On the Use of EMD for Automatic Newborn Cry Segmentation

Lina Abou-Abbas[1], Leila Montazeri[2], Christian Gargour[1], Chakib Tadj[1]

[1] Electrical Engineering department, École de Technologie Supérieure, Montreal, Canada
[2] Electrical Engineering Department, École Polytechnique, Montreal, Canada
Lina.abou-abbas.1@etsmtl.net, Leilamontazeri@hotmail.com

***Abstract*— Cry segmentation is an essential preprocessing step in any infant crying diagnosis system. Besides crying sounds consisting of expiration phases followed by short periods of inspiration episodes, each recording of newborn cries also includes silence sections as well as other sounds such as speech of caregivers, noise and sound of medical equipments. This paper is devoted to a newly developed Empirical Mode Decomposition (EMD) application to cry segmentation. The goal of the segmentation is to detect cry episodes automatically from unimportant acoustic activities existed inside the recorded signals. EMD decomposes a multicomponent non stationary signal into a set of monocomponent signals called Intrinsic Mode Functions (IMFs). The cry signals are segmented using Hidden Markov Models (HMMs) applied to the features extracted by employing EMD combined with Mel-Frequency Cepstral coefficients to the recorded cry signals. The performance of the proposed approach is evaluated on a database of 200 cry signals recorded in a real clinical environment. The experimental results demonstrate the effectiveness and suitability of the proposed method for the automatic cry segmentation.**

Keywords— Automatic cry segmentation; Empirical mode decomposition; Features extraction; Mel-frequency Cepstral coefficients; Classification; Hidden Markov Models. Introduction (*HEADING 1*)

I. INTRODUCTION

Although it might be hard sometimes to listen to infants crying, it is their only means of communication with their parents or caregivers, allowing them to express their needs for support. While cry might be considered as a simple behavior, it is actually, rather complicated. Studies show that newborns' cries take on different patterns depending on the reason of cries. Different features of newborn cries such as raising-falling pitch pattern, rapid pitch shifts, ascending-descending melody, high intensity or intensity lower than normal, length of the cry and other time relations, can give some hints concerning different pathological conditions [1-3]. Therefore, it becomes essential to put effort into newborn cry signal analysis.

Automatic cry segmentation serves as a primary step towards infant cry signal analysis applications such healthy and pathological classification. A segmentation system should be able to identify cry segments and distinguish them from other audio types. In most studies, cry segmentation has been made manually by experts. This technique, while highly effective, is time-consuming and cannot serve the needs of a real-time newborn diagnostic tool. Reviewing the literature, it has been found some efforts on automatic segmentation of cry sounds using different voice activity detection methods like zero crossing rate and short term energy [4-6], Harmonic Product Spectrum [7] and methods based on HMM classification [8, 9]. Unlike most of these previous works on this topic which uses a database composed only of cry sounds, this paper is meant for segmenting a database recorded in a real clinical environment. In this work, a novel approach which relies on EMD-based features extraction technique has been proposed for cry segmentation. EMD has been widely used and applied by researchers to different areas and has demonstrated to be effective in the domain of biomedical signal and speech signal processing [10-12].

In this framework, the input signal is hierarchically decomposed using empirical mode decomposition (EMD) and the given signal is divided into a set of Intrinsic Mode Functions (IMFs) and residual. Mel-frequency Cepstral coefficients are then extracted from the obtained IMF sequence through FFT algorithm and inverse cosine transform. The performance of the features extracted from the IMFs components is evaluated using Hidden Markov Models classifier. It has been demonstrated in the proposed approach that cepstral characteristics derived from IMFs components are useful to draw a distinction between voiced cries segments and other acoustical activities. Voiced cry segment is associated to the audible expiratory or inspiratory phase during the cry.

The rest of the paper is organized as follows. Corpus used in this paper is presented in section II. EMD technique is explained in section III while section IV provides description of the proposed approach. Results are presented in section V and the paper is finally concluded in section VI.

