Voiced/Unvoiced Classification by Hybrid Method Based On Cepstrum and EMD

Bhumika Nirmalkar[1], Dr. Sandeep Kumar[2]

Department of Electronics and Telecommunication
Rungta College of Engineering and Technology Bhilai, India
bhuminirmalkar@gmail.com

Department of Electronics and Telecommunication
Rungta College of Engineering and Technology Bhilai, India
skumardr20@gmail.com

----------***----------

Abstract: ***This paper introduces an enhanced cepstrum and EMD based hybrid method for the voiced/un-voiced (V/UV) classification by simulation using MATLAB. Voicing choices are made utilizing a feature voiced/unvoiced characterization calculation taking into account factual investigation of cepstral peak, energy and zero crossing rates. Execution investigation on an extensive database shows in terms of classification accuracy and error. This algorithm is likewise appeared to be robust to large verity of voiced signals as compared to traditional cepstrum and EMD based classifiers.***

Keywords **—** Cepstrum, EMD, Hybrid method voiced/unvoiced classification, cepstral peak, Energy, zero crossing rate, MIR database.

1. Introduction

Speech being a natural mode of communication for humans can provide a convenient interface to control devices. Some of the speech recognition applications require speaker-dependent isolated word recognition. Current implementations of speech recognizers have been done for personal computers and digital signal processors. However, some applications which require a low-cost portable speech interface cannot use a personal computer or digital signal processor based implementation on account of cost, portability and scalability. Human speech is the foundation of self-expression and communication with others. In the past, ranges of speech based communication technologies have been developed. Speech is an acoustic signal produced from a speech production system. Producing speech sounds, the air flow from lungs first passes the glottis and then throat and mouth. Depending on speech sound the speech production can be broadly categorized into three activities:-Voiced speech, unvoiced speech, Silence region.

If the input excitation is nearly periodic impulse sequence, then the corresponding speech looks visually nearly periodic and is termed as voiced speech. During the production of voiced speech; the air exhaling out of lungs through the trachea is interrupted periodically by the vibrating vocal folds. Due to this, the glottal wave is generated that excites the speech production system resulting in the voiced speech. If the excitation is random noise-like, then the resulting speech will also be random noise-like without any periodic nature and is termed as Unvoiced Speech. During the production of unvoiced speech, the air exhaling out of lungs through the trachea is not interrupted by the vibrating vocal folds. However, starting from glottis, somewhere along the length of vocal tract, total or partial closure occurs which results in obstructing air flow completely or narrowly. This modification of airflow results in stop or frication excitation and excites the vocal tract system to produce unvoiced speech. Silence region, there is no excitation supplied to the vocal tract and hence no speech output.

While speech processing applications like multi rate speech coder[1,2],language identification[3],speech signal modeling[4]require classification of the speech signal in to voiced, unvoiced and silences regions, there are some prominent application like pitch frequency estimation[5,6], identification of the glottal closure instants(GCIs)[7],which require knowledge of only the voiced regions of the speech signals.

In this paper hybrid method for voiced/unvoiced classification has been simulated using MATLAB. Performance of this scheme has been compared with existing algorithms in terms of voiced/unvoiced classification accuracy and error rate.

The rest of the paper is organized as follows. In Section II, a review like comparative study on different voiced/unvoiced classification algorithm is described. In section III Close description of the implementation of enhanced hybrid voiced/unvoiced speech classifier is given. In Section V, the results of the performance analysis are presented. Concluding remarks are given in Section VI.

