International Journal of Research in Advent Technology (IJRAT) Special Issue
E-ISSN: 2321-9637
Available online at www.ijrat.org
*International Conference "*INTELINC 18*", 12th & 13th October 2018*

Gaussian Mixture Model Based Automatic Segmentation of Infant Cry Signal

S.Pavithra [1], G.Yamuna[2], T.Ravichandran[3],R.Durga[4]
[1]PG scholar, [2]Professor and Head, [3]Associate Professor,[4] Assistant Professor
[1]mspavithra1@gmail.com,[2] yamuna.sky@gmail.com,[3]travi132@yahoo.com,[4] rdurgaphd@yahoo.com
[1]9655879389,[2]9842399987,39629880680,[4]9787213522
Department of ECE, Annamalai University

Abstract: - Impulse extraction of audio regions of concern from footages grabbed in exposed health conditions is an essential pre-processing step in few cry analysis scheme. The baby cry indicator is assistance to carrel the preschool age to age actions made on the classes of the signal. Autism Spectrum Disorder (ASD) insurances a varied scale of indicators with the core ones reading to difficulties with common announcement and interface. The adaptive wiener filter is pragmatic to develop the baby cry indication. Audio inquiry of baby cry indicator has developed as a capable area, flat for pre-verbal offspring. As such, age, drive and voice superiority related types have been explored for primary ASD judgment. In this projected work, an application for response cry wide-ranging separation for submission in a cry-based diagnostic scheme has been wished-for. The boldness of a Gaussian mixture model/hidden Markov model (GMM/HMM)-based cry separation arrangement is studied by the contributions of time and frequency domain. A fully programmed system is grounded in to two methods: arithmetical scrutiny created on GMMs or HMMs classifiers and a post-processing scheme based on force, zero crossing rate, and frequency for extracting cry sound components namely expiration and inspiration.

Key Words: Autism Spectrum Disorder (ASD), Gaussian mixture model/hidden Markov model, feature extraction.

I. INTRODUCTION

Gratified investigation of all-purpose aural is afraid with inquiry of audial for unthinking mining of appropriate evidence such as audio scene appearances and sound sources present. Tenders series from simple sorting jobs for aiding milieu sentience, to retreat and observation organisms based on identified noises of interest, to party, indexing, group and quizzing of large audio catalogues. In our regular lives, we come upon a rich assortment of sound trials such as mutt yelp, footsteps, flute tinkle and roar. Sound event detection (SED), or audio event detection, deals with the mindless empathy of these sound trials.

The goal of SED is to notice the commencement and balance times for each sound incident in an audio recording and assistant a documentary descriptor, i.e., a label for every of these trials. It has been diagram a heaving amount of awareness in recent years with submissions comprising audio following, healthcare checking, built-up sound analysis, software event recognition and hen call recognition. In the fiction the lexicon varies between poets; communal terms being thorough event detection, appreciation, labeling and sorting. Sound dealings are sharp with

scheduled labels called rigorous event classes. In our slog, sound event cataloguing, sound event appreciation, or sound event labeling.

All refer to classification an aural soundtrack with the complete affair modules extant, nonetheless of the onset/equipoise spells. On the extra hand, an SED task embraces together onset/offset recognition for the modules existing in the demo and arrangement indoors the appraised onset/offset, which is certainly the requirement in a real-life set. Sound trials often ensue in amorphous environs in real-life. Causes such as conservation clamor and covering springs are contemporaneous in the amorphous milieus and they can host a high unit of deviation midst the thorough dealings from the same rigorous incident class. Also, there can be several sound causes that food sound events going to the equal class, e.g., a dog bark thorough event can be formed from some breeds of dogs with changed acoustic faces.

II. RELATED WORK

A solitary capable of being heard cry can ordinarily contain both voiced and unvoiced portions so The issue of cry division/location can't be viewed as an issue of voiced/unvoiced. VAD modules are not ready to understand the cry accounts caught in clinical condition which is the fundamental technique utilized in cry examination system.(Kuo, 2010; Ruiz et al., 2010; Várallyay, 2006; Zabidi et al., 2009). VAD just finds the discourse locale from other acoustic area of a sound flag.

The other regions include noise,silence or an alarm signal. However, the signal-to-noise ratio (SNR) is an important parameters which results in errors.In audio communication systems such as automatic speech recognition, mobile phones, personal digital assistants, and real-time speech transmission, VAD is mostly used. The two modules of VAD are (1) feature extraction and (2) decision rules features depends on energy of the signal, cepstral coefficients, zero-crossing rate (ITU, 1996), (Marzinzik and Kollmeier, 2002), entropy and wavelet transforms(Wang and Tasi, 2008; Juang et al., 2009).

