International Journal of Engineering Innovation & Research
Volume 5, Issue 3, ISSN: 2277 – 5668

Applying Hidden Markov Model Baby Cry Signal Recognition Based on Cybernetic Theory

Dr. Saeed Ghaffari
Faculty of Library and Information Science, Payam Noor University, Qom, Iran

Maryam Ashkaboosi
Department of Visual Communication Design in Art and Architecture, Islamic Azad University - Tehran Central Branch, Tehran, Iran

Abstract **– Baby cry signal recognition is branch of application of speech recognition. It studied translation of cry signal into text. Using tool to pick up a short time period from a signal, the tools can still assume the signal under processing to be stationary. It may constitute time invariant system and time invariant excitation when it is viewed in blocks of 10-30 msec. A processing is termed as Short Term Processing (STP). Then, using mel-frequency cepstral coefficients (MFCC) and bark frequency cepstral coefficients (BFCC) extract the feature of different known baby cry signal to make the codebook. Set up hidden markov model (HMM) to train estimate baby cry signal. Classify estimate baby cry achieve baby cry signal recognition.**

Keywords **– Cybernetic Theory, Speech Signal, Signal Processing, STP, HMM.**

I. INTRODUCTION

Based on many researches, the speech signal is non-stationary in nature. Most of the signal processing assumes time invariant system and time invariant excitation, stationary signal. If we use tool to pick up a short time period from a signal, the tools can still assume the signal under processing to be stationary. A processing is termed as Short Term Processing (STP). Short term processing includes short term energy (STE), short term magnitude (STM) and short term zero-across rate (ZAR) to separate voiced sound and unvoiced sounds, as following equation shows. The acoustic and spectral characteristics of baby cry sound through speech analysis have been evaluated. Through using mel-frequency cepstral coefficients (MFCC) extract the features of three types of cries. Time-frequency representations of features are proposed. Hidden Markov Mode is used in baby cry signal recognition.

In this project, we also tried the KDTreeSearcher object to recognize the cry signal. The KDTreeSearcher base on the k-nearest neighbor algorithm which is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method for classification and regression, that predicts objects' "values" or class memberships based on the k closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification.

II. DETAILED STUDY

Short term energy:

$$E_n = \sum_{m=-inf}^{m=inf} [x(m)w(n-m)]^2$$

Short term magnitude:

$$M_n = \sum_{m=-inf}^{m=inf} |x(m)|w(n-m)$$

Short term zero crossing Rate:

$$Z_n = \sum_{m=-inf}^{m=inf} |sgn[x(m)] - sgn[w(m-1)]|w(n-m)$$

Mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. MFCC are obtained from a band-based frequency representation (using the Mel scale by default), and then a discrete cosine transform (DCT), as Fig 1 shows. The DCT is an efficient approximation for principal components analysis, so that it allows a compression, or reduction of dimensionality. A small

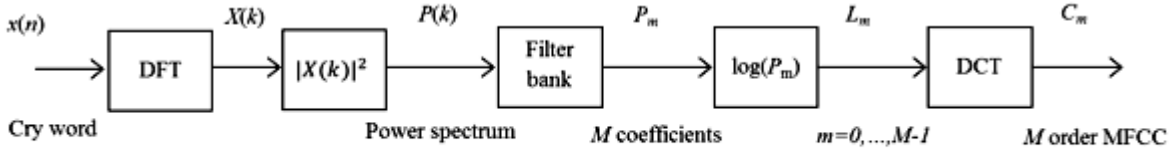
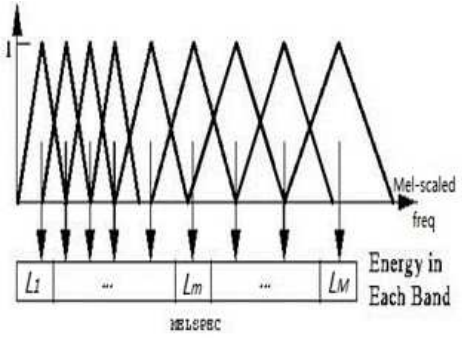