II. CORPUS

200 cry signals used in this study were recorded in the neonatology departments of several hospitals in Canada and Lebanon. In order to participate in the project, newborns had to have just born up to 6 month old, regardless of gender and prematurity. The corpus collection is still in its middle stage, and will continue for a while. The cry signals were recorded with a sampling rate of 44.1 KHz and a sample resolution of 16 bits. The recorder was placed at a distance of 10 to 30 cm from the baby's mouth. The details of the generated database are described in our previous work [8]. The recordings collected are produced by 64 babies including both healthy and sick cases, for a total of 200 signals, in different environments and conditions from silent to very noisy: the average duration of one signal is 90 seconds. To evaluate the proposed segmentation system, cry signals were segmented and labeled manually through combined of visual and auditory techniques by authors prior to the approach's development. Overall, our corpus had 66.5 % of voiced cry sounds and 33.5 % of other acoustic activities like speech, machines' sounds, noise, silence, etc.

978-1-4673-6516-1/15/$31.00 ©2015 IEEE

III. EMPERICAL MODE DECOMPOSITION

Huang et al. (1998) developed a novel decomposition technique which is known as empirical mode decomposition (EMD) and is used for analyzing nonlinear and non-stationary signals. EMD decomposes the given signal into a set of intrinsic mode functions (IMFs) and a residual function during a so-called sifting process. The main benefit of EMD is that basic functions are derived directly from the signal. An IMF represents simple oscillatory mode of the signal and is defined so as to draw position of the signal in time. An IMF must satisfy two basic criteria [13]:

- The number of zero crossing rate and the number of extremes in the whole data set are equal or differ by at most one.
- The mean value of the envelope defined by the local maxima and the envelope defined by local minima is zero at any point.

To extract an IMF from an observed signal, the following sifting approach is adopted [13]:

1) Identify the extrema of an observed signal $x(t)$ (local maxima and local minima separately)
2) Interpolate the local maxima to form the upper envelope $u(t)$
3) Interpolate the local minima to form the lower envelope $l(t)$
4) Calculate local mean value of the upper and lower envelopes $m(t)=\dfrac{u(t)+l(t)}{2}$
5) Retrieve the local mean value from the original signal $h(t)=x(t)-m(t)$
6) If h(t) satisfies the criteria to be an IMF, stop sifting else repeat the procedure with h(t) regarded as original signal.

This process could be repeated k times until $h(t)$ is an IMF, therefore: $c_1(t)=h_{1k}(t)$ is the first IMF of the original signal $x(t)$. The remainder or so-called residue $r_1(t)=x(t)-c_1(t)$ is regarded as the new signal to repeat the sifting process for the extraction of the second IMF and so on.

By summing IMFs and residue, the original signal can be reconstructed: $x(t)=\sum_{i=1}^{n} c_i(t)+r_n(t)$

In the figure 1 below the leading five IMFs are shown since all the cry signal components exist in these IMFs.

