A Modified Hilbert Analysis Method to Improve Voice Stress Analysis Systems

Waleed El Nahal
Electronics & Communications Department
Faculty of Engineering, MSA University 6th October, Egypt

Ashraf Mohamed Ali
Electronics & Communications Department
Faculty of Engineering, Sinai University, Egypt

Hatem M. Zakaria
Electrical Engineering Department
Benha Faculty of Engineering, Benha University, Egypt

ABSTRACT

Analyzing the cognitive load generated in the brain is the most important issue for specific applications such as voice stress analysis (VSA) systems in which the detection of stressed speech caused by an act of deception under law enforcement interview questioning or military interrogation.

The most widely used algorithm for VSA systems is the empirical mode decomposition (EMD). Currently EMD that uses the cubic spline interpolation technique to find the envelopes of the non-periodic signal takes a long processing time, and to achieve accurate results the process is very time consuming and expensive otherwise some tests tend to produce inaccurate results. On the other hand EMD that uses Hilbert analysis method to speed up the process and provide more accurate results, suffer from finding the envelopes of the non-periodic signal.

In this paper, a new algorithm is proposed for VSA, named fast Fourier transform (FFT) with a modified Hilbert analysis method (MH) for EMD algorithm, (FTT_MH_EMD), which provides a new technique that modifies the conventional Hilbert analysis method and combines it with the fast Fourier transform algorithm to overcome the previous limitations of using individually the FTT algorithm, the cubic spline interpolation technique or the conventional Hilbert analysis method and that can speed up the processing time and gives accurate results.

Simulations and results witness that the proposed algorithm provides higher accuracy than the other attempts and also the processing time has dropped by 10 times faster than those in the products currently available in the market for VSA.

General Terms

Lie Detection, Polygraph, Biometrics

Keywords

Voice Stress Analysis, Empirical Mode Decomposition, Cognitive Load.

1. INTRODUCTION

Detection of stress by voice analysis has numerous applications, most notably in military, law enforcement and emergency services. The concept of VSA comes from when a person is under stress, micro-muscle tremors which are present in every muscle in the body including the vocal chords will transmit this stress through the speech [1]. The physiologic tremors may be produced by anxiety, stress, fatigue, or metabolic derangements or by certain drugs [2].

Stress can be detected by analyzing the change in the micro-tremor frequency for each person's voice [3 and 4]. Conventional stress detection methods include Fast Fourier Transform (FFT) or McQuiston-Ford algorithm.

The advantages of using FFT for VSA are low cost when implemented in products and fast iterations method, and the disadvantage is the low resolution and therefore low accuracy $(60\% - 70\%)$ [1]. On the other hand the advantage of using Mcquiston Ford for VSA is the high accuracy results which are greater than 80%, and the disadvantage are complex algorithms and expensive implementation [5].

An important, barely audible, frequency is produced by the human vocal cords which could measure in a fairly accurate manner if the subject is being deceitful due to the effect the brain's cognitive load has on the vocal cords [1 and 5]. This frequency is called the micro-tremor frequency generated by the involuntary micro-tremor muscle fibers in the vocal cords. When the subject is being deceitful, more work is induced on the brain which as a result causes tremor muscles to tense up in a non-voluntary manner [1 and 6].

Different non voluntary muscles would tense up including the micro-tremor muscle fibers found on the vocal cords. This induces a higher micro-tremor frequency range notably showing a large quantity of induced cognitive load and therefore a deceitful subject [7].

The micro-tremor frequency of all muscles in the body, including the vocal chords, vibrates in the 8 to 12 Hz range in the normal case without any stress, and that the VSA vendors claim to be the sole source of detecting if an individual is lying or is under stress [1, 4 and 8].

The paper is organized from seven sections EMD, EMD with Hilbert Analysis, previous work and limitations, the proposed algorithm, the proposed VSA system model, results and simulations and finally the conclusion.

