CLASSIFICATION OF NORMAL AND PATHOLOGICAL INFANT CRIES USING BISPECTRUM FEATURES

Anshu Chittora and Hemant A. Patil

Dhirubhai Ambani Institute of Information and Communication Technology,
Gandhinagar, Gujarat, India
{anshu_chittora, hemant_patil}@daiict.ac.in

ABSTRACT

In this paper, bispectrum-based feature extraction method is proposed for classification of normal *vs.* pathological infant cries. Bispectrum is computed for all segments of normal as well as pathological cries. Bispectrum is a two-dimensional (*2-D*) feature. A tensor is formed using these bispectrum features and then for feature reduction, higher order singular value decomposition theorem (HOSVD) is applied. Our experimental results show *70.56 %* average accuracy of classification with support vector machine (SVM) classifier, whereas baseline features, *viz*., MFCC, LPC and PLP gave classification accuracy of *52.41 %* , *61.27* % and *57.41 %,* respectively. For showing the effectiveness of the proposed feature extraction method, a comparison with other feature extraction methods which uses diagonal slice and peaks and their locations as feature vectors is given as well.

Index Terms— Higher order signal processing cumulant, bispectrum, higher order singular value decomposition theorem.

1. INTRODUCTION

In our day-to-day life, to express our emotions and need we use speech as a communication medium. Similarly, infants use cry as a signal to communicate with the world. Accordingly, cry types are classified as hunger, pain, fear and birth cries, *etc*. Mel frequency cepstral coefficients (MFCC) feature has been used in [1-2], for classification of cry in normal and hypothyroidism and normal and deaf infants class, respectively. Fundamental frequency (F_0) has also been used to classify cry as normal and pathological infant cries [3]. It is shown by researchers that fundamental frequency also varies for hunger and pain cries. Over the years, spectrographic analysis of cry has been used by researchers and they have shown promising results in pathology classification using spectrograms. Some genetic diseases show distinguished acoustic features such as double harmonic break, biphonation and rising and falling or rising melody pattern*, etc.* that indicate presence of pathology [4-7]. In [8], different melody types are defined for the classification of infant cries for pathology identification. Cry is classified in four classes, *viz*., pain, pleasure, hunger and birth [9]. MFCC has also been used for classification of pathological infant cries along with hidden Markov models (HMM) and neural networks [10-12]. To the best of authors' knowledge, features used for this task are MFCC, Perceptual Linear Prediction coefficients (PLP) and linear prediction coefficients (LPC) [1-7]. In pathological cry classification, most of the work is done on classification of normal and hearing-impaired infants, where researchers got high classification accuracy with MFCC features (around *95* %). The reason for high performance of normal and deaf infants' cry classification is the absence of auditory feedback in deaf infants which results in higher pitch (*i.e*., distinct cry). Some disease-specific classification of infant cries is limited and is restricted to classification of one type of pathology with normal infants [1-4]. In this paper, we are attempting classification of pathological infant cry from normal infant cries. Even most of pathologies considered are not the severe pathologies.

Classification of normal and pathological infant cry is a challenging task because of higher fundamental frequency (e.g., *250* Hz- *1* kHz) and higher formant frequencies due to small vocal tract length (in new born, average vocal tract length is about *7 cm*.). The variability in these features is very high because of variations in vocal tract length and development of speech production system with growing age and weight. Cry characteristics also depends on gestation age and cause of cry (hunger, pain, etc).

Traditional speech signal processing methods consider that the speech production system is *linear* and speech is *stationary* (*i.e.,* Linear Time Invariant (LTI) system). The nonzero values of higher-order cumulant show that the speech production is *not* a linear process. Because higher-order spectral analysis can capture presence of nonlinearity embedded in speech production mechanism, it may give better results for infant cry classification. Moreover, the bispectrum features are used with logarithm to give good class separation of these features. Higher order singular value decomposition theorem (HOSVD) is proposed for dimensionality reduction. In this paper, comparison of