The further auditory commotion provinces can be any type such as clamor, calm, or an alarm caution. Still, the signal-to-noise ratio (SNR) denotes a crucial constraint and may outcome in chief errors. VAD is an crucial part of countless audio statement structures.The decision rule is based on threshold setting.

VAD systems used in G.729b and the Rabiner-Sambur system to detect cry slices (Benyassine et al., 1997; Rabiner and Sambur, 1975)where the setting of an threshold is not an easy task because cry and speech segments are not distinguished and also expiratory phase from inspiration phase will not be identified.

Algebraic methodologies seems to be a good clarification. Thats why present works are based on Algebraic methodologies (Abou-Abbas et al., 2015b; Abou-Abbas et al., 2015c). Cry segmentation will be of low consideration than speech and music.

We stated the art conversed in our topical mechanism (Abou-Abbas et al., 2015b; Abou-Abbas et al., 2015a).Only cries from silent pauses or respiratory phases are separated using the exsisting algorithms(Várallyay et al., 2008, 2009).

The databases are collected from laboratory and not from a clinical environment so the noise or other acoustic regions are included.

Due to the limited availability of resources only three classes expiration, inspiration, and silence are considered. Gaussian mixture models (GMMs) classifiers were used, to train the modules later Viterbi process was added to study the whole planning as HMM.

Two-dimensional linear frequency cepstral coefficients (STDLFCCs), based on linear frequency cepstral coefficients (LFCCs)was investigated at the work of Kim et al. (2013) .

Here low frequency as well as high frequency will be captured from the crying segments which is used for the better discrimination between cry and non-cry segments.
In Yamamoto et al. (2010, 2013), baby's say was detected using JULIUS a speech gratitude classification and a detection rate of 69.4% was achieved.

III. PROPOSED METHODOLOGY

The extant study is an scrutiny of shrieks flanked by pre-schooler's at high risk and low risk for ASD, to study if there is an audio quantity which may well act as an first risk gage.

The new-born cry indicator is cast-off to shape the new-born era to era events created on the skins of the indicator. Autism Field Disorder (ASD) covers a wide variety of indicators with the core ones concerning to hitches with social statement and interface.

In this wished-for work, The boldness of a Gaussian mixture model/hidden Markov model (GMM/HMM)-based cry separation arrangement is calculated by the contributions of time and frequency domain.

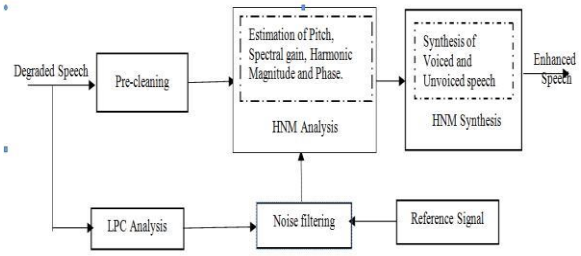
A fully programmed system is grounded in to two methods: arithmetical scrutiny created on GMMs or HMMs classifiers and a post-processing scheme based on force, zero crossing rate, and frequency for extracting cry sound components namely expiration and inspiration.

3.1 Suckling Cry signal development using Adaptive Wiener filter

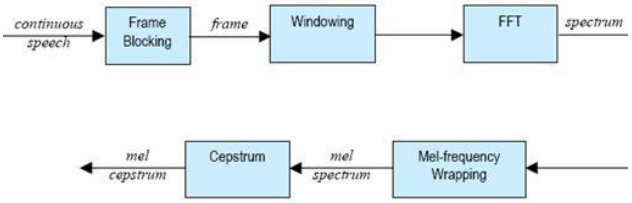
Toddler cry indicator development is a central research area, which is used to increase the tarnished speech. Model based speech development procedure, boosts the noisy speech using an analysis-synthesis agenda. Speech enrichment using ethereal clarifying will result in tuneful noise. Melodious clatter is very infuriating to anthropological hearers.

The ear-splitting indication is pre-cleaned via MMSE Estimator and then the
adaptive wiener filter analyses the piercing Infant cry signal and guesstimates the vital frequency (pitch), ethereal gain, the sung
amplitude, sung phase and the enunciated/hidden mixing. Adaptive wiener screen Amalgamation redoes the bull Pre-schooler cry signal. Ethereal packet guesstimate is upgraded by adaptive wiener which completes a restored speech enrichment when related with adaptive Wiener filter.

This process reservations the speech quality and diminishes the enduring clamour and results in lovely speech for anthropological listeners. The momentous saving in tuneful noise is shown with the support of spectrogram judgments. Based on the routine calculation of spectrogram and objective procedures (LSD, SNR, PESQ), it is unproven that the adaptive wiener filter process achieves weighty enhancement over unadventurous methods at changed blast ranks.