2. EMD

EMD algorithm extracts or decomposes the input signal into its frequency components; these components are called intrinsic mode functions (IMFs). The extraction is done from the highest frequency to the lowest frequency and the process stops when it satisfies a selected stopping criterion that could be a standard deviation between two consecutive results or else [9, 10]. So we can write the original signal as seen in (1)

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + R_n(t) \tag{1}$$

The micro tremor frequency which represents the stress induced by the subject is the frequency of the last IMF extracted and it is considered an accurate identifier to cognitive load presence which can lead to accurate analysis of the subject's case [1, 5, and 6].

The EMD process can be summarized as shown in Fig.1 and starts by extracting the upper and lower envelopes for the

input signal h(t). Then the mean of the envelopes is calculated (mn). The difference between the input "original" signal h(t) and the mean value is given by the value achieved from this calculation hn(t) is then tested to determine whether it matches the criteria of an IMF [11 and 12].
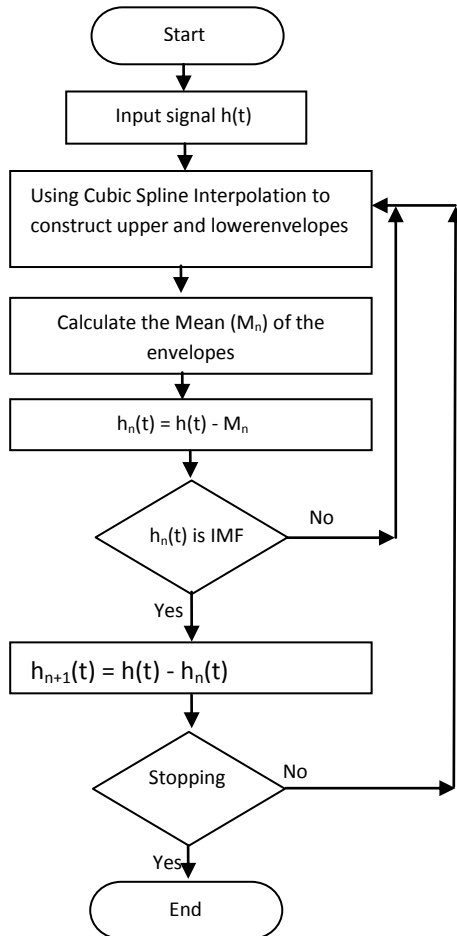

Fig.1. EMD flowchart.

The output signal is to be considered an IMF if the following are both true [1 and 2]:

1. The number of extreme points and the number of zero crossings are equal or differ by one at most in the whole residual signal.

2. The mean value of the maxima and minima envelopes becomes defined as zero.

3. EMD WITH HILBERT ANALYSIS

As seen in [2 and 13] Hilbert analysis is a process to find the amplitude envelopes of the signals. Hilbert analysis provides what is known as the analytic representation of a real-valued function or signal to produce the signal envelope. The basic idea is that the negative frequency components are to be discarded with no loss in information if the signal satisfies the Hermitian symmetry in its spectrum.

The Hermitian symmetry is a property indicating that the complex conjugate of a component is equal to the original function with the variable changed in sign.

This is represented through the following function:

$$f(-x) = \overline{f(x)} \qquad (2)$$

As a result of the analytic representation, attributes of the signal become more accessible. This is used to find the amplitude envelope of symmetrical signal as shown in Fig.2.

4. PREVIOUS WORK AND LIMITATIONS

VSA is accomplished by measuring fluctuations in the physiological micro-tremor present in speech. However, the accuracy of VSA is disputed among experts. Algorithms such as Fast Fourier Transform (FFT), Mcquiston Ford algorithm and EMD are some of the algorithms used. The latter has many advantages over the others which are the higher accuracy and faster processing time for constant phased signal.

Limitation of EMD using Cubic Spline Interpolation is that this technique when provided with a non-periodic or non-symmetrical signal suffers from long processing times. A five second signal could be processed in hours based on its complexity which mainly depends upon two main factors, the number of frequencies present in the signal and the process of producing the envelopes for the input signal and its iterations [11 and 12]. A 3 second signal will take over 30 minutes of processing time to compute using the average 2.4 GHz processors. To overcome this, a new approach is considered using Hilbert Analysis to find the envelopes of the Voice Signal [5].