Fig.1 Mel-frequency Cepstrum (MFCC) flow chart

International Journal of Engineering Innovation & Research
Volume 5, Issue 3, ISSN: 2277 – 5668



$$L_m = log\left(\sum_{k=0}^{N-1}|X(k)|^2 H_m(k)\right),\ 0 \le m < M$$

$$C_m = \sum_{n=0}^{M-1} L_m cos\left(\frac{\pi m(n+0.5)}{M}\right), 0 \le mM$$

$$\text{Mel Frequency: } Mel(f) = 2595\, log_{10}\left(1+\frac{f}{700}\right)$$

Fig.2 Mel-frequency Cepstrum (MFCC) principle

number of features (the coefficients) end up describing the spectrum. The MFCCs are commonly used as timbre descriptors[2]. Hidden Markov Models (HMMs) provide a simple and effective frame-work for modeling time varying spectral vector sequences. As a consequence, almost all present day large vocabulary continuous speech recognition (LVCSR) system are based on HMMs. [1] The mainly principal of HMMs. It is very good for the application to classify the different feature through build Markov Models. First, using a known data initialize and train the HMM. Secondly, the trained HMM will be used to predict future value over a selected period. Finally, compute the accuracy and value of model. A random walk model will be compare to the generated HMM and a trading strategy will be implemented over a period [4].

Applying Hidden Markov Models have five following steps:

1) The model should be built the number of N hidden states. Each state corresponds to a unique state proved by model.
2) The amount of M unique observations each state as$Z = [Z_1, Z_2, \dots, Z_M]$.
3) State transition probability distributions $A = \{a_{ij}\}$ where $a_{ij} = P(q_{t+1} = S_j | q_t = S_i), 1 \le i, j \le N$
4) The emission probability distribution in state j, B== $\{b_j(k)\}$ where $b_j = P(v_k att | q_t = S_j), 1 \le j \le N, 1 \le k \le M$
5) The prior probability $\pi_i = \{\pi_i\}$ of being in state i at the beginning of the observations where $\pi_i = P(q_1 = S_i), 1 \le i \le N$.

The value of N, M, A, B and π can be used to generate the observation sequence $O = O_1 O_2 O_3 \dots O_T$ where O_T is an observation from V, and T is the number of observations in the sequence (Md, stock). To initiate a HMM, an initial state will be chosen based on the prior distribution π and t is set at 1. The model moves to state $q_{t+1} = S_j$ based on the transition probability distribution of S_i. This process will continue as t increments or until termination. More formally, this process is denoted by $\lambda = (A, B, \pi)$ where: $\sum_j a_{ij} = 1$, $\sum_t b_i(O_t) = 1$, $\sum_i \pi_i = 1$, $a_{ij}, b_i(O_t), \pi_i \ge 0$ for all i, j, t[4].

For the KDT researcher object, the training examples are vectors in a multidimensional feature space, each with a class label [5-7]. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point[8], [9]. A commonly used distance metric for continuous variables is Euclidean distance. For discrete variables, such as for text classification, another metric can be used, such as the overlap metric (or Hamming distance). Often, the classification accuracy of k-NN can be improved significantly if the distance metric is learned with specialized algorithms such as Large Margin Nearest Neighbor or Neighborhood components analysis.

A drawback of the basic "majority voting" classification occurs when the class distribution is skewed [12-17]. That is, examples of a more frequent class tend to dominate the prediction of the new example, because they tend to be common among the k nearest neighbors due to their large number. One way to overcome this problem is

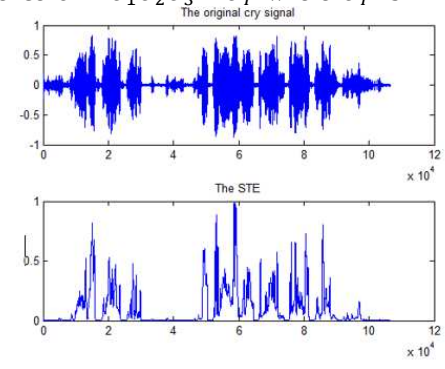


Fig. 3 Time-domain and short-term energy of attention cry signal

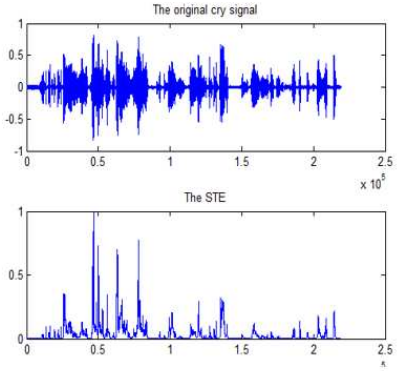


Fig. 4 Time-domain and short-term energy of diaper cry signal

to weight the classification, taking into account the distance from the test point to each of its k nearest neighbors [7-15].

III. Simulation Result

Baby cry signal detection uses short term energy to calculate word boundaries returns and the number of word utterances based off of energy threshold requirements. Speech is a random signal and is therefore extremely difficult to model. However, over short time intervals 10-30ms, speech can be considered a stationary signal (easier for analysis) because of the physiological limits of human speech production. Analyzing four different baby cry sounds, as attention, diaper requirement and hungry, detect the unvoiced or voiced for each signal, at first. Fig 3, Fig 4 and Fig 5 shows the time-domain and short-term energy for each voices.