3.2 Feature Extraction



To increase the precision and adeptness of the mining methods, communication signals are customarily pre-processed already countryside are take out. Tongue indicator pre-processing shelters ordinal sieving and talking gesture exposure.

3.2.1 Pre-emphasis filter

Pre emphasis should be done to reduce the high dynamic range and noise that affects the speech signal. The first-order FIR high-pass filter will do this emphasis. In the time realm, with input $x[n]$ and $0.9 \leq a \leq 1.0$, the filter equation
$y[n] = x[n] - a \,.\, x[n-1]$.

An adaptive or fixed coefficient filter is used to implement the pre-emphasizer, depending upon the autocorrelation values the coefficients are adjusted with time.

The amount of energy in high frequencies will get boosted in this stage where it is affected by the glottal pulse naturally. Energy present in the higher frequencies will have information about the higher formants which is now available for the acoustic model. Before Windowing, pre-emphasis should be done to the input signal.

3.2.2 Framing and windowing

Instead of analysing the whole speech signal at a time, we split the signal in to many number of frames, so that it can be analysed in short time.
Of range 10-30ms because most of the speech signal are stationary. The overlapping will be done to these frames where hop-size are introduced. Hop size is defined as half of the frame size.

3.2.3 Windowing

It is necessary to work with short term or frames of the indication. This is to first-rate a slice of the indicator that can judiciously be supposed inactive. Windowing is made to sidestep unnatural discontinuities in the speech piece and falsehood in the basic gamut. The special of the hole-in-the-wall is a balance amongst some factors. In orator appreciation, the most generally used space shape is the con window.

The development of the tongue wave by the hole function has two things:-

1-It steadily diminishes the breadth at both locks of hair of pulling out pause to avoid an rapid modification at the endpoints. 2-It harvests the involvedness for the Fourier convert of the opening job and the dialog field.

The use for hamming windows is unpaid to the datum that mfcc will be used which comprises the regularity province (hamming openings will lessening the leeway of tall regularity gears in each mount due to such rapid slicing of the signal).

3.2.4 Fast Fourier Transform

To renovate the indicator as of time sphere to occurrence realm preparing to the succeeding stage (Mel frequency wrapping). The basis of performance Fourier transform is to alter the involvedness of the glottal pulse and the verbal tract whim retort in the stint province into reproduction in the rate province.

Super natural study shows that altered qualities in speech indicators links to unalike energy spreading over rates. Thus we habitually perform FFT to attain the degree rate comeback of each frame.

When we make FFT on a mount, we accept that the gesture in the interior a mount is interrupted, and nonstop when wrap from place to place. If this is not the event, we can still make

FFT but the in continuousness at the frame's first and last ideas is prospective to introduce undesirable effects in the frequency response. To deal with this unruly, we have two policies:

Increase each border by a Sham window to growth its permanence at the leading and last points. Take a border of a inconstant mass such that it always comprehends an digit many numeral of the major periods of the tongue indicator.

The second approach chance meeting difficulty in preparation since the documents of the ultimate passé is not a minor unruly. Also, hidden thuds do not have a essential historical at all. Accordingly, we habitually espouse the main approach to reproduce the mount by a Hamming window earlier performance FFT (as mentioned before).

3.2.5 Mel-scaled filter bank

Tones with different frequencies are present in the speech signal. Frequency which is measured in HZ and pitch measured in MEL scale for every tone of the speech signal. A linear spacing below 1KHz and a logarithmic spacing above 1KHz will be in a mel frequency scale.

mel(f)= 2595*log10(1+f/700).

3.2.6 Mel frequency analysis

Mel-Frequency examination of dialogue is based on anthropological discernment investigates**.** Human ears, for regularities lower than 1 kHz, hears tendencies with a linear scale instead of logarithmic scale for the regularities developed than 1 kHz. The material agreed by low regularity gears of the dialog signal is more vital related to the tall rate gears.

In command to residence extra underline on the low regularity workings, mel clambering is achieved. Mel filterbanks are non-uniformly set apart on the occurrence bloc, so we obligate extra screens in the low incidence counties and less figure of screens in high regularity districts.

We useful mel mesh banks , the indicator managed in such left like that of anthropoid ear answer:

$$\tilde{S}(l) = \sum_{k=0}^{N/2} S(k) M_l(k)$$

Where :

S(l) :Mel spectrum.

S(K) :Original spectrum.

M(K) :Mel filterbank.