FFT provides a fast and easy method to achieve the frequency spectrum of a time domain signal, due to having easily accessible built in functions in almost every coding platform. However, FFT suffers from a major disadvantage in this area of research [5]. FFT provides a low resolution and that makes it very difficult to extract the micro-tremor frequency.

Limitation of Hilbert analysis is to find envelopes of asymmetric signals, where a speech signal is the most obvious example of an asymmetric signal [14]. Even if the speech signal is inputted symmetric, it will lose this property after passing through filters. Therefore, performing Hilbert analysis on the signal will cause a loss of information as the upper and lower envelope will not match due to the fact that a speech signal does not follow the Hermitian symmetry in its spectrum [16]. As a result, in its original mathematical format, Hilbert analysis cannot be used to find both the upper and lower envelopes of the signal causing EMD iterations not possible [2 and 3].

5. THE PROPOSED ALGORITHM

The proposed algorithm FTT_MH_EMD is coming to solve the problems and the limitations of the previous methods, to speed up the processing time. At first the proposed algorithm modifies the Hilbert analysis method to detect the signal envelope rather than using the Cubic Spline Interpolation or conventional Hilbert analysis to overcome the long processing time problem for envelope detection, and then it combines FTT with the modified Hilbert analysis to speed up VSA process.

So we will discuss at first the proposed modified Hilbert analysis EMD (MH_EMD) algorithm, then we will demonstrate how FTT method will be used with MH_EMD algorithm to introduce the proposed FTT_MH_EMD algorithm to speed up the VSA process by decomposing the speech signal and finding the micro tremor frequency as fast as possible.

5.1. The Proposed MH_EMD Algorithm.

The first novel of the proposed algorithm is to modify the Hilbert analysis method to detect the envelope of the input voice signal instead of the Cubic Spline Interpolation to overcome the long processing time problem.

Although the Hilbert analysis can't detect the envelope of asymmetrical signals, so the proposed algorithm modifies it by using a unique signal processing technique to find the positive and negative envelopes of not only of the symmetrical signals but also of the asymmetrical signals.

The conventional Hilbert analysis for detecting the signal envelope shown in **Fig.2,** assumes that the signal is symmetric and its rules are only applied to the positive components.
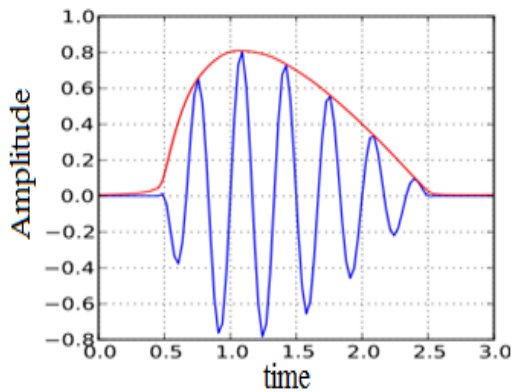

Fig.2. Symmetrical signal with amplitude envelope

The proposed FTT_MH_EMD algorithm provides a modified Hilbert analysis for the EMD algorithm (MH_EMD) that assumes any input signal is an asymmetric one, gets the positive and the negative components of the input signal then the Hilbert analysis is applied to each component individually to get the positive and the negative envelopes and finally finds the mean of the two envelopes as shown in **Fig.3.**

A novel technique is introduced where Hilbert analysis is applied to signal which do not follow the Hermitian symmetry property to find the signal envelopes.

Hilbert transform of x(t) is H[x(t)] = y(t) is defined as seen in [14 and 16] by

$$H[x(t)] = y(t) = \lim_{\varepsilon \to 0}[\int_{-\infty}^{\infty} x(t)dt < 0 \tag{3}$$

then

$$H[x(t)] = y(t) = \frac{1}{\pi} p.v. \int_{-\infty}^{\infty} \frac{x(u)}{(t-u)} du \tag{4}$$

where p.v. as the Cauchy principle value of the integral.

Moreover, z(t) is the analytical signal of x(t):

$$z(t) = x(t) + iy(t) = A(t)e^{i\theta(t)} \tag{5}$$

$$A(t) = \sqrt{x^2(t) + y^2(t)} \tag{6}$$

According to the flowchart of EMD in **Fig.1** and the above equations of the Hilbert transform, the proposed MH_EMD algorithm can be obtained as shown in **Fig.3**.