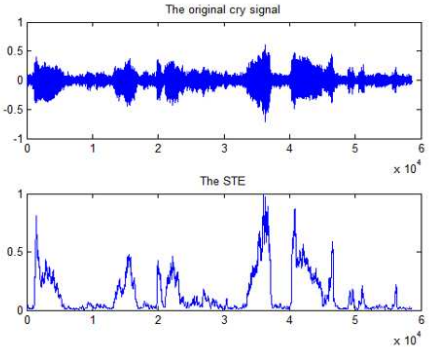


Fig. 5 Time-domain and short-term energy of hungry cry sign

Fig 6,7 and 8 shows the feature of these cry signal. We extract these features from the first peak of voiced. According to the result of detection, the attention file has 11 times cries, the diaper file has 8 times cries and the hungry file has 6 times cries. The first voiced fragment of attention cry is chosen to calculate the LPC, LPCC, MPCC and BFCC coefficients. We generate 10th order LPC,

LPCC, MPCC and BFCC.

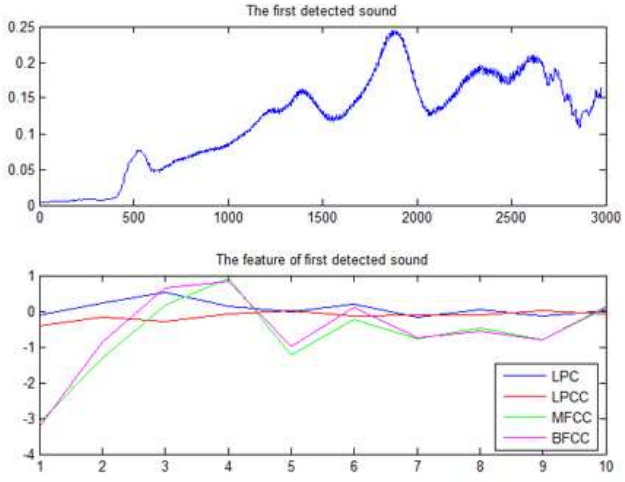


Fig. 6 the short time energy of the fragment and the feature (attention)

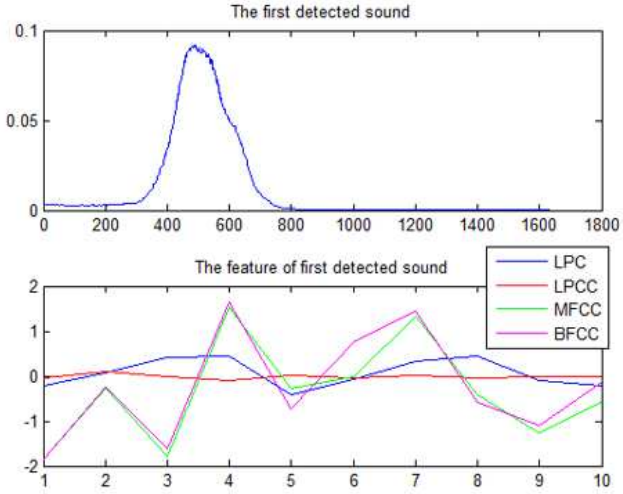


Fig. 7 The short time energy of the fragment and the feature (Diaper)

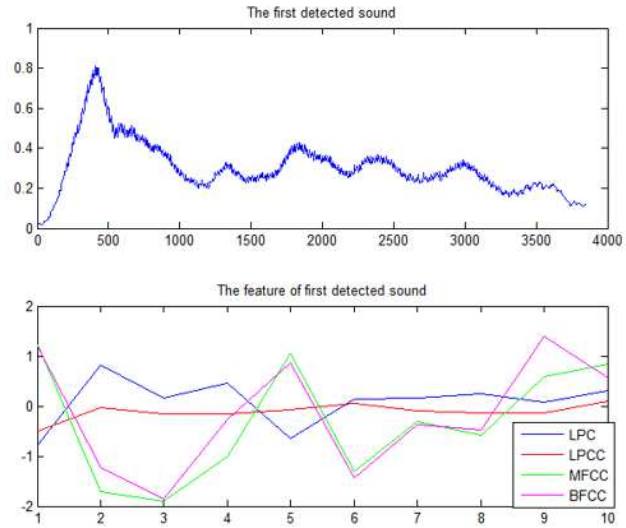


Fig. 8 the short time energy of the fragment and the feature (Diaper)