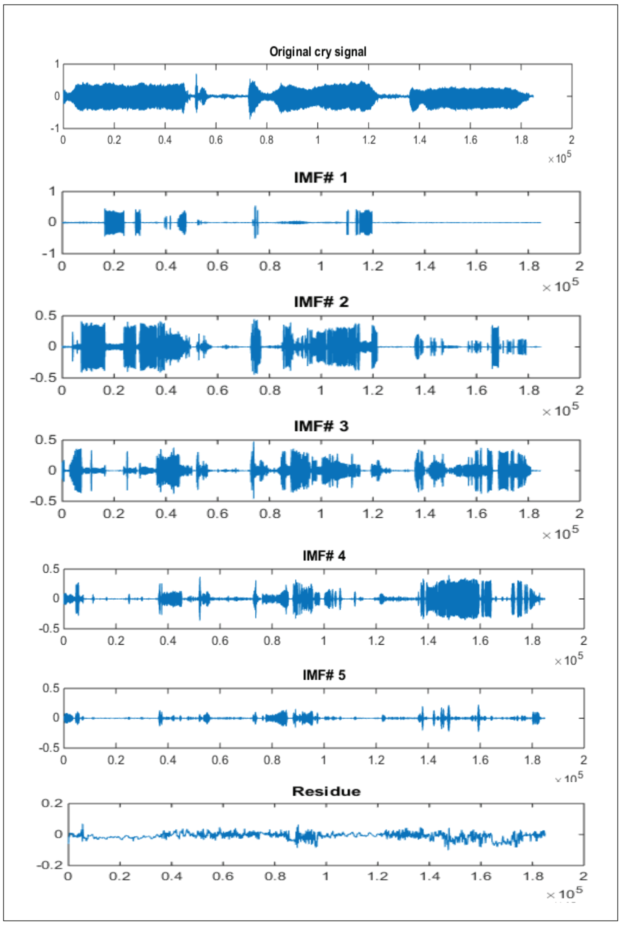

Fig. 1. EMD of a cry signal: from top to bottom, the original signal, five IMFs and the residue

IV. SYSTEM IMPLEMENTATION

The proposed segmentation system is divided into three essential stages: Empirical mode decomposition, features extraction and classification into two distinct classes: Voiced cry and other sounds. The block diagram of the proposed system is illustrated in figure 2.

To segment subjected cry signals, original signals have been broken down into several small frames of size 50ms with overlap of 35ms. Frames have been decomposed into five IMFs using EMD technique and following parameters: resolution of 50dB and residual energy of 40 dB. Once these IMFs obtained, sets of different IMFs are computed. The 12 Mel-Frequency Cepstral components and their respective energy are derived from the first five IMFs and from the different combination of IMFs. Note that the first IMF corresponds to the fastest oscillation while the last IMF corresponds to the slower one.

The stochastic analysis of the characteristics' space has been carried out using an HMM classifier of 5 states and 32 Gaussians in order to detect voiced cry segments among other acoustic activities. A training step has been achieved using the expectation maximization algorithm for the estimation of the maximum likelihood probabilities. The obtained results will be compared to the standard MFCC characteristics extracted directly from the original signal [8].

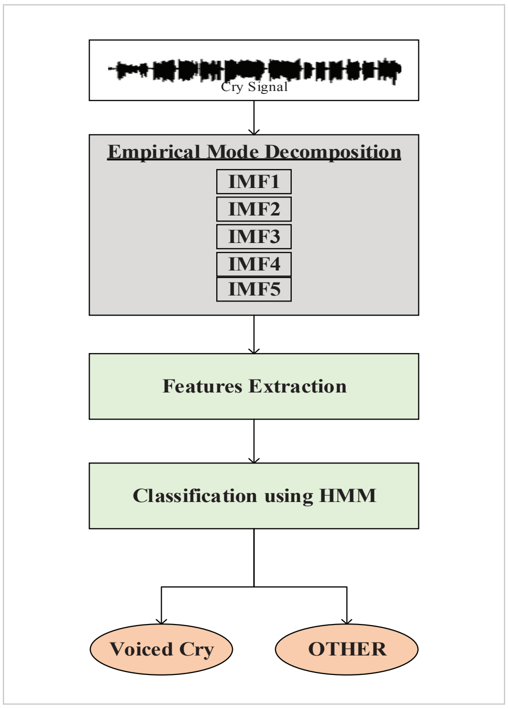

Fig. 2. Block diagram of the segmentation system

For the features extraction phase, signals obtained from different set of decompositions from one, two and three IMFs resulting in total 9 sets have been considered and presented in table 1. From each set, 39 MFCC have been extracted and then applied to HMM for the classification stage. Two groups of recordings were used. A training group used to develop and optimize the model and a testing group used for test. Both groups included recordings contaminated with various types of noise such as respiration sounds, speech and machine equipments. Experiments were designed in order to identify the best IMF combinations.