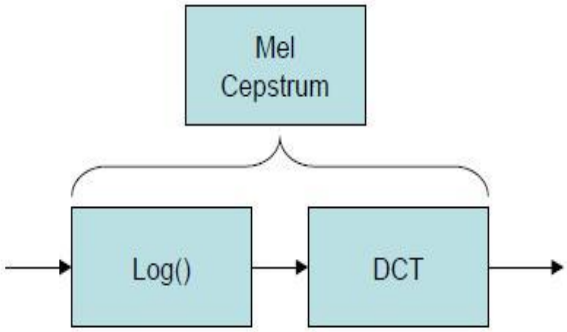
L=0 ,1 , ..., L-1 , Where L is the over-all digit of mel filterbanks

N/2 = Half FFT size.

3.2.7 Cepstrum

In the ultimate step, (MFCCs) are calculated where the log mel band has to be changed back to time. The result is called the mel regularity cepstrum numbers (MFCCs).

The local spectral properties of a signal will be given by Cepstral representation.. For the mel spectrum numbers are real numbers(and so are their logarithms), they could be new to the interval dominion using the Disconnected Cosine Transform (DCT).

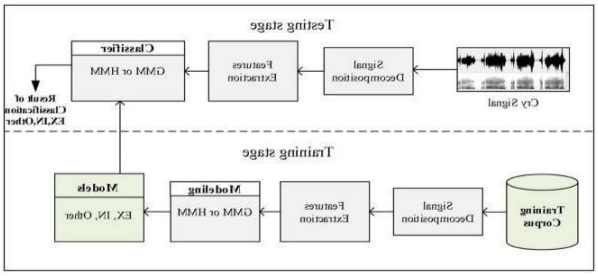


Since the dialog signal embodied as a involvedness flanked by slowly changeable vocal tract whim rejoinder (filter) and hurriedly changeable glottal pulse (source), so, the dialog field consists of the ethereal incase(low frequency) and the ethereal niceties(high frequency).

At the end of mfcc :
After the former argument we set the ensuing to gizmo the mfcc system:

1-pre-mphasis.
2-framing with frame size=256 sample.
3- Windowing by increasing each border with hamming window.
4-Using 40 mel filterbanks.
5-Extracting 12 mfc numbers for every one frame.

3.3 GMM/HMMMODEL FOR SEGMENTATION



In command to scrutinize toddler cry cassettes arrested in true locations having meddlesome fonts, the objective is to sector cry signal in to three classes inspiratory, expiratory and residual. All other areas of the acoustic region other than the area covered by the above two classes are referred as residual.

For segmentation Hidden Markov models (HMM) with Gaussian mixture model (GMM) are used as an pattern recognizer. In automatic speech recognition system HMM are broadly used to find the unpredictability in dialogue caused by unlike speakers, dialogue styles, vocabularies, and settings. Each cry soundtrack are divided in to frames and windowed overlapped as short time.

For every one such frame,the set of probabilities will be given as an output by HMM pattern recognizer. Viterbi system will decode these prospectus. It finds the best path from the HMM states that gives maximum likelihood of features used for cry investigation.

Finally the region of interest will be identified from a output class label given for each frames(i.e., expiratory and inspiratory phases) in the cry signal. Thus Feature extraction, HMM training and viterbi decoding are the three stages for overall implementation.

3.4 Classification of Infant cry signal to justify ASD

We done network working out by improving clear-cut cross-entropy. The grid bulks were adjusted with Glorot unbroken data, and stochastic slope origin (SGD) with Nesterov thrust was used as an optimizer, with a erudition rate of 0.02. We decremented the knowledge rate by 0.0001 for each era and the mini-batch size was 128. Arbitrarily particular fifteen in a hundred of the working out figures be situated used for endorsement. To foil the web from over-fitting the training data, we done training after the confirmation loss reassured; i.e., whereby the endorsement loss unveiled no further decrease during 15 epochs. The convolutional neural grid will give the conclusion manufacture on ASD.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this segment, we contemporary the MATLAB fallouts for gauging the performance of our proposed mechanism. The estimation of CNN, GMM and HMM methods are showed using the hyper constraints. We offer the mean and the typical deviation of these experimentations in this segment. Best performance system is decorated with bold aspect in the slabs of this slice. The

procedures whose best show midst the ten runs is indoors one customary abnormality of the best acting.