We made use of the K-NN search algorithm to classify the baby cry. The MFCC of all the known signal are calculated. Then the test signal is analyzed and the MFCC of it are extracted. These are then sent into the K-NN algorithm which calculates the k-nearest neighbor distances (which in general is Euclidean distance between similar values). Then using the training set a threshold is calculated which is used to classify the test signal. We plots some examples as following:

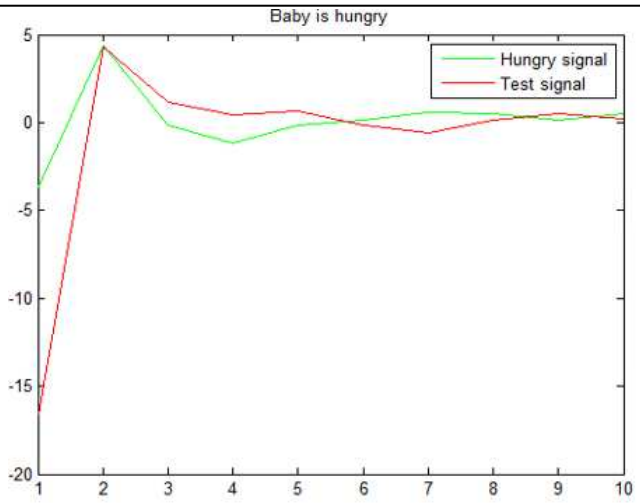


Fig. 8 Test signal T116(3) is found out to be Hungry cry

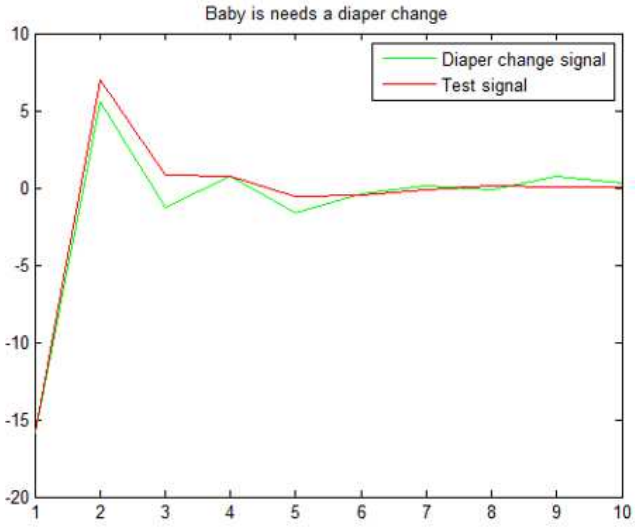


Fig. 8 Test signal T117(1) is found out to be Diaper cry

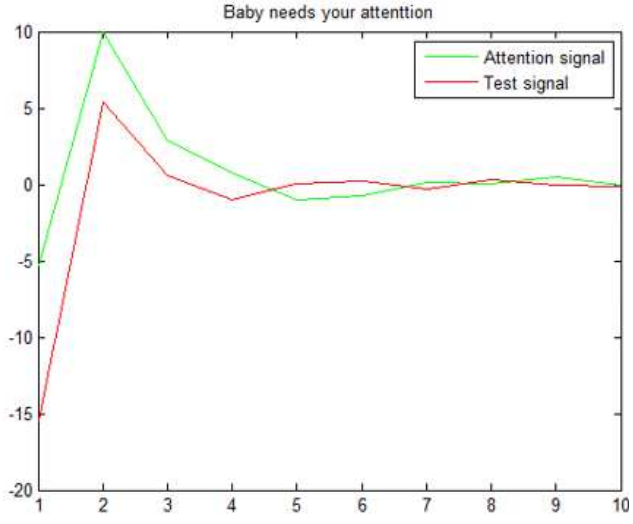


Fig. 8 Test signal T120 file is found out to be attention cry

The statistics of unknown signals

Cry signal	Attention	Diaper	Hungry	Other
File name: <name>.wav	T33(1)	T35(2)	T113	T36(1)
	T33(2)	T112(1)	T114	T36(2)
	T34	T112(3)	T116(3)	T104
	T35(1)	T117(1)	T117(2)	T105(1)
	T37	T112(1)	T121	T105(2)
	T107(1)	T123(2)	T127(1)	T122(1)
	T107(2)	T127(2)		
	T112(2)			
	T116(1)			
	T116(2)			
	T120			
	T112(2)			
	T123(1)			
	T124(2)			
Total:34	15	7	6	6

IV. CONCLUSION

Identification of the reason for the baby to cry is very much helpful in understanding the needs of the baby automatically and attending to them. Many helpful techniques can be designed which can be useful in both medical as well as household purpose. To do this we extracted the cry of an infant in different situations and then analyzed them. We then make use of the short-term signal processing to make the non-stationary signal to stationary. Later we use the MFCC to do the feature extraction of the cry part of the signal. We then made use ok K-NN algorithm for pattern matching. The results we achieved from our project is about 80% accurate. But if we had access to more number of signals we can make you of some of the most modern pattern recognition techniques like HMM, neural networks to improve the accuracy to around 95%.

