Analysis of Stress in Speech Using Empirical Mode Decomposition

Nyaga Mbitiru, Peter Tay, James Z. Zhang, Robert D. Adams
Department of Engineering and Technology
Western Carolina University
Email: zhang@email.wcu.edu

Abstract

Voice stress analysis (VSA) is accomplished by measuring fluctuations in the physiological microtremor present in speech. In this paper, Empirical Mode Decomposition is compared to traditional Fast Fourier Transform in the analysis of the physiological microtremor. The results are expected to show that EMD is better suited in the detection of stress in voice.

Introduction

Stress is a constant factor in most peoples' lives, often at times it is undesirable but other times it is beneficiary. Detecting stress and the level can be crucial in applications that are stress sensitive such as voice activated military equipment, psychological testing and deception detection. Since the discovery of the physiological microtremor by Lippold [1], research into the relationship between psychological stress and its effect on the microtremor has been conducted. While the evidence supporting a direct correlation between the two is disputed, certain companies have been successful in marketing lie detector software based on the microtremor. Voice stress analysis (VSA) is accomplished by measuring fluctuations in the physiological microtremor present in speech. However, the accuracy of VSA is disputed among experts. This is mostly due to the different algorithms in use as well as the effectiveness of the examiners. Algorithms such as Fast Fourier Transform (FFT) as well as McQuiston-Ford algorithm are some of the algorithms used. The latter includes a scoring system. An analysis method that is capable of consistently determining the stress level of an individual at low to medium stress levels regardless of examiner effectiveness is needed.

Physiological Microtremor

Stress is defined as the disruption of (homeostasis) by physical or psychological stimuli. Physical factors such as noise, excessive heat/cold, and psychological factors such as emotion and sleep deprivation alter the internal equilibrium of the body causing a stress response. The General Adaptation Syndrome (GAS) is a model by Hans Seyle [3] that identifies the various stages of the stress response. The first stage is known as the alarm stage, here the body identifies the stressor or threat and goes into a state of alarm. Adrenaline is produced in order to prepare the body for fight or flight, this cause's blood flow to be diverted to the large muscles of the body as the body prepares to run away or fight. In addition to adrenaline, another hormone known as cortisol is also produced. Cortisol is known as the 'stress hormone' and increases blood pressure and blood sugar in order to restore the body's homeostasis after stress. The second stage is known as the resistance stage, during this stage the body attempts to cope with the stress by adaptation. As the body tries to cope with the stress it uses up it’s resources. The third stage is appropriately known as the exhaustion stage. This occurs when the body's resources are used up and it is unable to maintain normal function.

The first and second stages of stress are of particular interest as increased muscle tension occurs during these stages. This increased tension affects all muscles in the body including the vocal chords. An increase in tension may directly or indirectly affect the production of speech. In particular increased tension affects the physiological microtremor that is present in speech. A microtremor is a low amplitude oscillation of the reflex mechanism controlling the length and tension of a stretched muscle caused by the finite transmission delay between neurons to and from the target muscle [1]. Microtremors are present in every muscle in the body including the vocal chords and have a frequency of around 8 – 14 Hz. During times of increased stress this microtremor increases in frequency and decreases in amplitude [1]. This change in frequency transfers from the muscles in the vocal tract to the voice produced. Stress can thus be detected by analyzing the change in microtremor frequency of an individual’s voice.

Voice Stress Analysis

Detection of stress by voice analysis has numerous applications most notably in military, law enforcement and emergency services. Military applications are of greatest interest as individuals are often under some sort of stress, be it physical or emotional. Physical stress can stem from sleep depravation, environmental extremes and exhaustion. Emotional stress from fear or confusion from conflicting information. Elevated stress levels can adversely affect the performance speech recognition equipment which is a concern especially when military equipment is concerned. Additional coding can be incorporated into the equipment to identify stressed speech and perform the appropriate corrections in order to maintain optimum performance [2].

Deception detection is a valuable application in both military and law enforcement. Voice stress analysis can be used to detect elevated stress levels caused by deception. Voice stress analysis is non-intrusive and does not require any physical connection to the subject in question so is therefore also suitable for clandestine functions.

Voice stress analysis can be used in emergency services to direct calls to priority operators based on the stress level of the individual. This allows for priority responses to priority calls therefore increasing the quality of service [2].

Most VSA products use the physiological microtremor in combination with other voice features such as pitch, tone, and fundamental frequency as a descriptor of an altered psychological state. Using signal processing as well as knowledge of microtremors, the products claim to be able to identify if an individual is in a stressed state.