2. Literature review

In [8] a cepstrum-based pitch detection using a new statistical voiced/unvoiced classification. In this method voicing decision are made using multi feature voiced unvoiced classification based on statistical analysis of cepstral peak, zero crossing rate and energy of short time segment of speech signal. The performance was improved under noisy conditions. In [9] robust voiced/unvoiced speech classification using empirical mode decomposition and periodic correlation model. This method analyzes the signal by nonlinear and non stationary signal which is used as a filter for additive noise in speech signal. The use of EMD improves the classification performance and efficiency is noticeable. In[10] pitch detection algorithms and voiced/unvoiced classification for noisy speech. This algorithm is based on cepstral analysis; time auto co-relation, spectral temporal auto co-relation (STA) and average magnitude difference function. All of the algorithms of voiced/unvoiced classification give good performance for clean speech. In[11] a new method for voiced/unvoiced classification based on epoch extraction. This method uses zero frequency filtered speech signal is used to extract the epochs. Features of this method are depends on excitation source information. This method was better than the normalized cross correlation based voiced/unvoiced classification. In[12] a new method of voiced/unvoiced classification based on clustering. Their algorithm is based on analysis of cepstral peak, zero crossing rate, and autocorrelation function (ACF) peak of short-time segments of the speech signal by using some clustering methods The advantage of this clustering based method is getting rid of determining a threshold. So it is highly speaker independent. better satisfactory performance for identification of voiced and unvoiced segments of speech. In[13] Adaptive thresholding approach for robust voiced/unvoiced classification. They introduced a robust voiced/unvoiced classification method by using linear model of empirical mode decomposition (EMD) controlled by Hurst exponent. This algorithm improves the classification performance. In[14] a pseudo Wigner-Ville distribution (PWED based method) for the voiced/unvoiced detection in noisy speech signals. The marginal energy density with respect to time (MEDT) which is used as a feature to provide voiced/unvoiced classification and allowed instantaneous detection of voiced regions. Also this method does not require knowledge of pitch frequency. The performance of algorithm was improved for clean and noisy signals. In [15] identifying the voice/unvoiced and silence chunks in speech. This algorithm is based on Zero crossing rate, Short time energy, and fundamental frequency for identifying the speech signals. They found better accuracy and data collected four different speakers. In [16] Voiced-unvoiced classification of speech using autocorrelation matrix. This method is based on, signal energy, the peak-to-peak difference of the autocorrelation function, number of zero crossings of the autocorrelation function and the unit delay autocorrelation coefficient all together. The accuracy of the proposed method found 100% for women and 98% for men. In[17] a comparative performance study of several time domain features for voiced/unvoiced classification of speech. They have considered five classification schemes based on Energy (E), Zero crossing rate (ZCR), Autocorrelation Function (ACF), Average Magnitude Difference Function (AMDF), Weighted ACF (WACF) and Discrete Wavelet Function (DWT) for their study. They evaluated the performance of five voiced/unvoiced classification scheme one or two features without any pre or post processing approaches. In[18] voiced/unvoiced speech classification using adaptive thresholding with bivariate EMD. They proposed an effective method of voiced/unvoiced classification without any use of training data and prior knowledge, to achieve robust and data adaptive voiced/unvoiced classification technique which is suitable for real world speech processing application. The classification efficiency was better than that of the recently reported algorithms. In [19] instantaneous detection of voiced/non-voiced detection based on the method variation mode decomposition

(VMD). This method does not require prior information of the pitch. It's provided better accuracy and performance of voiced and unvoiced classification. In[20] voiced/unvoiced classification in compressively sensed speech signals. This method is based on compressive sensing (CS)/Sparse coding for detection of voiced/unvoiced classification. the sparse vector contains the source characteristics of the speech signals, if a suitable dictionary is chosen. Using an information theoretic based criterion the behavior of sparse vector is quantified. An adaptive threshold selection scheme used for final voiced/unvoiced classification. This method selected (voiced) region, which can be used for application of speaker verification. In[21] voice and unvoiced classification using fuzzy logic. This algorithm is based on features Zero crossing rate, Short time energy for classification of voice, unvoiced and silence. This method successfully classified the speech signals.

3. Problem identification

Cepstrum is the conventional and convenient method for voiced/unvoiced classification although it is best suitable for classification of voiced signal only. There are some limitations of cepstrum of Voiced classification as it does not give sharp spectrum of signal and it is best suitable for static signal only. EMD is used for voiced/unvoiced classification which is an advanced method but it is most suitable for unvoiced signal classification. Also EMD is highly nonlinear method i.e. it is very time consuming. Thus we can conclude that none of the method is more reliable when it comes to voiced/unvoiced classification for both voiced and unvoiced signals.