V. RESULTS

As we have considered a classification approach that implies training and testing stages, we have decomposed our database equally into two parts: 50% for the training stage and 50% for the testing stage. To evaluate the performance of our system, a comparison between the automatic transcription file and the reference file (manually segmented) is done. Three types of errors are calculated: insertion (I), deletion (D) and substitution (S). Readers are referred to our previous work [8] for more details about these errors. The accuracy rate of the system could be estimated as follows:

$$AR = \frac{N - D - I - S}{N} \times 100\%$$

Where N represents the total number of labels in the reference file.

Table 1 List of different EMD decomposition used

Set 1	IMF1
Set 2	IMF2
Set 3	IMF3
Set 4	IMF4
Set 5	IMF5
Set 6	IMF34=IMF3+IMF4
Set 7	IMF45=-IMF4+IMF5
Set 8	IMF234=IMF2+IMF3+IMF4
Set 9	IMF345=IMF3+IMF4+IMF5

The results presented in table 2 represent that our approach applied to IMF3+IMF4+IMF5 have an improvement of accuracy rate of 8.76% compared with the standard one which is: signal without EMD considered in a previous work [8]. We considered each IMF separately. We noted that IMF3, IMF4 and IMF5 contained more discriminant information than the other IMFs and allowed some improvement of accuracy rate in comparison with the standard approach without EMD. The summation of the IMF3 or IMF5 to the IMF4 allows also some improvement in comparing to the other IMFs considered separately. From other side, the summation of three IMFs: 2, 3, 4 and 3, 4, 5 was considered and we noted that the combination of IMF345 achieves the best compromises for the accuracy rate.

Table 2 Accuracy Rate of the overall system for the different experiments

Experiments	Accuracy Rate
Without EMD	77.93
IMF1	68.36
IMF2	73.85
IMF3	77.26
IMF4	79.84
IMF5	78.62
IMF3+IMF4	84.05
IMF2+IMF3+IMF4	79.6
IMF4+IMF5	83.62
IMF3+IMF4+IMF5	86.69

VI. CONCLUSION

This paper introduces a novel Empirical Mode Decomposition based decision HMM approach for crying signals segmentation. EMD deals with nonlinear and nonstationary signals by decomposing them into intrinsic Mode Functions. Features extracted from the IMFs of crying signals have been found useful in detecting the voiced cry parts.