Authors would like to thank Department of Electronics and Information Technology (DeitY), Government of India, New Delhi, India and DA-IICT, Gandhinagar for supporting this research work and providing the necessary resources. We are also thankful to authorities of Civil Hospital, Ahmedabad, for their cooperation in data collection.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

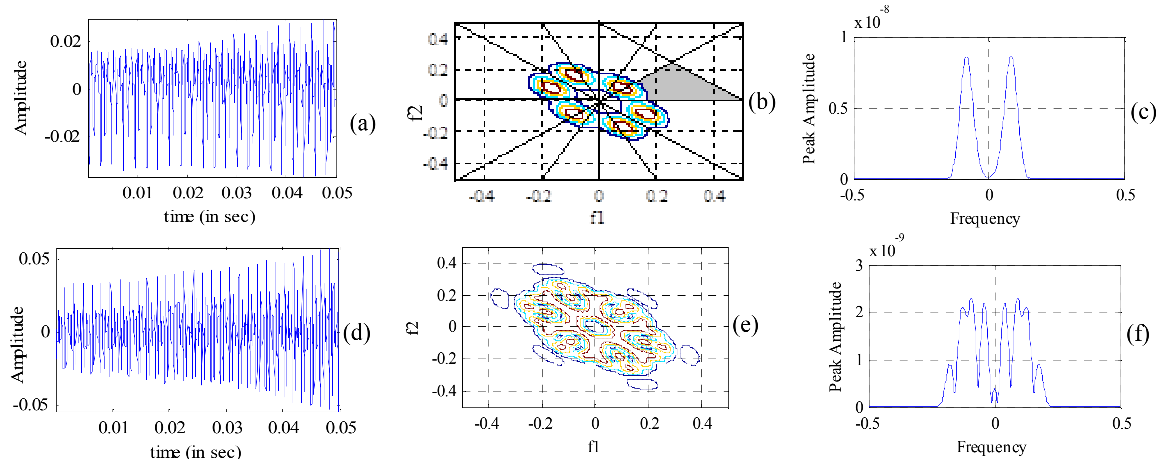

Fig.1: (a) Time- domain waveform of normal infant cry, (b) corresponding bispectrum, (c) diagonal slice of (b) , (d) time - domain waveform of pathological infant cry, (e) corresponding bispectrum, and (f) diagonal slice of (e).

proposed feature extraction method with other feature extraction methods is also reported.

2. BISPECTRUM

Speech signal is considered as a zero-mean random process and it is described by its second order moment, *i.e.*, autocorrelation function or power spectrum. Higher order spectral analysis (HOSA) yields information about the non-Gaussianity of a signal. HOSA generally deals with bispectrum and trispectrum. In this paper, bispectrum is used for feature extraction for infant cry classification. Bispectrum is defined as the 3^{rd} order spectrum of a signal. It is calculated by taking Fourier transform of the 3^{rd} order cumulant of the signal. The n^{th}-order cumulant of a signal is defined by [13-14]

$$c_n^x(\tau_1,\tau_2,...,\tau_{n-1}) = m_n^x(\tau_1,\tau_2,...,\tau_{n-1}) - m_n^G(\tau_1,\tau_2,...,\tau_{n-1}), \quad (1)$$

where $m_n^x(\tau_1,\tau_2,...,\tau_{n-1})$ is the n^{th} -order moment function of signal $x(k)$ and $m_n^G(\tau_1,\tau_2,...,\tau_{n-1})$ is the n^{th}-order moment function of an equivalent Gaussian signal that has the same *mean* value and *autocorrelation* sequence as $x(k)$. Eq. (1) is valid only for $n=2,3$ and 4. The bispectrum is written as:

$$B_3^x(\omega_1,\omega_2) = \sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty} c_3^x(\tau_1,\tau_2)\exp(-j\,|(\omega_1\tau_1+\omega_2\tau_2)|), \qquad (2)$$

for $|\omega_1| \le \pi$, $|\omega_2| \le \pi$ and $|\omega_1+\omega_2| \le \pi$.

