In order to represent the poles of the vocal tract transfer function adequately, the linear predictor memory must be equal to twice the time required for sound waver to travel from the glottis to the lips (nasal opening for nasal sound), for example, if the vocal tract is 17 cm in length, the memory of the predictor should be roughly 1 ms in order to represent the poles of transfer function of the vocal tract. *p* is natural function of the sampling frequency f_s and is roughly proportional to f_s. Considering a sampling interval of 0.1 ms with the sampling frequency of 10000 Hz results in *p* value of 10, an additional two poles are required, one accounting for glottal flow and another for the radiation. Hence the optimum value for *p* for a sampling rate of 0.1 ms is 12. In our analysis, the optimum value of *p* corresponding to minimum error is found to be around 10. Greater distinction between various severities cases are noticed on unit polar plot for *p* value equal to 12. Hence in our study also *p* value is taken equal to 12.

LPC plot in Fig. 2 shows a unit polar plot for all the cases. The locations of captured 12 poles from LPC signal analysis are shown in Fig. 2(a) to Fig. 2(d) for the cases of normal, subject1, subject2 and subject3. It can be seen that the real poles did not exist on the real axis of the unit circle in most severe case 'subject3'. Furthermore it was noticed that some poles were moved towards periphery of the unit circle, indicating the severity of the disease.

4. Discussion

The exact value of frequency at which the abnormality occurs could not be determined from LPC. Hence there is a need for the usage of other tools. FFT analysis can be used to study the spectrum splitting thus identifying the frequency component. FFT can be used to distinguish various severe cases.

The procedure extraction of real and imaginary components of CEPSTRUM is shown in Fig. 3(a) and Fig. 3(b), respectively. Real CEPSTRUM can be obtained by computing the inverse discrete Fourier transform of the logarithm of magnitude of the discrete Fourier transform of the input speech sequence X(n) and Complex CEPSTRUM can be obtained by computing the inverse discrete Fourier transform of the Complex logarithm of the discrete Fourier transform of the input speech sequence X(n).

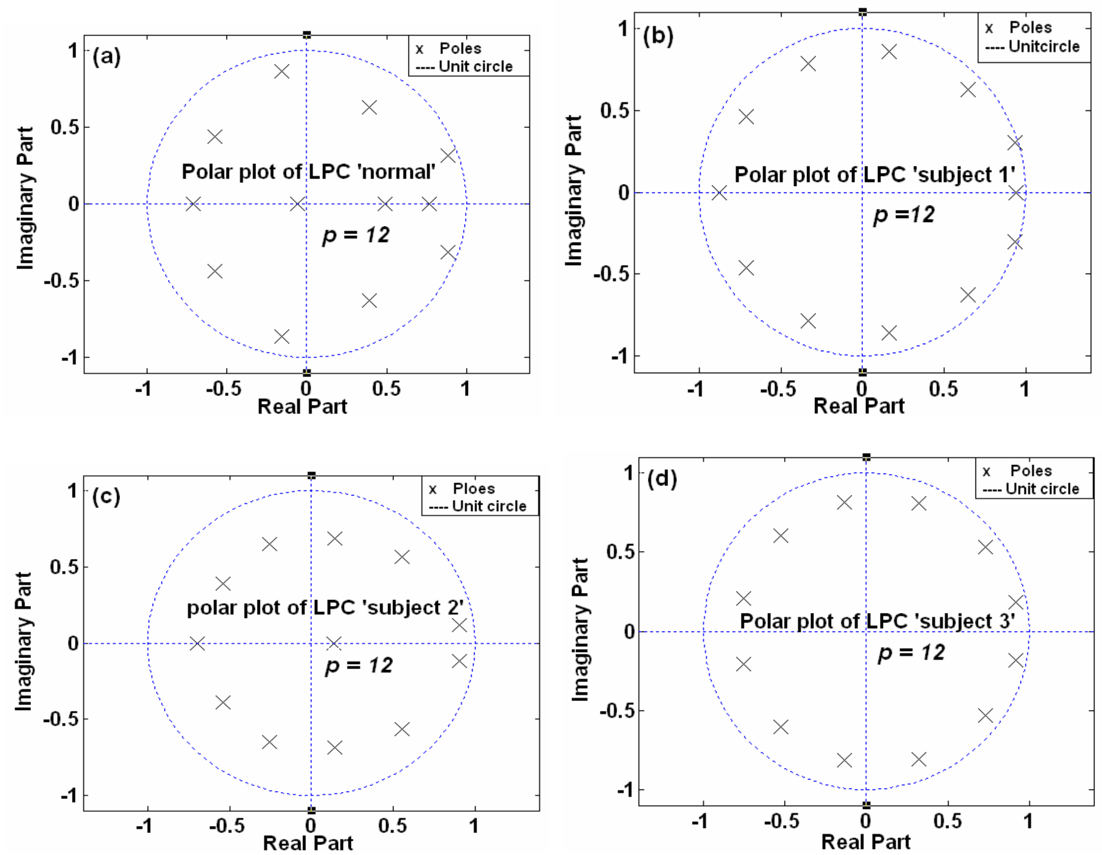

Fig. 2. The LPC pole-zero plots for (a). normal (b). Subject1 (c). Subject2 and (d). Subject3. P = 12.