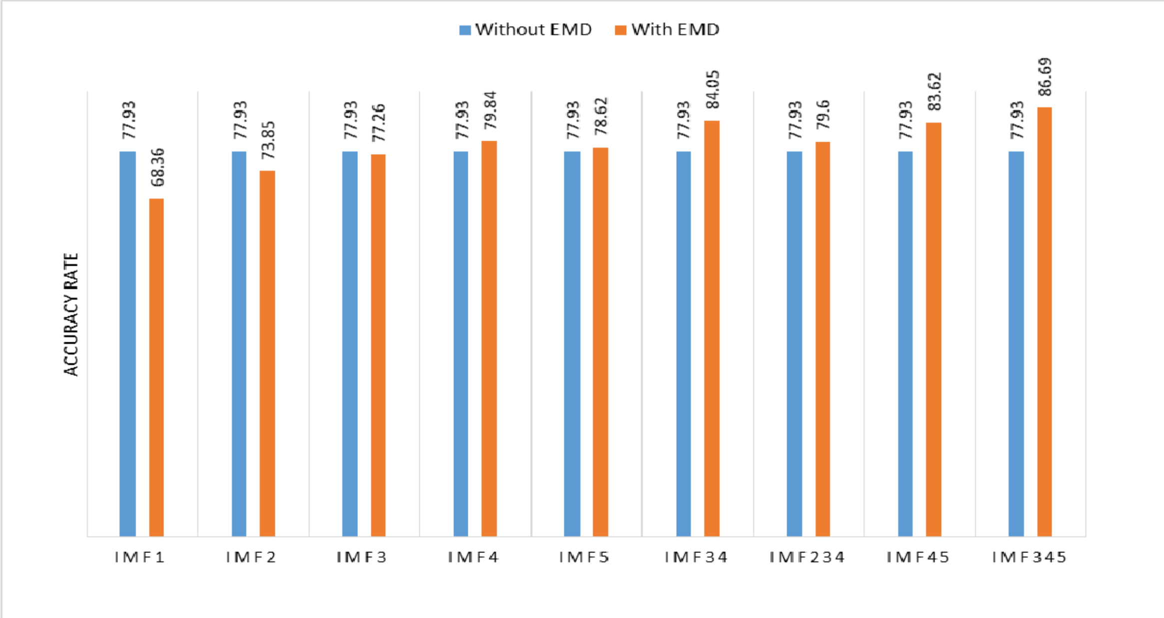

Fig. 3 Performance evaluation of segmentation of crying signals- Comparison between experiments based on EMD and without EMD

We concluded that the combination of EMD with MFCC analysis gives interesting features for the classification of voiced crying parts and other acoustics parts. The classification results indicated that studied approach had provided 86.69% accuracy. Future direction of research may include application of EMD combined with wavelet to improve the accuracy of the segmentation system.

Acknowledgment

This work is supported by Bill and Melinda gates foundation. The authors thank the personnel at the Department of neonatology at Sainte Justine hospital and Sahel Hospital for their cooperation in the data collection process.

References

[1] H. L. Golub, "A physioacoustic model of the infant cry and its use for medical diagnosis and prognosis," *The Journal of the Acoustical Society of America,* vol. 65, pp. S25-S26, 1979.

[2] J. A. Davis, "Pathological Cry, Stridor, and Cough in Infants," *Archives of Disease in Childhood,* vol. 58, pp. 319-320, 1983.

[3] J. Orozco-García and C. Reyes-García, "A Study on the Recognition of Patterns of Infant Cry for the Identification of Deafness in Just Born Babies with Neural Networks," in *Progress in Pattern Recognition, Speech and Image Analysis*. vol. 2905, A. Sanfeliu and J. Ruiz-Shulcloper, Eds., ed: Springer Berlin Heidelberg, 2003, pp. 342-349.

[4] Rui, x, M. A. z, L. C. Altamirano, C. A. Reyes, and O. Herrera, "Automatic identification of qualitatives characteristics in infant cry," in *Spoken Language Technology Workshop (SLT), 2010 IEEE*, 2010, pp. 442-447.

[5] A. Zabidi, W. Mansor, K. Lee Yoot, R. Sahak, and F. Y. A. Rahman, "Mel-frequency cepstrum coefficient analysis of infant cry with hypothyroidism," in *Signal Processing & Its Applications, 2009. CSPA 2009. 5th International Colloquium on*, 2009, pp. 204-208.

[6] K. Kuo, "Feature extraction and recognition of infant cries," in *Electro/Information Technology (EIT), 2010 IEEE International Conference on*, 2010, pp. 1-5.

[7] G. Várallyay, A. Illényi, and Z. Benyó, "Automatic infant cry detection," in *MAVEBA*, 2009, pp. 11-14.

[8] L. Abou-Abbas, H. Fersaie Alaie, and C. Tadj, "Automatic detection of the expiratory and inspiratory phases in newborn cry signals," *Biomedical Signal Processing and Control,* vol. 19, pp. 35-43, 5// 2015.

[9] L. Abou-Abbas, H. Fersaei Alaei, and C. Tadj, "Segmentation of voiced newborns' cry sounds using Wavelet Packet based features," presented at the Electrical and Computer Engineering (CCECE), 2015 IEEE 28th Canadian Conference Halifax, Canada, 2015.

[10] V. Bajaj and R. B. Pachori, "EEG signal classification using empirical mode decomposition and support vector machine," in *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011*, 2012, pp. 623-635.

[11] R. J. Martis, U. R. Acharya, J. H. Tan, A. Petznick, R. Yanti, C. K. Chua, *et al.*, "Application of empirical mode decomposition (emd) for automated detection of epilepsy using EEG signals," *Int J Neural Syst,* vol. 22, p. 1250027, Dec 2012.

[12] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis,* vol. 1, pp. 1-41, 2009.

[13] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 1998, pp. 903-995.