Audio 1: **Pressure Level in frequency domain 141.5466**

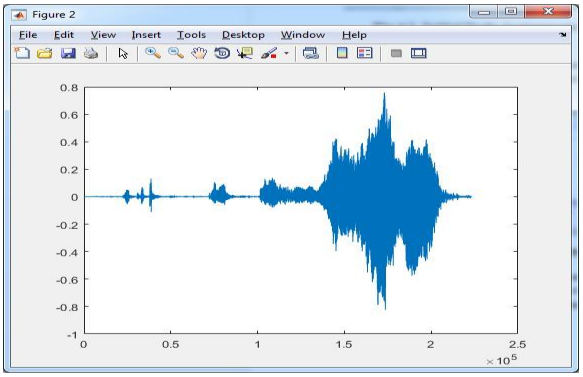


Fig.4.1. Input Audio Signal

Fig. 4.1 shows the Input Audio Signal, with n number of samples and amplitude.

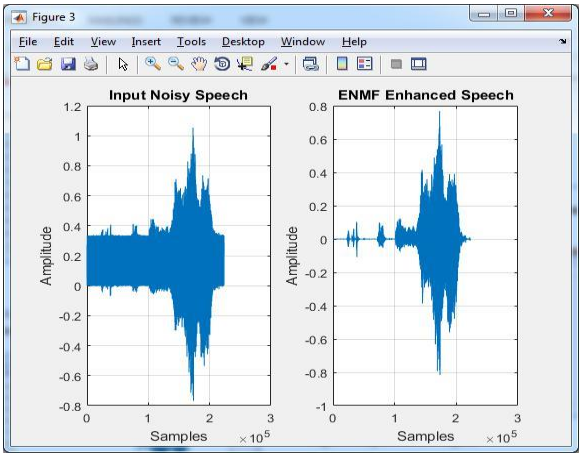


Fig.4.2. Audio Rate

Fig.4.2. shows the Input Noisy Speech and ENMF Enhanced Speech

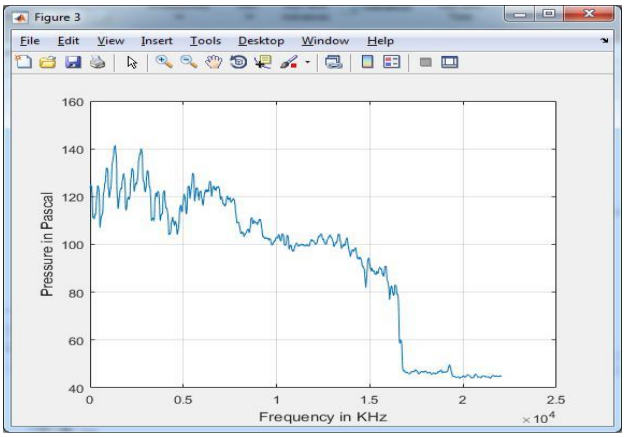


Fig.4.3 Pressure Pascal

Fig. 4.3.shows the Pressure Pascal and Frequency level of input audio file

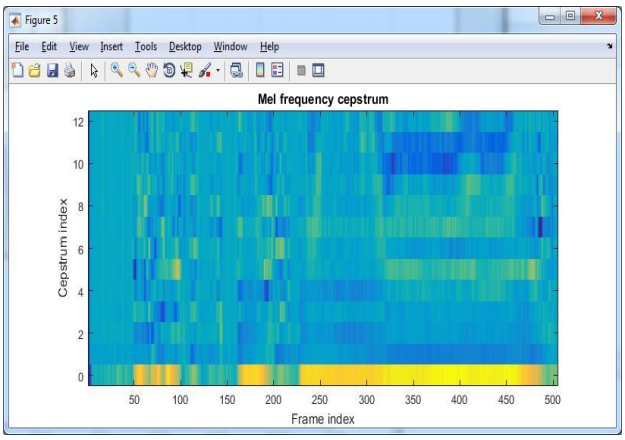


Fig.4.4 Mel Frequency Cepstrum

Fig. 4.4.shows the Mel Frequency Cepstrum of input audio file

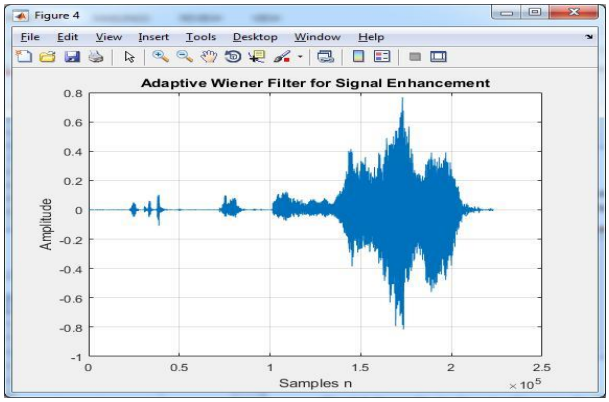


Fig.4.5 Adaptive Wiener for Signal Enhancement

Fig. 4.5.shows the Adaptive Wiener for Signal Enhancement of input audio file

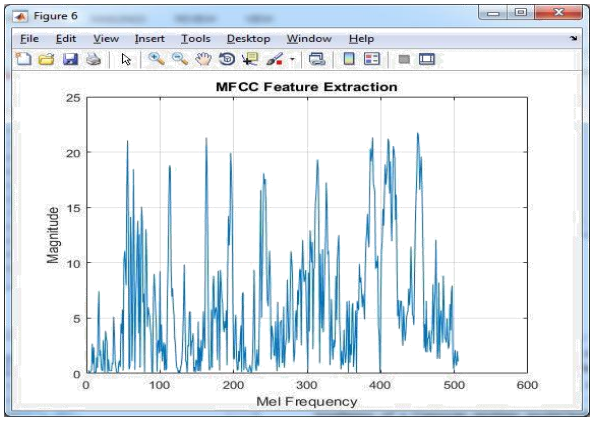


Fig.4.6 MFCC Feature Extraction

Fig. 4.6.shows the MFCC Feature Extraction of input audio file

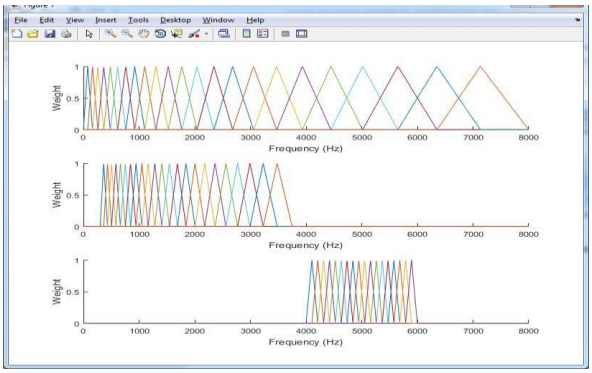


Fig.4.7 Weight and Frequency

Fig. 4.7.shows the Weight and Frequency of input audio file

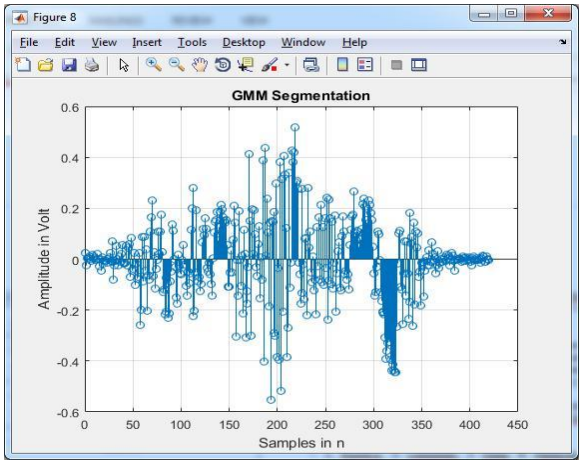


Fig.4.8 GMM Segmentation

Fig. 4.8. GMM Segmentation of input audio file

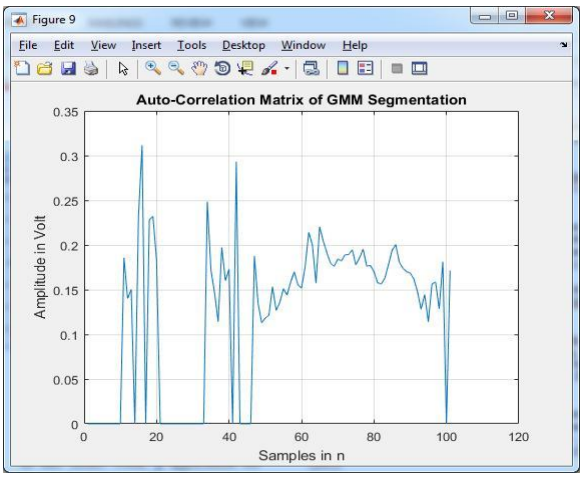


Fig.4.9 Auto Correlation Matrix of GMM Segmentation

Fig. 4.9. Auto Correlation Matrix of GMM Segmentation of input audio

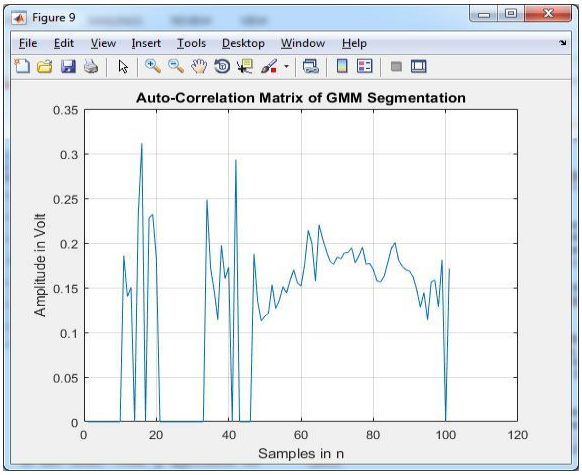