Contour plot of bispectrum of an infant cry is shown in Fig. 1. Time-domain waveforms of normal and pathological cries are shown in Fig. 1 (a) and Fig. 1(d), respectively, and their corresponding bispectrum are shown in Fig. 1(b) and Fig. 1(e). From these bispectrum plots, it can be seen that bispectrum features show significant changes between normal and pathological infant cry. In bispectrum of normal cry, we can observe that contours are *smooth* while for pathological cry samples, it shows *irregularity*. Diagonal slices of the corresponding bispectra (which are shown in Fig. 1(c) and Fig. 1 (f)) show peaks corresponding to bispectrum pattern. These plots indicate that in pathological cry, the number of peaks is higher than the normal cry. These differences in bispectrum, motivated authors to use bispectrum-based features for normal and pathological infant cry classification. For the computation of bispectrum, direct method of bispectrum calculation using FFT is used (HOSA toolbox of MATLAB is used [15]). Bispectrum using the direct method is defined as

$$B(f_1,f_2) = X(f_1).X(f_2).X^*(f_1+f_2), \qquad (3)$$

where $X(f)$ is the *1-D* FFT of a given discrete series $x(n)$ of M samples and $X^*(.)$ is the complex conjugate of $X(.)$.

3. EXISTING FEATURE EXTRACTION METHODS

In this Section, three methods to extract features from bispectrum are presented.

3.1 Method A-Using triangular symmetry of bispectrum

In [16], *triangular symmetry* of bispectrum (as shown in Fig. 1(b)) is used and features in the triangular region of the first quadrant are calculated. The features are defined as follows:

$E_1^j = \{\sum_i |B(\omega_1^i,\omega_2^j)|\}$ where

$$i = \begin{cases} \frac{N}{2}-j+1,...,\frac{N}{2}, if\ j=1,...,\frac{N}{4}; \\ j,...,\frac{N}{2},\ if\ j=\frac{N}{4}+1,...,\frac{N}{2}. \end{cases} \qquad (4)$$

$$\{E_2^{i-N/4}\} = \left\{\sum_j |B(\omega_1^j,\omega_2^i)|\right\}, \text{ where } j=\frac{N}{2}-i+1,...,i, \qquad (5)$$

$$if\ i=\frac{N}{4}+1,...,\frac{N}{2} \text{ and } E=\left\{E_1^j, E_2^{i-\frac{N}{4}}\right\} \qquad (6)$$

In both the equations, N is the number of points in bispectrum. This method gives very large dimension feature vectors (in our case 1x512).

3.2 Method B- Diagonal slice of bispectrum

In most of the bispectrum applications, diagonal slice of bispectrum is used as a feature vector [17]. Diagonal slice is defined as bispectrum calculated on points where $\omega_1 = \omega_2$, *i.e.,*

$$D(\omega) = |B(\omega_1, \omega_2)|_{\omega_1 = \omega_2}, \tag{7}$$

Diagonal slice of bispectrum are shown in Fig. 1 (c) and Fig. 1 (f). This feature reduces the computational load at the cost of losing useful information in bifrequency plane.

3.3 Method C- Peaks, Peak Locations and Entropy

In this method, first diagonal slice is obtained. On this diagonal slice, peak amplitudes and their locations are recorded as $[\alpha_i, \omega_i]$ where *i=1, 2, 3.*Top three peaks (a_i) and their corresponding locations (ω_i) are recorded. Normalized entropy of the bispectrum is defined as [19]: $P_1 = -\sum_n p_n \log p_n$, where $p_n = |B_x(\omega_1, \omega_2)| / \sum_\Omega |B_x(\omega_1, \omega_2)|$ and normalized bispectrum squared entropy is defined as $P_2 = -\sum_i p_i \log p_i$, where $p_i = |B_x(\omega_1, \omega_2)|^2 / \sum_\Omega |B_x(\omega_1, \omega_2)|^2$, distance between two peaks is defined as *d*, impulse width at *-3dB* is denoted by *w* and energy of slice spectrum *E* is defined as $E = \sum_{i=1}^{M} |B_x(\omega_1, \omega_2)|^2_{\omega_1 = c\omega_2}$. The feature vector is defined as:

$$F = \{a_1, \omega_1, a_2, \omega_2, a_3, \omega_3, d, w, E, P_1, P_2\}. \tag{8}$$

In [18], authors have considered two frequencies corresponding to a peak. However, as can be seen from bispectrum, both frequencies will have same value due to *symmetry* property. Hence, we have considered only one frequency component in order to remove redundancy. This method has the advantage of having small computation time due to smaller dimension of feature vector.