The selected stopping criterion of the proposed MH_EMD algorithm is the standard deviation (SD) between two consecutive IMFs; we choose $0.2 \le SD \le 0.3$.

5.2. The Proposed FFT_MH_EMD Algorithm

In this section, the proposed FTT_MH_EMD algorithm will be introduced to find the micro tremor frequency of the input speech signal accurately in a short time.

The second novel of the proposed algorithm is to reduce the complexity of the signal as much as possible and then speeding up the processing time by using FFT with MH_EMD algorithm as shown in **Fig.4**.
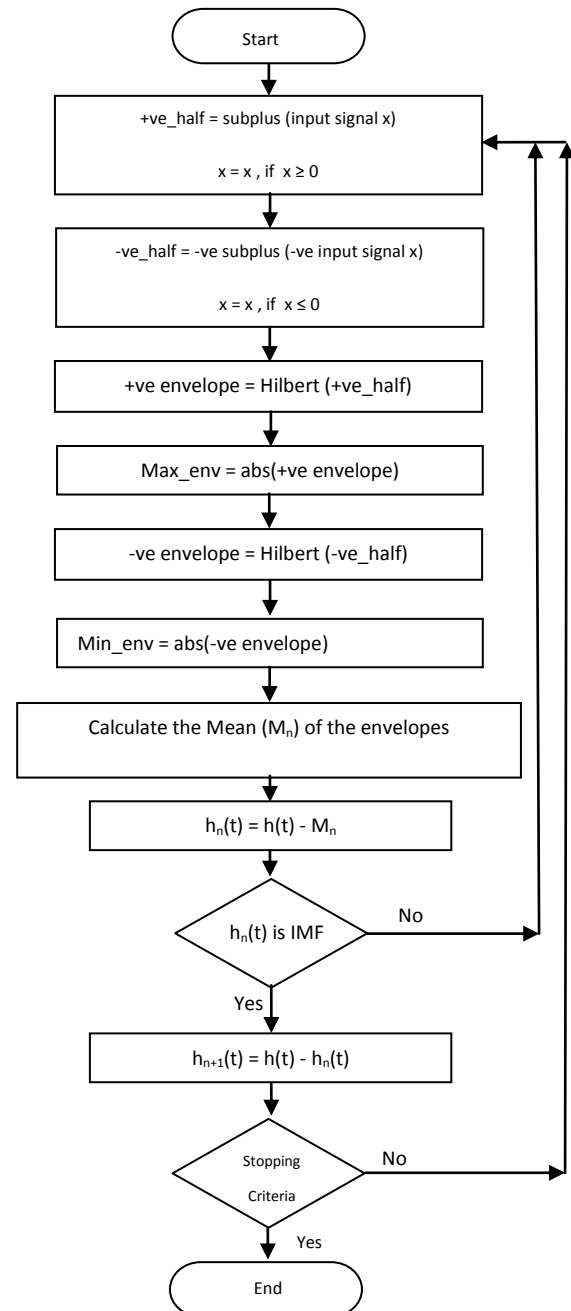