Fig.4.9 Auto Correlation Matrix of GMM

Segmentation

Fig.4.9.Auto Correlation Matrix of GMM Segmentation of input audio

V. CONCLUSION

In this frame work, A cry based diagnostic system has been proposed where the cry sound will be automatically segmented.. The boldness of a Gaussian mixture model/hidden Markov model (GMM/HMM)-based cry separation arrangement is studied by the contributions of time and frequency domain. We have used the GMM for conservation sound acknowledgment by means of dataset, and it reached correctness that bettered state-of-the-art convolutional neural webs using more bounds. Future work will study changed demand n from turn to turn the layers and subterranean GMM/HMM styles. We will also ponder rub on the GMM/HMM to time-based indications other than comprehensive.

REFERENCES

[1] MJ Corwin, BM Lester, C Sepkoski, M Peucker, H Kayne, HL Golub, Newborn acoustic cry characteristics of infants subsequently dying of sudden infant death syndrome. Pediatrics.96(1), 73–77 (1995)

[2] HL Golub, MJ Corwin, Infant cry: a clue to diagnosis. Pediatrics.69(2), 197–201 (1982)

[3] C Manfredi, M D'Aniello, P Bruscaglioni, A Ismaelli, A comparative analysis of fundamental frequency estimation methods with application to pathological voices. Med. Eng. Phys.22(2), 135–147 (2000)

[4] K Michelsson, O Michelsson, Phonation in the newborn, infant cry. Int. J. Pediatr. Otorhinolaryngol.49, 297–301 (1999)