4. HIGHER ORDER SINGULAR VALUE DECOMPOSITION THEOREM (HOSVD)

Here, HOSVD is proposed for feature extraction from bispectrum of infant cries. The higher-order singular value decomposition (HOSVD) theorem proposed in [19] is used to reduce the dimensionality of the feature space. HOSVD is a generalization of SVD applied to a tensor. Initially, all the features are *stacked* together one after another to form a *3-D* tensor *A*. The *2-D* feature, *i.e*., bispectrum is in the space $F \in R^{I_1 \times I_2}$ and let the number of samples be I_s .The tensor *A* (of dimension $I_1 \times I_2 \times I_s$) can be represented in HOSVD form as (as shown in Fig. 2):

$$A = S \times_1 U_{I_1} \times_2 U_{I_2} \times_3 U_S, \tag{9}$$

where *S* is the core tensor with the same dimension as *A*.

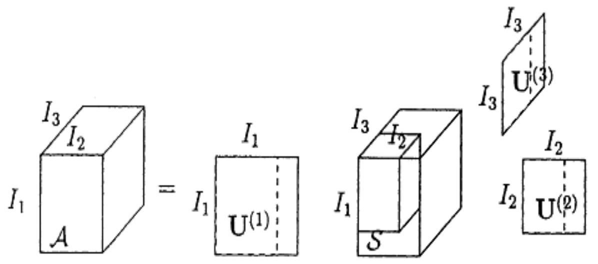

Fig 2. Visualization of HOSVD for tensor. After [19].

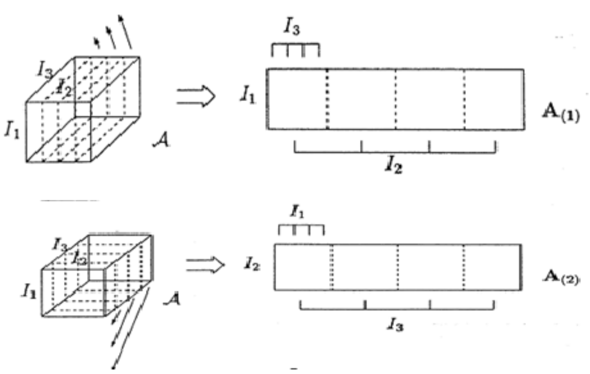

Fig. 3. Unfolding of tensor A to matrix A1 and matrix A2. After [19].

$U_{I_1} \in R^{I_1 \times I_1}$, $U_{I_2} \in R^{I_2 \times I_2}$ and $U_{I_s} \in R^{I_s \times I_s}$ are the *unitary* matrices of the corresponding subspaces of I_1, I_2 and I_s. The matrices U_{I_1} and U_{I_2} contains *n* mode singular vectors, *i.e.,*

$$U^{(n)} = [U_1^{(n)} \quad U_2^{(n)} \quad .... \quad U_{I_n}^{(n)}], \tag{10}$$

The matrices U_{I_1} and U_{I_2} can be obtained from the *matrix unfolding* of *A*. The unfolded matrices $A_{I_1} \in R^{I_1 \times I_2 I_S}$ and $A_2 \in R^{I_2 \times I_1 I_S}$ are obtained (as shown in Fig. 3) and they are decomposed in their *SVD* representations to give U_{I_1} and U_{I_2} . Only first R_1 and R_2 principal components are retained from these unitary matrices, respectively. Next, $\hat{U}_{I_1} \in R^{I_1 \times R_1}$ and $\hat{U}_{I_2} \in R^{I_2 \times R_2}$ are obtained, which gives dimension reduced feature set, *viz.*,

$$Z = B \times_1 \hat{U}_{I_1}^T \times_2 \hat{U}_{I_2}^T = \hat{U}_{I_1}^T . B . \hat{U}_{I_2} , \tag{11}$$

where $Z \in R^{R_1 \times R_2}$ and $B \in R^{I_1 \times I_2}$ which is taken from *A*.