Fig.3. The proposed MH_EMD algorithm flowchart

VSA systems can be classified into two categories, energy-based systems and frequency-based systems. In this paper we concern with the frequency-based VSA systems where IMFs have 3 frequency bands: Neutral band (8-12Hz), Medium-Stress band (12-15Hz), and High-Stress band (15-17Hz) [5]. The human vocal chords vibrate in the range of 8Hz to 12Hz without exposing to any stress, these micro tremor frequencies that the VSA systems search for, to be the only source of detecting if an individual is lying [5].
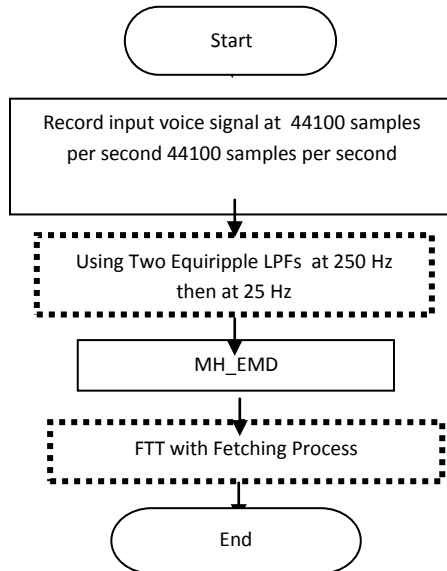
```
        ┌─────────────┐
        │    Start     │
        └─────────────┘
               │
  ┌──────────────────────────────┐
  │ Record input voice signal at 44100 │
  │ samples per second 44100 samples   │
  │ per second                          │
  └──────────────────────────────┘
               │
  ┌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌┐
  ┊ Using Two Equiripple LPFs at 250 Hz ┊
  ┊ then at 25 Hz                         ┊
  └╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌┘
               │
  ┌──────────────────────────────┐
  │           MH_EMD               │
  └──────────────────────────────┘
               │
  ┌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌┐
  ┊      FTT with Fetching Process     ┊
  └╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌╌┘
               │
        ┌─────────────┐
        │     End      │
        └─────────────┘
```

Fig.4. The proposed FTT_MH_EMD algorithm flowchart.

The proposed FTT_MH_EMD algorithm starts with a recording of the signal at a sampling rate of 44100 samples per second. It is recorded for a pre-specified period of time, a 3 second time interval as usually an answer with a simple reply would not require any more time for VSA systems. The following step of the proposed FTT_MH_EMD algorithm, using two LPF filters (250Hz then 25Hz), marked by a dashed line box in **Fig.4,** comes to speed up the processing time of the MH_EMD algorithm to decompose the speech signal and hence decrease the whole processing time of the proposed FTT_MH_EMD. Two LPF filters step is used to decrease the MH_EMD iteration numbers as it selects frequencies lower than 25 Hz only from the input speech signal to decompose it rather than decomposing all the frequencies of the speech signal. Frequencies higher of that value are either of no importance. The type of filter to be used is an equiripple filter as responds quickly and accurately to the voice signals. The filtering process carried on the signal is zero-phase filtering which increases accuracy due to filtering the data in both the forward and reverse directions.

For a speech signal, finding the micro tremor frequency by using EMD is extremely difficult and near to impossible due to various factors that prevent isolation of only the speech signal.

To overcome this, the next step after the proposed MH_EMD algorithm, using FTT with a fetching process, marked by a dashed line box in **Fig.4,** is introduced here to get the micro tremor frequency as fast as possible with higher accurate result. That is done by converting the last signal achieved by the MH_EMD algorithm is converted into the frequency domain by using FFT and plotted in a periodogram displaying frequency against Power in watts and by choosing a standard

deviation (SD) $0.1 \leq SD \leq 0.3$, we are capable of separating the surrounding vibration frequencies or the adjacent frequencies from the actual voice signal, , then the proposed FTT_MH_EMD algorithm can find the micro tremor frequency by using a fetching process that searches between 8-12Hz and selects the minimum frequency with minimum amplitude.as only one frequency representing a speech signal as shown in **Fig.7**.
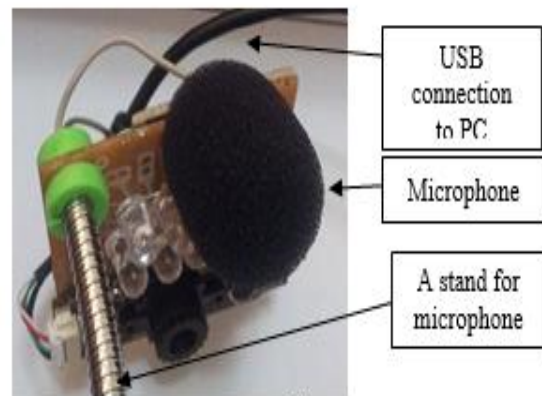
6. VSA SYSTEM MODEL FOR SIMULATION

In this section a system model for simulation which is divided into two main parts; a hardware and a user friendly GUI is proposed.

The proposed hardware used shown in **Fig.5,** could be simply defined as a microphone and the processing unit of HP latitude with a processing speed of 2.8 GHz. The proposed user friendly GUI shown in **Fig.8** and is done by using Matlab R2013a program for testing.