[5] G Esposito, P Venuti, M Bornstein, Assessment of distress in young children: a comparison of autistic disorder, developmental delay, and typical development. Res. Autism Spectrum Disorders.5(4), 1510–1516 (2011)

[6] LL LaGasse, AR Neal, BM Lester, Assessment of infant cry: acoustic cry analysis and parental perception. Ment. Retard. Dev. Disabil. Res. Rev. 11(1), 83–93 (2005)

[7] R.Durga,G.Yamuna and R.Barkavi(2018),,Video segmentation using short term hierarchial fastwatershed algorithm,Proceedinhgs of ieee international conference on

[8] Communication & signal processing ICCSP'18,pp.285-289 Apr 3-5,2018

[9] P Zeskind, M McMurray, CL ET, K Grewen, K Garber, J Johns, Translational analysis of effects of prenatal cocaine exposure on human infant cries and rat pup ultrasonic vocalizations. PloS ONE.9(10), 110349–110349 (2013)

[10] A Fort, C Manfredi, Acoustic analysis of newborn infant cry signals. Med. Eng. Phys.20(6), 432–442 (1998)

[11] BM Lester, Developmental outcome prediction from acoustic cry analysis in term and preterm infants. Pediatrics. 80(4), 529–534 (1987)

[12] S Orlandi, C Manfredi, L Bocchi, M Scattoni, inProc. Annual Intl. Conf. IEEE Engineering in Medicine and Biology Society. Automatic newborn cry analysis: a non-invasive tool to help autism early diagnosis (IEEE, 2012), pp. 2953–2956