Empirical Mode Decomposition

The physiological microtremor, much like the human voice is a nonlinear, non-stationary process. The frequency and amplitude of the microtremor varies considerably in time and current methods of voice stress analysis that utilize signal processing methods such as the FFT lose time resolution as they provide and average over time rather than instantaneous values. This considerably reduces the effectiveness of stress detection, providing an overview rather than a complete picture. Empirical Mode Decomposition (EMD) is a new mathematical method that provides a means of decomposing nonlinear, non-stationary signals into the sum of a series of stationary signals; this allows fluctuations in frequency and amplitude to be detected in time. [3]

EMD works by identifying time scales that uncover the physical characteristics of the signal. The signal is decomposed into numerous stationary signals that represent these physical characteristics, these are known as Intrinsic Mode Functions (IMFs). Specifically, IMFs are signals that satisfy a certain set of criteria:

1) The number of extrema and number of zero crossings throughout the dataset must either be equal or differ by one at most.
2) At any point, the mean value of the envelope defined by local maxima and the envelope defined by the local minima is zero.

While the first condition is similar to the traditional narrow band requirements for stationary Gaussian processes, the second condition alters the classic global requirement to a local one; this ensures that the instantaneous frequencies will not have unwanted fluctuations induced by asymmetric wave forms. An IMF can be both amplitude and frequency modulated and, can even be a non-stationary signal. By utilizing EMD frequency contents of a signal can be observed without sacrificing time resolution as is typical with time domain to frequency domain transformation. [3]

EMD is an iterative process which produces IMFs until a stopping criterion has been satisfied. The stopping criterion can be the standard deviation between two consecutive results. The process can be summarized as follows:

1) Upper and Lower Envelopes of the signal are constructed with its maxima and minima using cubic spline function
2) Mean of the envelopes is subtracted from the signal to obtain a new signal
3) Determine if the new signal is an IMF using the criteria described above.
4) If the new signal is indeed an IMF, it is subtracted from the original signal and the resulting new signal goes through the above process until an IMF is obtained

Having obtained IMFs from the EMD process, the Hilbert Huang Transform (HHT) is used to represent the original signal as a linear combination of the real parts of the analytic function constructed with the i^{th} IMF and a residue term r_n. Amplitude can be represented as a function of frequency and time; this distribution is designated as the Hilbert Spectrum. The measure of total amplitude contribution from each frequency and the instantaneous frequency can be obtained from this spectrum. [3]

Speech under Simulated and Actual Stress

Speech samples from the Speech under simulated and Actual Stress (SUSAS) database were used to isolate and identify features unique to stressed voice. The database contains voice samples of 44 speakers, both male and female speaking in five different domains; Talking Styles, Single Tracking Task, Dual Tracking Task, Actual Speech Under Stress, Psychiatric analysis. The scope of the database allows for an analysis of stress under different types of stress and different speaking styles. Voice samples from the SUSAS database have a 16 bit sample depth and 8 kHz sample rate. The voice samples are stored in uncompressed Pulse Code Modulation (PCM) format to preserve as much information as possible.

Signal Analysis

Both the traditional FFT and the newer EMD were used to analyze the microtremor frequency component of the voice samples. In order to compare their effectiveness in detecting fluctuations in the microtremor each process was used to analyze voice samples of a speaker saying ‘help’ in 3 different styles under both simulated and actual stress conditions. In addition to this, the energy and correlation coefficient of the sample to a baseline neutral sample were also computed. The complexity of the human voice as well as the numerous harmonics requires that more than one feature of the voice be used to detect stress. Correlation and total energy are considered supporting features.

Results

In order to test the effectiveness of FFT and EMD six voice samples with neutral, low task stress and high task stress speaking styles were used. In the first three samples, the speaker is in a quiet environment and the stress is simulated whilst in the latter the speaker is recorded while on a roller coaster and the stress considered actual.

For testing FFT effectiveness the 8 – 14 Hz component of the voice samples was filtered out and data edges trimmed. The resulting signal was then plotted and compared to the other samples. The charts below show the difference between the simulated neutral sample and the high task stress sample.