The CAD 900 tiny microphone in **Fig.5** is selected primarily with an Omni-directional feature to decrease the noise value as signals recorded from the subject's voice. The microphone was also preferred over other types due to its higher specifications and smaller size and weight.
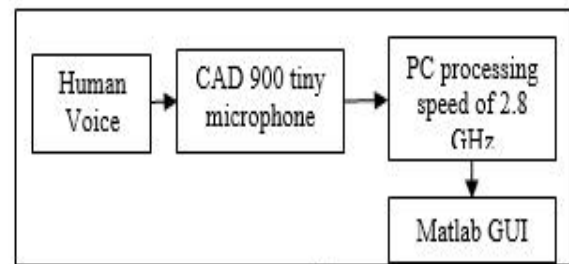
Fig.5. VSA system model for simulation, (a) Block diagram, (b) Implementation.

7. RESULTS AND SIMULATIONS

Fig.6 and **Fig.7** show the periodogram of different decomposed voice signals for different cases. In **Fig.6**, there was no input signal therefore no micro-tremor frequency produced. The proposed algorithm detects that the input was only noise vibrations from the surrounding. In **Fig.7**, there was a voice input signal and the proposed algorithm produces the micro-tremor frequency.

As seen from **Table1**, the processing time of the proposed algorithm FTT_MH_EMD is faster than that of the conventional EMD where the total time taken is around 10 seconds which includes operations not concerned with the code such as accessing the browser. The self-time of the code is 8 seconds. Due to the novel technique of (FTT_MH_EMD) is combined of FFT and EMD with Modified Hilbert Analysis, the processing time of the proposed algorithm FTT_MH_EMD was brought down into the 3 second range which is much faster than the inspected research papers.
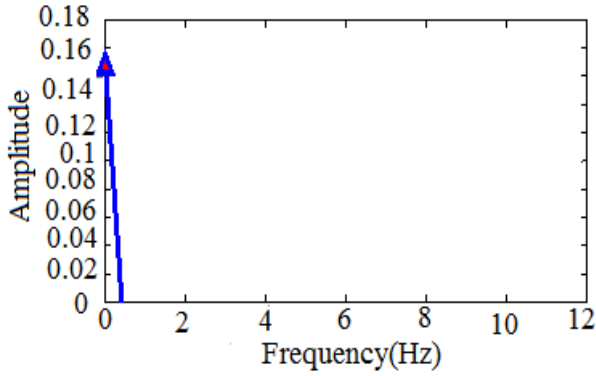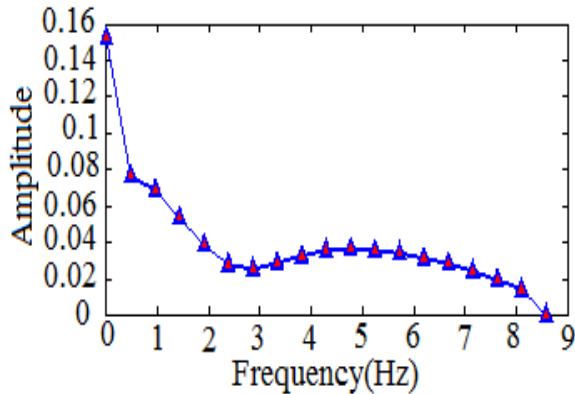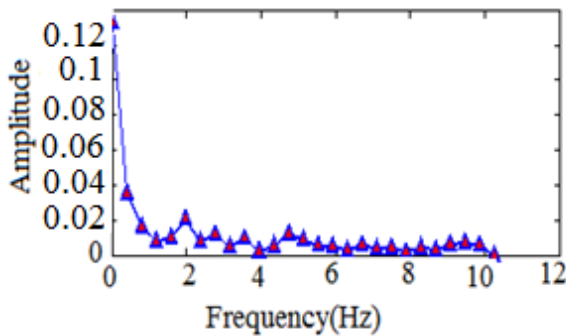

Fig.6. Output of Noise input or no voice input signal.

Table 1. Processing time for a 3 second speech signal.