The solution of the problem is that cepstrum is best suitable for the voiced classification and EMD is best suitable for the unvoiced classification because of its decomposing nature. This leads to the realization of a hybrid method which carries property of both methods of voiced/unvoiced classification i.e. cepstrum and EMD to overcome the problem of voiced/unvoiced classification. In this hybrid method we can take advantage of advantages of both methods as cepstrum has some advantages that it is non-decomposing in nature and comparatively simpler then the available method. EMD gives sharp spectrum and it is more suitable for dynamic signals also.

4. Methodology

A hybrid method for voiced/unvoiced classification based on cepstrum and EMD will be used. For voiced/unvoiced Classification signals will be taken from suitable well defined source.

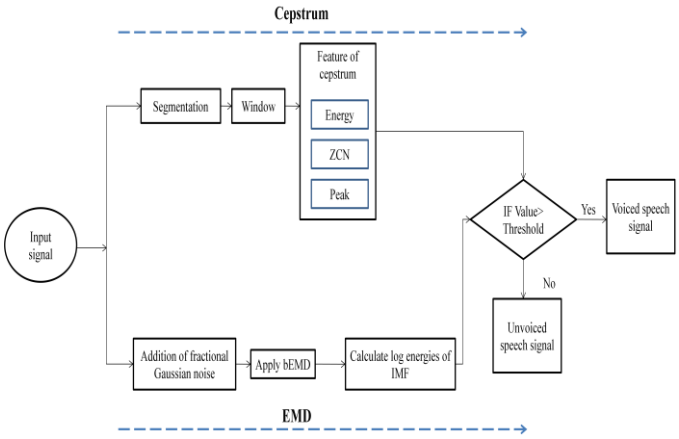

Fig 1 methodology flow chart

The entire work is mix of cepstrum and EMD method thus generating a hybrid method. In cepstrum module signals will be segmented by segmentation method. In segmentation speech signal must be segmented into several frames for the ease of analysis. Then it will be passed through hamming window. If w = hamming (L) returns an L-point symmetric Hamming window in the column vector w. L should be a positive integer. The coefficients of a Hamming window are computed from the following equation.

$$w(n) = 0.54 - 0.46 \cos(2\pi(n/N)), 0 \le n \le N \qquad (1)$$

The window length is L=N+1.Each frame has to be multiplied with a hamming window in order to keep the continuity of the first and the last points in the frames. After that feature of cepstrum will be calculated. Since A zero-crossing is a point where the sign of a mathematical function changes in feature of cepstrum energy, zero crossing number and peak will calculate. After that threshold extraction will be performed. As per our algorithm if cepstral peak value exceeds the threshold level then section is marked as voiced otherwise unvoiced.

In EMD method, combine speech signal and fractional Gaussian noise (fGN) to the complex signal after that apply bivariate EMD. Fractional Gaussian noise (fGn) is a generalization of ordinary discrete white Gaussian noise, and it is a versatile model for broad- band noise .In bivariate EMD, which is a decomposing method; it decomposes the non linear and non stationary signals into set of band limited components known as intrinsic mode function(IMF).After that calculate the log energies of IMF.

If sub band log energies IMF will exceeds the threshold level then section is marked as voiced otherwise unvoiced.

5. Result and discussion

A hybrid method has been generated with is evaluated with compare to the conventional methods i.e. cepstrum and EMD. The speech signals are taken from the MIR database. Out of 1000 speech signals selected 10 speech signals has been analyzed. In this experiment the speech material is sampled at 16 kHz and segmented into frame of length 256 with -3ms sifting. The silence part is removed. Approximately 187 each frames are used to evaluate the accuracy and bit rate each frame is accurately labeled for voiced/unvoiced. The performance of the proposed method is measured in terms of the error rate and accuracy. The performance of the proposed algorithm is compared with the EMD and cepstrum based method. For evaluation of error rate and accuracy parameter were calculated for signals from the data base. The error rate is obtained by, performing the XOR operation between the actual classification and classification obtained from cepstrum. The sum of the result obtained from the XOR operation is divided by the total length of the signal similar operation has been performed with EMD and derived Hybrid algorithm. The result obtained has been sown in table I.