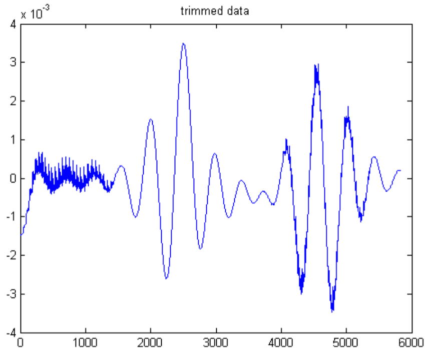

Fig.1: Neutral Sample

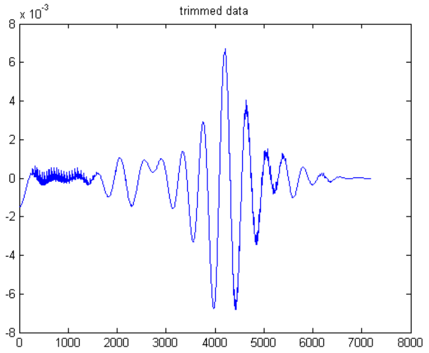

Fig. 2: High Task Stress Sample

The energy and correlation to the other five samples was then computed. Table 1 shows the energy in watts and correlation coefficient between the simulated neutral sample and the other five samples.

Table 1.

Voice Sample	**Energy (Watts)**	**Correlation Coefficient**
Neutral – Simulated	0.0079	1.0000
Low Task Stress – Simulated	0.0238	0.2278
High Task Stress – Simulated	0.0221	0.1744
Neutral – Actual	1.9046	0.1248
Low Task Stress – Actual	1.9123	0.1159
High Task Stress – Actual	1.0192	0.1636

EMD in combination with the HHT is especially powerful for combining frequency and amplitude in time as seen in figure 3 below.

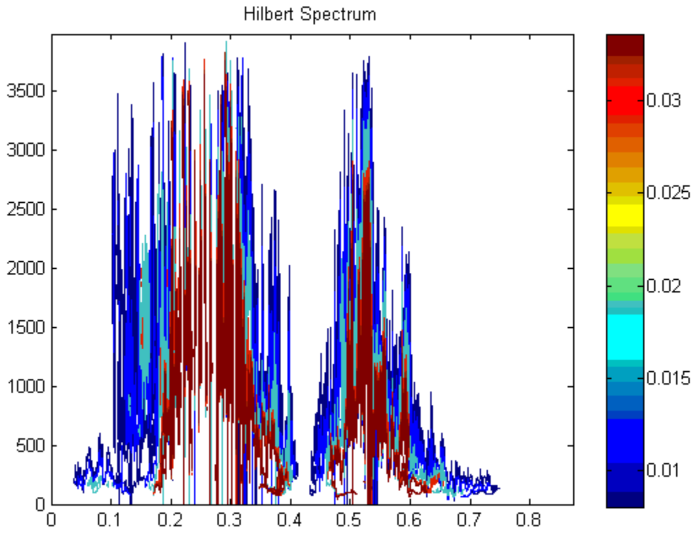

Figure 3. High Task Stress

Conclusion

EMD provides a lot of detail and can be modified to suit the application at hand. While FFT has remained and will continue to remain a staple in signal analysis, EMD provides magnitudes on a time-frequency plane. This allows for specific fluctuations to be identified in time further improving the detection of stress. Further investigation into features unique to stress as well as modification of the EMD algorithm will increase the success rate and capability of detecting stress in voice.

References

[1] O. Lippold, “Physiological Microtremor,” Scientific American, 224(3): pg. 65-73. 1971.

[2] D. A. Cairns, J. H. L. Hansen, “Nonlinear Analysis and Detection of Speech Under Stressed Conditions,” Journal of the Acoustical Society of America, vol. 96, no. 6, pp. 3392-3400, 1994.

[3] H. Seyle, “The General Adaptation Syndrome,” Annual Review of Medicine, Vol. 2: pg. 327-342, 1951.

[4] N. E. Huang, Z. Shen, S.R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” Proc. Roy. Soc. Lond. A, 1998, pp. 903-1005

Biography

NYAGA MBITIRU is currently a second year graduate student in the M. S. Technology program of the Kimmel School of Construction Management and Technology at Western Carolina University.

PETER TAY is currently a visiting professor at the Kimmel School of Construction Management and Technology at Western Carolina University.

JAMES Z. ZHANG is currently an associate professor and M. S. Technology program director at the Kimmel School of Construction Management and Technology at Western Carolina University.

ROBERT D. ADAMS is currently an assistant professor at the Kimmel School of Construction Management and Technology at Western Carolina University.