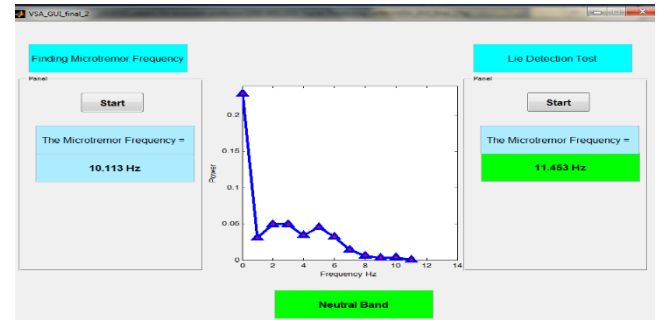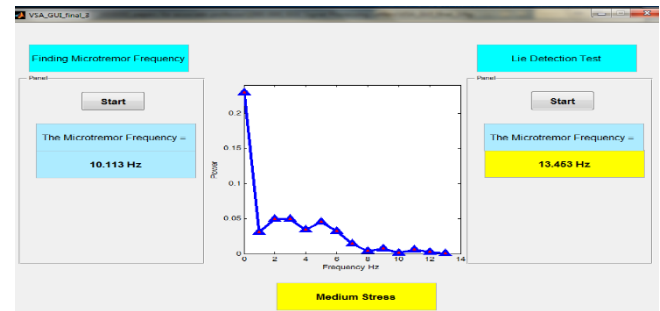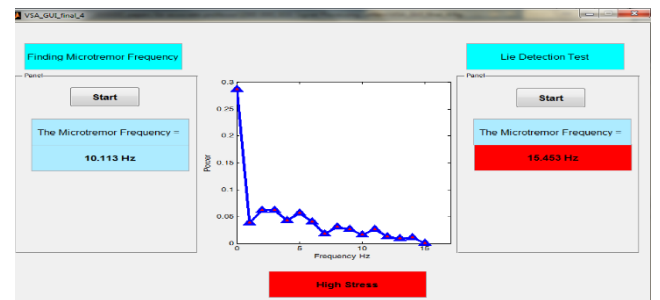| Algorithm	Self-time	Total-time
Conventional EMD using Cubic Spline Interpolation	≈ 6 minutes	≈7 minutes
Proposed FTT_MH_EMD	≈ 8 seconds	≈ 10 seconds

(a)

(b)

Fig.7. Output micro-tremor frequency FTT_MH_EMD, (a) person1: 8.79Hz, (b) person 2:10.22Hz.

(a)

(b)

(c)

Fig.8. The proposed VSA GUI, (a) Neutral band (8-12Hz), (b) Medium-Stress band (12-15Hz), (c) High-Stress band (15-17Hz).

Fig.8 shows the GUI of the proposed FTT_MH_EMD for the proposed VSA system model. When pressing the VSA pushbutton in the left hand-side region, the GUI clearly shows that recording will start after the beeping sound. A loud beep is then produced after which the subject should answer the question. The voice signal is then processed by the methodology of the proposed algorithm FTT_MH_EMD. Results, graphs and a decision are then produced automatically and quickly.

As seen from **Fig.8,** the proposed VSA GUI during the examining process can interact with the user by showing to him if the examined person is under stress or not by sending a "Neutral Band with green color" massage for micro-tremor frequency range (8-12Hz), or "Medium Stress with yellow color" massage for micro-tremor frequency range (12-15Hz) to indicate the honest subjects, otherwise sending a "High Stress with red color" massage for micro-tremor frequency range (15-17Hz) to indicate the deceptive subjects, and that because of the short processing time of the proposed algorithm

FTT_MH_EMD. **Table2** shows the summary of the achieved results. A test was implemented to test the validity, robustness and accuracy of the proposed VSA system model. Fifteen subjects were chosen over a period of two weeks. The subjects were seated into typical rooms which had an average illumination and low surrounding noise. The room contained a flipped over document and the subjects were allowed to enter with their mobile phones. The subjects are then told that they are allowed to either read the document or completely ignore it. The examiner leaves the room for a period of 5 minutes and returns for the examination process.

Table 2. Testing Results.