TABLE 1 : The error rate of the hybrid algorithm at different speech signals by comparison with cepstrum and EMD.

S. N	Cepstrum	EMD	Hybrid	deviation w.r.t. cepstrum	deviation w.r.t. EMD
1	0.32796	0.38172	0.30108	-0.089278	-0.26783
2	0.25134	0.38172	0.3172	0.207629	-0.20340
3	0.59893	0.61497	0.57219	-0.046732	-0.07476
4	0.57219	0.59893	0.54011	-0.059395	-0.10890
5	0.46277	0.61702	0.41489	-0.115404	-0.48718
6	0.38172	0.47312	0.34946	-0.092313	-0.35386
7	0.60638	0.67021	0.56383	-0.075466	-0.18867
8	0.43316	0.60963	0.4492	0.035707	-0.35714
9	0.47059	0.62032	0.48128	0.022211	-0.28889
10	0.34177	0.58742	0.32691	-0.045455	-0.79688

Table I Demonstrate the error rate of the hybrid algorithm at different speech signals by comparison with cepstrum and EMD. The overall voiced/unvoiced classification accuracy with hybrid method is always better than the existing algorithm .Since in our whole operation the 100% in constitute of error and accuracy i.e

Accuracy of the result + Error of the result = 1 (2)

The accuracy is obtained by subtracting the values of the error from table I by factor 1 and corresponding values has been show in table II.

TABLE 2 : The accuracy of the hybrid algorithm at different speech signals by comparison with cepstrum and EMD.

S. N	Cepstrum	EMD	Hybrid	deviation w.r.t. cepstrum	deviation w.r.t. EMD
1	0.67204	0.61828	0.69892	0.0384593	0.115378
2	0.74866	0.61828	0.6828	-0.096455	0.094493
3	0.40107	0.38503	0.42781	0.0625043	0.099997
4	0.42781	0.40107	0.45989	0.0697558	0.127900
5	0.53723	0.38298	0.58511	0.0818307	0.345456
6	0.61828	0.52688	0.65054	0.0495895	0.190088
7	0.39362	0.32979	0.43617	0.0975537	0.243895
8	0.56684	0.39037	0.5508	-0.029121	0.291267
9	0.52941	0.37968	0.51872	-0.020608	0.268044
10	0.65823	0.41258	0.67309	0.0220772	0.387035

Table II Demonstrate the accuracy of the hybrid algorithm at different speech signals by comparison with cepstrum and EMD. MATLAB 15.0 is used for our calculation choose MATLAB as our programming environment as it offers many advantages. It contains a variety of signal processing and statistical tools, which help users in generating a variety of signals and plotting them. MATLAB excels at numerical computation, especially when dealing with matrices of data.

The performance of the proposed algorithm is compared to the cepstrum based method and EMD based method. The same reference data base is used to evaluate the performance of the both algorithm. Finally, observe that based on the mentioned performance of the existing algorithms. The proposed method proves its superiority in

voiced/unvoiced classification. The principle advantages of the proposed algorithm are that is it works with less error rate and with more accuracy.

6. Conclusion

In this work an improved voiced/unvoiced classification algorithm is has been obtained. This paper present voiced/unvoiced classification based on combining features of two most popular conventional methods of voiced/unvoiced classification i.e. EMD and Cepstrum. Experimental result gives on speech database shows the superiority of the obtained Hybrid method over existing methods. The use of the proposed method for real world speech processing applications will be more reliable as both of these methods has their shortcomings. On carefully observing the error table it is evidently clear that the generated Hybrid method is giving much better result.

References

[1] B. Yin, E. Ambikairajah, F. Chen, Voiced/unvoiced pattern-based duration modeling for language identification, in: IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 2009, pp. 4341–4344.

[2] E. Paksoy, J. Carlos de Martin, A. McCree, C.G. Gerlach, A. Anandakumar, W.M. Lai, V. Viswanathan, An adaptive multi-rate speech coder for digital cellular telephony, in: IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Phoenix, USA, vol. 1, 1999, pp. 193–196.

[3] A.M. Kondoz, Digital Speech: Coding for Low Bit Rate Communication Systems, Wiley, England, 2004.