[13] AM Chilosi, P Cipriani, C Pecini, D Brizzolara, L Biagi, D Montanaro, M Tosetti, G Cioni, Acquired focal brain lesions in childhood: effects on development and reorganization of language. Brain Lang.106(3), 211–225 (2008)
[14] F Cunha, JJ Heckman, The economics and psychology of inequality and human development. J. Eur. Econ. Assoc.7(2-3), 320–364 (2009)
[15] PS Douglas, PS Hill, A neurobiological model for cry-fuss problems in the first three to four months of life. Med. Hypotheses.81(5), 816–822 (2013)
[16] M Hadders-Algra, Challenges and limitations in early intervention. Dev. Med. Child Neurol.53(s4), 52–55 (2011)
[17] KE Pape, Developmental and maladaptive plasticity in neonatal SCI. Clin. Neurol. Neurosurg.114(5), 475–482 (2012)
[18] AM Goberman, MP Robb, Acoustic examination of preterm and full-term infant cries: The long-time average spectrum. J. Speech Lang. Hearing Res.42(4), 850–861 (1999)
[19] K Wermke, MP Robb, Fundamental frequency of neonatal crying: does body size matter? J. Voice.24(4), 388–394 (2010)
[20] A Branco, SM Fekete, LM Rugolo, MI Rehder, The newborn pain cry: descriptive acoustic spectrographic analysis. Int. J. Pediatr. Otorhinolaryngol.71(4), 539–546 (2007)
[21] FL Porter, SW Porges, RE Marshall, Newborn pain cries and vagal tone: parallel changes in response to circumcision. Child Dev.59, 495–505 (1988)
[22] A Messaoud, C Tadj, inProc. Intl. Conf. on Image and Signal Processing.A cry-based babies identification system (Springer, 2010), pp 192-199
[23] B Reggiannini, SJ Sheinkopf, HF Silverman, X Li, BM Lester, A flexible analysis tool for the quantitative acoustic assessment of infant cry. J. Speech Lang. Hearing Res.56(5), 1416–1428 (2013)
[24] G Várallyay, A Illényi, Benyó, inProc. Intl. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications. Automatic infant cry detection, (2009), pp. 11–14
[25] S Orlandi, L Bocchi, G Donzelli, C Manfredi, Central blood oxygen saturation vs crying in preterm newborns. Biomed. Signal Process. Control.7(1), 88–92 (2012)
[26] S Orlandi, PH Dejonckere, J Schoentgen, J Lebacq, N Rruqja, C Manfredi, Effective pre-processing of long term noisy audio recordings: an aid to clinical monitoring. Biomed. Signal Process. Control.8(6), 799–810 (2013)
[27] R Cohen, Y Lavner, inProc. IEEE 27th Convention of Electrical and Electronics Eng. in Israel. Infant cry analysis and detection (IEEE, 2012), pp. 1–5
[28] SM Grau, MP Robb, AT Cacace, Acoustic correlates of inspiratory phonation during infant cry. J. Speech Lang. Hear. Res.38(2), 373–381 (1995)
[29] J-J Aucouturier, Y Nonaka, K Katahira, K Okanoya, Segmentation of expiratory and inspiratory sounds in baby cry audio recordings using hidden Markov models. J. Acoust. Soc. Am.130(5), 2969–2977 (2011)
[30] L Abou-Abbas, HF Alaie, C Tadj, Automatic detection of the expiratory and inspiratory phases in newborn cry signals. Biomed. Signal Process. Control.19, 35–43 (2015)
[31] L Abou-Abbas, C Tadj, C Gargour, L Montazeri, Expiratory and inspiratory cries detection using different signals' decomposition techniques. J. Voice.31(2), 259–13 (2017).
[32] D Lederman, E Zmora, S Hauschildt, A Stellzig-Eisenhauer, K Wermke, Classification of cries of infants with cleft-palate using parallel hidden Markov models. Med. Biol. Eng. Comput.46(10), 965–975 (2008)
[33] D Lederman, A Cohen, E Zmora, K Wermke, S Hauschildt, A Stellzig-Eisenhauer, inProc. IEEE 22nd Convention of Electrical and Electronics Eng. in Israel. On the use of hidden markov models in infants' cry (IEEE, 2002), pp. 350–352

[34] D Reynolds, inEncyclopedia of biometrics. Gaussian mixture models (Springer, USA, 2009), pp. 659–663
[35] L Rabiner, B-H Juang, An introduction to hidden Markov models. ASSP Mag. IEEE.3(1), 4–16 (1986)
[36] M Gales, S Young, The application of hidden Markov models in speech recognition. Foundations Trends Signal Process.1(3), 195–304 (2008)
[37] S Davis, P Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. Acoust. Speech Signal Process. IEEE Trans.28(4), 357–366 (1980)
[38] JT Foote, inVoice, Video, and Data Communications. Content-based retrieval of music and audio (International Society for Optics and Photonics, 1997), pp. 138–147
[39] O-W Kwon, K Chan, J Hao, T-W Lee, inProc. European Conf. on Speech Communication and Technology (Eurospeech). Emotion recognition by speech signals (Citeseer, 2003), pp. 125–128
[40] De Cheveigné, H Kawahara, YIN, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Am.111(4), 1917–1930 (2002)
[41] De Cheveigné, H Kawahara, inProc. European Conf. on Speech Communication and Technology (Eurospeech). Comparative evaluation of F0 estimation algorithms, (2001), pp. 2451–2454.
[42] R.Durga, G.Yamuna and P.Abirami (2018), Multilevel Thresholding for Video Segmentation Using Cuckoo Search Algorithm, *International Journal of Scientific Research and Reviews*, Volume. 7, Issue. 1, pp. 118-135,Jan. – March, 2018.