5. EXPERIMENTAL SETUP AND RESULTS

Database: Infant cry data was collected with the help of senior doctors from Civil Hospital, Ahmedabad, India. In the present work, data was recorded at sampling frequency of *44.1 kHz* at *24 bits/sample*. In the database, there are *61* normal infant cries and *38* pathological cries of *87* infants (few days-1 year) were recorded from the hospital [20]. For database, only hunger and pain cries are collected. In the database, pathological cries include infant pathologies, *viz.*, upper respiratory tract infection (URTI), septicemia, malnutrition, any surgery, epilepsy, diarrhea, hydrocephalus, hypo calcium, congenital heart disease, jaundice, and bronchitis. This database is then divided into *train* and *test* dataset with a ratio of 75:*25*. From the cry signal, voiced segments are extracted using an *energy-based* algorithm (*viz., l^2* energy). The voiced data is segmented in frames of length *50 ms* each. Each frame is normalized by subtracting the mean then bispectrum is calculated.

Bispectrum is a two-dimensional (*i.e.*, *2-D*) feature of size *512* x *512* as shown in Fig. 1. It can be observed from Fig. 1, that bispectrum plot has *twelve symmetry* regions. The *symmetry* property of bispectrum follows from the properties of *moments* [16]. From Fig. 1, it can be observed that the information of bispectrum in either *first* quadrant or in *third* quadrant is sufficient to consider it as a *feature*. This saves the memory space and reduces the feature extraction time as well. In our experiments, we have considered information contained in third quadrant only. It reduces the feature size from *512* x *512* to *256* x *256* and then a tensor (per infant) is formed for both *train* and *test* datasets. On these tensors, HOSVD is applied which reduces the size of bispectrum from *128* x *128* to *10* x *10*. These features are then stored as logarithm of bispectrum feature vectors (*1* x *100*) for classification purpose. On the reduced feature set (*i.e., 1* x *100*), entropy is used as feature selection criterion. Using this criterion, feature size is reduced to [10 20 30 40 50 60 70 80 90 100]. On the reduced feature set, classification accuracy is determined using support vector machine (SVM) classifier with polynomial function kernel of order *3*. LIBSVM tool is used in our work [21].

In this work, classification accuracy is defined as the ratio of total number of correctly classified frames to the total number of frames in test dataset. We have *957* and *318* feature vectors of *45* and *16* infants, respectively, in the training and testing dataset of normal infant cries. In the pathological cries, we have *581* and *193* feature vectors of *28* and *10* infants for train and test datasets, respectively. Classification accuracy with different feature sizes has been determined and best classification accuracy is observed at feature vector size of *1*x*90*, which is *70.65 %*. Similar performance is observed for the proposed feature for radial basis function (RBF) kernel with $\gamma=0.01$ (classification accuracy of *70.86 %).*

Feature	Classification Accuracy (%)
MFCC	52.41
LPC	61.27
PLP	57.41
Bispectrum	**70.65**

Table 1. Classification accuracy (in %) with MFCC, LPC, PLP and bispectrum features.

From Table 1, we can infer that classification with proposed feature *outperform* with baseline features. MFCC gives a classification accuracy of *52.41 %,* LPC and PLP gives average classification accuracy of *61.27 %* and *57.41 %, respectively,* whereas bispectrum gives a classification accuracy of *70.65 %*. Bispectrum captures nonlinearity in signal as feature, hence it performs better. The HOSVD theorem retains the *principal* components of the bispectrum. Since the principal components are used as feature vector, performance is better than other existing methods. Confusion matrix for classification of normal and pathological infant cries using the proposed bispectrum- based features (for feature size *1*x *90*) is shown in Table 2. It can be observed that *72.64* % frames of normal infant cries and *65.66* % frames of pathological infant cries from test dataset is classified correctly.