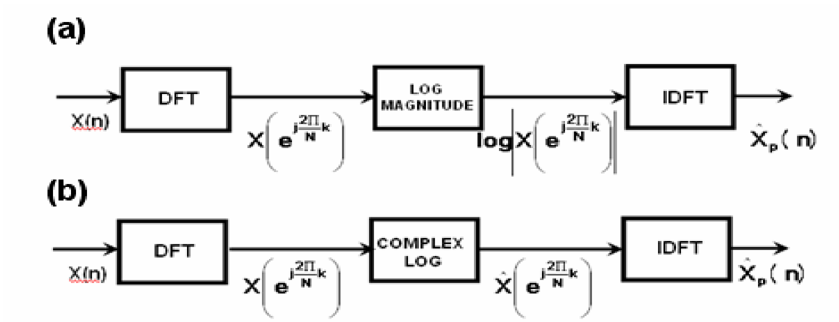

Fig. 3. Procedure for Extraction of CEPSTRUM Components
(a). Real CEPSTRUM (b). Complex CEPSTRUM.

From Fig. 4, the extent of severity can be easily recognized by comparing normal case amplitude in (a) and the most severe case in (b). It is seen the net energy is more negative in the case of subject3 indicating the difficulty in speech transmission and irregularity in gottal flow.

The Fast Fourier Transform (FFT) may also be used to evaluate the frequency response of discrete-time systems. The frequency response of discrete- time system is simply the Fourier transform of its impulse response. To obtain a smooth frequency response, it is important to take a sufficient number of impulse response values and/or to zero-pad the impulse response values before the FFT is taken. Using FFT function of MATLAB, frequency response of normal subject1, subject2 and subjects3 are obtained and the plots for only the normal case and subject 3 are shown in Fig. 5 (a) and (b) respectively.

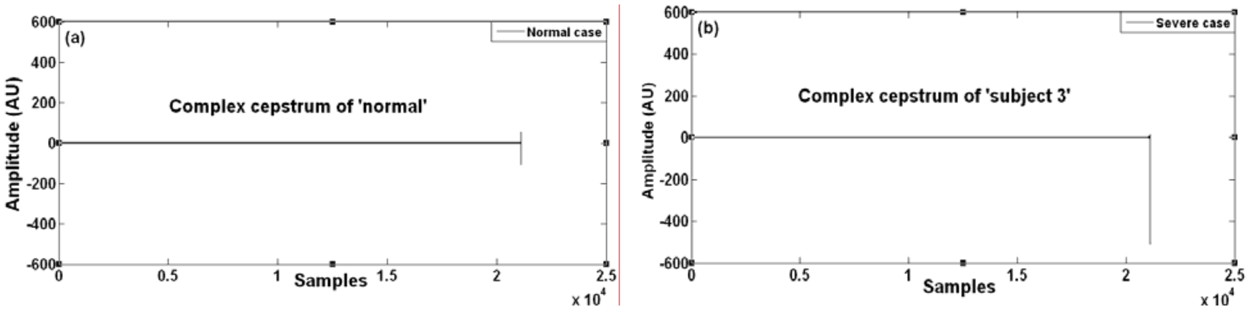

Fig. 4. Plotted Complex component of CEPSTRUM for (a) Normal and (b) subject3.

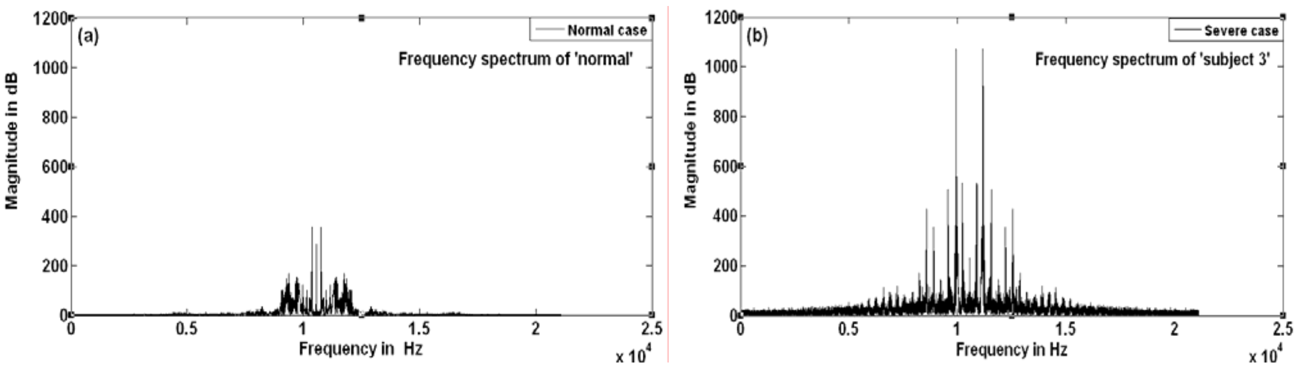

Fig. 5 Frequency Spectrum for (a) Normal and (b) subject3.