Question Number	Answers Number	Deceptive Subjects	Detected	Honest Subjects	Detected
1	15	9	9	6	5
2	15	12	10	3	3

The subjects are first asked irrelevant easy questions then are asked the two main questions. These are "Have you looked at the document?" and "Have you used your mobile phone?" Subjects were informed before the examination that they have the freedom of being deceptive or honest. After analyzing the table above, it can be seen that out of the 30 relevant questions, 27 where accurately detected whereas 3 where not detected. For an initial prototype device a 91.3% accuracy rate is very promising. Using higher quality hardware components and placing the subjects in a stronger interrogation atmosphere will likely cause the accuracy to increase.

8. CONCLUSIONS

In this paper, a new algorithm was proposed known as FTT_MH_EMD, in which a new technique was used. This technique used a modified Hilbert analysis method to detect the input voice signal envelope and the FTT method to speed up and facilitate the VSA processing time. The proposed FTT_MH_EMD algorithm was manipulated where a novel technique was used to produce the signal envelopes and speed up the iteration process up to ten times compared with the conventional EMD algorithm.

9. REFERENCES

[1] Hopkins, Clifford S., et al., "Evaluation of voice stress analysis technology", Proceedings of IEEE International Conference on System Sciences, 38th Annual Hawaii International Conference (HICSS'05), 2015.

[2] M. H. Beers and Robert Berkow, "The Merck Manual of Diagnosis and Therapy", 17th Edition, John Wiley & Sons 1999.

[3] Zhang, James Z., et al. "Analysis of stress in speech using adaptive empirical mode decomposition", Signals, Systems and Computers, 2009 Conference Record of the Forty- Third Asilomar Conference on. IEEE, 2017.

[4] Hopkins, Clifford S., et al. "Evaluation of voice stress analysis technology", System Sciences, 2015. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on. IEEE, 2015.

[5] James Z. Zhang and et al., "Analysis of stress in speech using adaptive empirical mode decomposition", Proceedings of IEEE International Conference on Signals, Systems, and Computers, 43rd Asilomar Conference, Pacific Grove, CA, pp.361-365, 2009.

[6] N. Mbitiru and et al., "Analysis of stress in speech using empirical mode decomposition", Proceedings of The 2008 IAJC-IJME International Conference, ISBN 978-1-60643-379-9, 2018.

[7] Olof Lippold, "Physiological Tremor", Technical paper, Scientific American, Volume 224, Number 3, March 1971.

[8] D. Haddad and et al., "Investigation and Evaluation of Voice Stress Analysis Technology", In-House Technical Memorandum, pp. 5-7, November 2001.

[9] R. Deering and J. F. Kaiser, "The Use of a Masking Signal to Improve Empirical Mode Decomposition", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.4, pp.485–488, 2015.

[10] N. E. Huang and et al., "The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis," Proceedings of the Royal Society London A., vol. 454, pp.903–915, 2009.

[11] K. Zeng and M-X. He., "A Simple Boundary Process Technique for Empirical Mode Decomposition", Proceedings of IEEE International Geoscience and Remote Sensing Symposium, vol.6, pp. 4258–4261, 2018.

[12] Rilling, G., P. Flandrin & P. Gonçalvès, "On Empirical Mode Decomposition and its algorithms", IEEE-EURASIP Workshop on Non-linear signal and Image Processing NSIP-03, Grado, Italy.

[13] N. E. Huang and et al., "A Confidence Limit for the Empirical Mode Decomposition and Hilbert Spectral Analysis," Proceedings of the Royal Society London A., vol. 31, pp.417–457, 2003.

[14] Norden E. Huang and Zhaohua Wu, "A review on Hilbert-Huang transform: Method and its applications to geophysical studies," AN AUG Journal, American Geophysical Union, Reviews of Geophysics, 46, Paper number 2007RG000228, 2018.

[15] Huang, Norden E., "Introduction to the Hilbert-Huang transform and its related mathematical problems", Interdisciplinary Mathematics vol.5, pp.1-26, 2015.

[16] Bendat, J.S., Piersol, A.G., "Random Data: Analysis and Measurement Procedures", John Wiley & Sons, Inc., (1986)