	Identified as ⟶	**Normal**	**Pathological**
Actual	**Normal**	231	87
	Pathological	63	130

Table 2. Confusion matrix of classification of normal and pathological cries using bispectrum as a feature.

Feature	Feature Size	Average Accuracy (%)	Confidence Interval (%)
Method *A*	1x512	59.295	±4.23
Method *B*	1x512	60.425	±4.21
Method *C*	1x11	60.425	±4.21
Bispectrum	**1x90**	**70.65**	**±3.91**

Table 3. Comparison of classification performances (in %) of bispectrum features extracted from existing methods under same experimental setup.

Table 3 shows the classification performances of existing feature extraction methods with proposed method. Results show that proposed method of feature extraction using HOSVD gives excellent performance compared to existing methods along with a comparatively small feature dimension. Comparison with *method B* and *method C,* indicates that diagonal slice or features derived from it are not of sufficient importance in classification of normal and pathological infant cries. To quote *statistical significance* of our experimental results, *95 %* confidence interval is also reported. The confidence interval at feature size of *1* x *90* is ± *3.91 %* around classification accuracy of *70.65 %* which indicates better performance of the proposed features compared to state-of-the-art methods.

6. SUMMARY AND CONCLUSIONS

In this paper, it was found that features derived from bispectrum can classify normal and pathological infant cries better than the conventional state-of-the-art spectral features. The motivation behind using bispectrum is to use a feature which can capture nonlinearity in speech production mechanism as speech is *a non-stationary* process and speech production system is *nonlinear*. In addition, bispectrum has ability to detect noise in speech signal. In case of pathological cries, the amount of noise is even higher (which is also apparent from use of jitter and shimmer for pathological signal classification). It also imparts the important result that peaks and peak locations of bispectrum may not be always a good feature for classification task. Some of the pathologies considered here as disease, are not severe diseases that were not identified by the bispectrum features. Increasing the number of cry samples in the pathological database may increase the classification accuracy (in %). In future, we would like to apply this method for classification of adult pathologies.

REFERENCES

[1] Petroni, M.; Malowany, M.E.; Johnston, C. C.; Stevens, B. J., "A crosscorrelation-based method for improved visualization of infant cry vocalizations," Canadian Conf. on Electrical and Computer Eng., Halifax, NS, vol. no. *2*, pp. *453-456, 25-28* Sep *1994*.

[2] Petroni, M.; Malowany, M.E.; Johnston, C. C.; Stevens, B. J., "A new, robust vocal fundamental frequency (F_0) determination method for the analysis of infant cries," Proc. IEEE *7th* Symposium on Computer-Based Medical Systems, , Winston-Salem, NC, pp. *223-228, 10-12* Jun *1994*.

[3] Petroni, M.; Malowany, A. S.; Johnston, C. C.; Stevens, B. J., "A robust and accurate cross-correlation-based fundamental frequency (F_0) determination method for the improved analysis of infant cries," *IEEE 17th* Annual Conf. on

Engineering in Medicine and Biology Society , Montreal, Que, vol.*2*, pp. *975-976* , *20-23* Sep *1995*.

[4] Scalassara, P.R.; Dajer, M.E.; Marrara, J.L.; Maciel, C.D.; Pereira, J.C., "Analysis of voice pathology evolution using entropy rate," Tenth IEEE Inter. Symp. on Multimedia, ISM, Berkeley, CA, pp. *580-585, 15-17* Dec. 2008.

[5] Reyes-Galaviz, O.F.; Cano-Ortiz, S.D.; Reyes-Garcia, C.A., "Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies," *7*th Mexican Inter. Conf. on Artificial Intell. MICAI , Atizapan de Zaragoza, pp. *330-335, 27-31* Oct. 2008.