[4] P. Sircar, R.K. Saini, Parametric modeling of speech by complex AM and FM signals, Digital Signal Processing 17 (6) (2007) 1055–1064.

[5] B. Resch, M. Nilsson, A. Ekman, W.B. Kleijn, Estimation of the instantaneous pitch of speech, IEEE Transactions on Audio, Speech and Language Processing 15 (3) (2007) 813–822.

[6] D. Joho, M. Bennewitz, S. Behnke, Pitch estimation using models of voiced speech on three levels, in: IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Honolulu, USA, vol. 4, 2007, pp. 1077–1080.

[7] P.A. Naylor, A. Kounoudes, J. Gudnason, M. Brookes, Estimation of glottal closure instants in voiced speech using the DYPSA algorithm, IEEE Transactions on Audio, Speech and Language Processing 15 (1) (2007) 34–43.

[8] Ahmadi, S., &Spanias, A. S. (1999). Cepstrum-based pitch detection using a new statistical V/UV classification algorithm.*Speech and Audio Processing, IEEE Transactions on*, *7*(3), 333-338.

[9] Molla, M. K. I., Hirose, K., &Minematsu, N. (2009). Robust voiced/unvoiced speech classification using empirical mode decomposition and periodic correlation model. In *INTERSPEECH* (pp. 2530-2533)

[10] Verteletskaya, E., Sakhnov, K., &Šimák, B. (2009). Pitch detection algorithms and voiced/unvoiced classification for noisy speech. In *Systems, Signals and Image Processing, 2009.IWSSIP 2009. 16th International Conference on* (pp. 1-5). IEEE.

[11] Dhananjaya, N., &Yegnanarayana, B. (2010). Voiced/nonvoiced detection based on robustness of voiced epochs. *Signal Processing Letters, IEEE*, *17*(3), 273-276.

[12] Radmard, M., Hadavi, M., &Nayebi, M. M. (2011). A new method of voiced/unvoiced classification based on clustering. *Journal of Signal and Information Processing*, *2*(04), 336.

[13] Molla, M. K. I., Hirose, K., Roy, S. K., & Ahmad, S. (2011,). Adaptive thresholding approach for robust voiced/unvoiced classification. In *Circuits and Systems (ISCAS), 2011 IEEE International Symposium on* (pp. 2409-2412). IEEE.

[14] Jain, P., &Pachori, R. B. (2013). Marginal energy density over the low frequency range as a feature for voiced/non-voiced detection in noisy speech signals. *Journal of the Franklin Institute*, *350*(4), 698-716

.

[15] Sharma, P., &Rajpoot, A. K. (2013). Automatic identification of silence, unvoiced and voiced chunks in speech.*Journal of Computer Science & Information Technology (CS & IT)*, *3*(5), 87-96.

[16] Senturk, Z., Yetgin, O. E., &Salor, O. (2014). Voiced-unvoiced classification of speech using autocorrelation matrix. In *Signal Processing and Communications Applications Conference (SIU), 2014 22nd* (pp. 1802-1805). IEEE.

[17] Faycal, Y., &Bensebti, M. (2014). Comparative performance study of several features for

voiced/non-voiced classification.*Int. Arab J. Inf. Technol.*, *11*(3), 293-299.

[18] Molla, M. K. I., Hirose, K., &Hasan,(2015) M. K. Voiced/non-voiced speech classification using adaptive thresholding with bivariate EMD. *Pattern Analysis and Applications*, 1-6.

[19] Upadhyay, A., &Pachori, R. B. (2015).Instantaneous voiced/non-voiced detection in speech signals based on variational mode decomposition.*Journal of the Franklin Institute*.

[20] Abrol, V., Sharma, P., & Sao, A. K. (2015). Voiced/nonvoiced detection in compressively sensed speech signals. *Speech Communication*, *72*, 194-207.

[21] Algabri, M., Alsulaiman, M., Muhammad, G., Zakariah, M., Bencherif, M., & Ali, Z. (2015). Voice and Unvoiced Classification Using Fuzzy Logic.In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)* (p. 416). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).