REFERENCES

[1] Octavian Cheng, Waleed Abdulla, “Performance Evaluation of Front-end Processing for Speech Recognition Systems”, School of Engineering report. web

[2] Li Tan and Montri Karnjanadecha “Modified mel-frequency cepstrum coefficient”, web.

[3] Yang Li,“Multi-function enhanced active noise control system for infant incubator”, ppt

[4] Taylor Sauder, “The implementation of a Hidden Markov Model in MATLAB for the Prediction of Commodity Prices”, Bradley University, Dec 7, 2011

[5] M. Rakhshan, N. Vafamand, M. Shasadeghi, M. Dabbaghjamanesh, A. Moeini, “Design of networked polynomial control systems with random delays: sum of squares approach”. International Journal of Automation and Control, Vol. 10 (1), pp. 73-86, 2016.

[6] M. Dabbaghjamanesh, A. Moeini, M. Ashkaboosi, P. Khazaei, K. Mirzapalangi, “High performance control of grid connected cascaded H-Bridge active rectifier based on type II-fuzzy logic controller with low frequency modulation technique”, International Journal of Electrical and Computer Engineering (IJECE), Vol 6(2), 2015.

[7] P. Khazaei, S.M. Modares, M. Dabbaghjamanesh, M. Almousa, A. Moeini, “A high efficiency DC/DC boost converter for photovoltaic applications”, International Journal of Soft Computing and Engineering (IJSCE), Vol. 6 (2), 2016

[8] A. Sahba, R. Sahba, and W.-M. Lin, "Improving IPC in Simultaneous Multi-Threading (SMT) Processors by Capping IQ Utilization According to Dispatched Memory Instructions," presented at the 2014 World Automation Congress, Waikoloa Village, HI, 2014.

[9] A. Sahba, Y. Zhang, M. Hays and W.-M. Lin, "A Real-Time Per-Thread IQ-Capping Technique for Simultaneous MultiThreading (SMT) Processors", In the Proceedings of the 11th International Conference on Information Technology New Generation (lTNG 2014), April 2014.

[10] M. Bagheri, M. Madani, R. Sahba, and A. Sahba, "Real time object detection using a novel adaptive color thresholding method", International ACM workshop on Ubiquitous meta user interfaces (Ubi-MUI'11), Scottsdale, AZ, November 2011.

[11] Hajinoroozi, Mehdi, et al. "Prediction of driver's drowsy and alert states from EEG signals with deep learning" Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop.

[12] Hajinoroozi, Mehdi, et al. "Feature extraction with deep belief networks for driver's cognitive states prediction from EEG data." Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on. IEEE, 2015.

[13] Grigoryan, Artyom M., and Mehdi Hajinoroozi. "Image and audio signal filtration with discrete Heap transforms." Applied Mathematics and Sciences: An International Journal (MathSJ) 1.1 (2014): 1-18.

[14] Grigoryan, Artyom M., and Mehdi Hajinoroozi. "A novel method of filtration by the discrete heap transforms." IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2014.

[15] Jenkinson, J., Grigoryan, A., Hajinoroozi, M., Diaz Hernandez, R., Peregrina Barreto, H., Ortiz Esquivel, A., ... & Chavushyan, V. (2014, October). Machine learning and image processing in astronomy with sparse data sets. In Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on (pp. 200-203). IEEE.

[16] Najjar, Mohammad, Amirhossein Moeini, Mohammad Kazem Bakhshizadeh, Frede Blaabjerg, and Shahrokh Farhangi. "Optimal Selective Harmonic Mitigation Technique on Variable DC Link Cascaded H-Bridge Converter to Meet Power Quality Standards."

[17] Moeini, Amirhossein, Zhao Hui, and Shuo Wang. "High efficiency, hybrid Selective Harmonic Elimination phase-shift PWM technique for Cascaded H-Bridge inverters to improve dynamic response and operate in complete normal modulation indices." In 2016 IEEE Applied Power Electronics Conference and Exposition (APEC), pp. 2019-2026. IEEE, 2016.