The cross-correlation becomes important when the case of two independent and random data sequences have to be compared. In this case the sum of the products will tend to vanishingly small random number as the number of pairs of points is increased. This is because all numbers, positive and negative, are equally likely to occur so that the product pairs tend to be self-canceling on summation. The cross- correlation r_{12} (n) between two data sequences $x_1(n)$ and $x_2(n)$ each containing N data which is written as

$$r_{12} = \sum_{n=0}^{N-1} x_1(n) x_2(n) \tag{3}$$

In this work, cross- correlation is applied between the speech signal samples of normal child ($x_1(n)$ normal) and speech signal samples of Pathological subjects (($x_2(n)$ pathological subjects). (Three respective Pathological subject speech signal samples cross-correlation with normal child speech signal the correlation coefficients or the correlation index values obtained.). Similarly, Correlation coefficients are obtained between the normal FFT values and subject1, subject2 and subject3 FFT values, and are tabulated in Table 2.

Table 2. FFT correlation study.

S No.	Comparative Study	FFT Correlation coefficients
1	'normal' and 'subject1'	0.42
2	'normal' and 'subject2'	0.28
3	'normal' and 'subject3'	0.21

Table 3. CEPSTRUM correlation study.

S No.	Comparative Study	CEPSTRUM correlation coefficient
1	'normal' and 'subject1'	0.85
2	'normal' and 'subject2'	0.78
3	'normal' and 'subject3'	0.76

It is seen that various data presented can serve as severity indicators. However for the speech therapy tools to be in place, the techniques of transformation of speech signal from severe cases to normal cases to be explored. The authors believe that this procedure is not only necessary but also will make the current topic a complete study. In an attempt to this, the severe case of subject 3 was taken and observed by adjusting the value of prediction coefficient and thus making the polar plot nearly coinciding to the normal case for the *p* value equal to 17.

5. Conclusions

Together FFT, LPC and CEPSTRUM form a complete tool kit for severity detection and threshold determination. More specifically three indices are identified, namely *p* value, FFT co-relation and CEPSTRUM co-relation index for the quantification of speech severity. For the collected sample space of four children, the threshold value for *p* and threshold co–relation indices are found to be 12. The threshold values of the correlation index of FFT, CEPSTRUM are 0.4 and 0.8 respectively. Work is on-going to make a complete toolkit of identifying the threshold parameters and correction tools.

Acknowledgements

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.

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, *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, A Survey of Digital Speech Processing Techniques, *IEEE Transaction on Audio and Electroacoustics*, 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, Partly Hidden 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].http://www.hindu.com (dated 1st Nov 2008).
[18].http://www.goldwave.com/

___

Guide for Contributors

Aims and Scope

Sensors & Transducers Journal (ISSN 1726-5479) provides an advanced forum for the science and technology of physical, chemical sensors and biosensors. It publishes state-of-the-art reviews, regular research and application specific papers, short notes, letters to Editor and sensors related books reviews as well as academic, practical and commercial information of interest to its readership. Because it is an open access, peer review international journal, papers rapidly published in *Sensors & Transducers Journal* will receive a very high publicity. The journal is published monthly as twelve issues per annual by International Frequency Association (IFSA). In additional, some special sponsored and conference issues published annually. *Sensors & Transducers Journal* is indexed and abstracted very quickly by Chemical Abstracts, IndexCopernicus Journals Master List, Open J-Gate, Google Scholar, etc.

Topics Covered

Contributions are invited on all aspects of research, development and application of the science and technology of sensors, transducers and sensor instrumentations. Topics include, but are not restricted to:

- Physical, chemical and biosensors;
- Digital, frequency, period, duty-cycle, time interval, PWM, pulse number output sensors and transducers;
- Theory, principles, effects, design, standardization and modeling;
- Smart sensors and systems;
- Sensor instrumentation;
- Virtual instruments;
- Sensors interfaces, buses and networks;
- Signal processing;
- Frequency (period, duty-cycle)-to-digital converters, ADC;
- Technologies and materials;
- Nanosensors;
- Microsystems;
- Applications.

Submission of papers

Articles should be written in English. Authors are invited to submit by e-mail editor@sensorsportal.com 8-14 pages article (including abstract, illustrations (color or grayscale), photos and references) in both: MS Word (doc) and Acrobat (pdf) formats. Detailed preparation instructions, paper example and template of manuscript are available from the journal's webpage: http://www.sensorsportal.com/HTML/DIGEST/Submition.htm Authors must follow the instructions strictly when submitting their manuscripts.

Advertising Information