[6] Patil H.A., "Cry Baby": Using spectrographic analysis to assess neonatal health from an infant's cry, in A. Neustein (Ed.) "Advances in Speech Recognition, Mobile Environments, Call Centers and Clinics", Springer-Verlag, pp. *323-348, 2010.*

[7] Reyes-Galaviz, O.F.; Cano-Ortiz, S.D.; Reyes-Garcia, C.A., "Validation of the cry unit as primary element for cry analysis using an evolutionary-neural approach," Mexican International Conference on Computer Science, ENC, Baja California, pp*. 261,267, 6-10* Oct. 2008.

[8] Ole Wasz Hockert, Katarina Michelson and John Lind, Ch. 4 ―Infant Crying: Theoretical and Research Perspective, plenum pub. Corp, NY.

[9] Petroni, M.; Malowany, M.E.; Johnston, C. C.; Stevens, B. J., "Classification of infant cry vocalizations using artificial neural networks (ANNs)," Inter. Conf. on Acous., Speech, and Signal Proces .(ICASSP), Detroit, MI, vol.*5,* pp. *3475-3478 , 9-12* May 1995.

[10] Petroni, M.; Malowany, M.E.; Johnston, C. C.; Stevens, B. J., "A comparison of neural network architectures for the classification of three types of infant cry vocalizations," IEEE *17*th Annual Conference on Engineering in Medicine and Biology Society (EMBS), Montreal, Que, pp. *821-822* , *20-25* Sep *1995*.

[11] Garcia, J.O.; Reyes Garcia, C.A., "Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks," in Proc. of the Inter. Joint Conf. on Neural Networks, vol.*4*, pp. *3140-3145, 20-24* July *2003*.

[12] Lederman, D.; Cohen, Arnon; Zmora, Ehud; Wermke, K.; Hauschildt, S.; Stellzig-Eisenhauer, A., "On the use of hidden Markov models in infants' cry classification,", The *22*nd Convention of Electrical and Electronics Engineers in Israel, pp. *350-352*, 1 Dec. *2002.*

[13] Nikias, C.L.; Raghuveer, Mysore R., "Bispectrum estimation: A digital signal processing framework," Proc. of the IEEE , vol.*75*, no.*7*, pp. *869-891*, July *1987*.

[14] Nikias, C.L.; Mendel, J.M., "Signal processing with higher-order spectra," IEEE Signal Proc. Mag., vol.*10,* no.*3*, pp.*10-37*, July 1993.

[15] Higher order spectral analysis toolbox: http://www.mathworks.nl/matlabcentral/fileexchange/loadFile.do?objectId=3013&objectType=file {Last Accessed on 1st Jun., 2015}.

[16] Haitao Yu; Wang Yingmin; Zhanlin Xie; Wei Liu, "Feature extraction and classification based on bispectrum for underwater targets," Inter. Conf. on Intelligent System Design and Engineering Application (ISDEA), Changsha, vol.*1*, pp. *742-745, 13-14* Oct. *2010.*

[17] Ji Hongbing; Li Jie; Xie Weixin; He Wei, "Bispectrum-based radar target classification," 4th International Conference on Signal Processing Proc. (ICSP), Beijing, vol. *1*, pp. *419-422, 1998*.

[18] Li Shengjun; Liu Yi, "Feature extraction of lung sounds based on bispectrum analysis," *3rd* Inter. Symp. on Information Processing (ISIP), Qingdao, pp. *393-397, 15-17* Oct. 2010.

[19] Lathauwer, L. D., Moor, B. D. and Vandewalle, J. "A multilinear singular value decomposition". in SIAM J. Matrix Anal. Appl., vol. *21*, no.*4*, pp. *1253-1278, 2000*.

[20] Chittora A. and Patil H.A., "Data collection and corpus design for analysis of normal and pathological infant cry," in Oriental COCOSDA held jointly with International Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), New Delhi, pp. *1-6, 25-27* Nov. 2013.

[21] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: A library for support vector machines, in ACM Transactions on Intelligent Systems and Technology, 2*:27:1--27:27, 2011.* {Last Accessed on 1st Jun, 2015}.