INTERNATIONAL JOURNAL OF RESEARCHES IN BIOSCIENCES, AGRICULTURE & TECHNOLOGY
© VISHWASHANTI MULTIPURPOSE SOCIETY (Global Peace Multipurpose Society) R. No. MH-659/13(N)
www.vmsindia.org

Recognition of Fatigue Detection from Voice

[1]Mayura Mathankar and [2]Sandhya Dhabe
[1]Sydenham College of commerce and Economics B Road ,Churchgate ,Mumbai 20
[2]Institute of science R T road ,Civil lines Nagpur

Abstract
The working conditions of professional drivers are characterized by long working hours, movement restriction, dim light levels, background noise, and infrasonic vibration. All of these factors are known to cause fatigue, driving without awareness and even microsleep events. Thus, the prediction and warning of professional drivers against impending critical fatigue could play an important role in preventing accidents and the resulting human and financial costs. Hence, many efforts have been reported in the literature for measuring fatigue related states. But these efforts still do not fulfill the demands of an everyday life measurement system. The major drawbacks are lack of robustness against environmental and individual-specific variations (e.g. bright light, wearing correction glasses, angle of face or being of asian race) and lack of comfort.
In this paper the study was carried out to construct and validate a non-obtrusive fatigue detection system based on speech as used in operator communication systems. The main findings may be summarized as following. First, acoustic features, that were extracted from speech and subsequently modelled with pattern recognition methods, contain a substantial amount of information about the speaker's fatigue state. Our acoustic measurements showed differences between alert and fatigue speech in fundamental frequency, intensity, formants, and duration features.

1. INTRODUCTION

Military and civilian experience has shown that long duration assignments present increased risk of performance failures as the mission progresses. This is due to interruption of normal sleep cycles and psychological pressures of the work environment. There continues to be a need for a non-intrusive fatigue assessment system to successfully monitor the level of alertness of personnel during critical missions and activities. Experimental results on human voice show that specific phones have a predictable dependence on fatigue. Hence, precise phonetic identification and alignment are important to voice-based fatigue detection. This project explores techniques for detecting fatigue from voice using speech recognition to obtain phonetic alignments. In this project we restricted our analysis to dealing with out-of-vocabulary words.

Non-intrusive fatigue assessment systems are crucially needed to successfully monitor the level of alertness of all personnel during critical mission or life-threatening activities. This project explored the use of automatic speech recognition (ASR) to detect fatigue from voice. There are numerous challenges which have to be overcome in order to have reliable fatigue detection systems based on voice. However, advances in speech recognition technology have made it possible to obtain good performance even in noisy environments, and hence, the technology has found widespread application in recent years.

2. FATIGUE DETECTION FROM VOICE

2.1 Using Voice to Detect Fatigue

Applications of speech recognition have grown from simple speech to text conversion to other more challenging tasks. A relatively new application using voice is in the field of cognitive analysis. The main goal is to detect the mental preparedness of a worker before critical missions, based on cognitive measures such as fatigue. Speech is one attribute of human behavior that can be used to measure fatigue. People working in stressful environments such as military and aviation are more susceptible to fatigue than others, and accidents by such workers are often fatal. Complex instrumentation often creates cognitive overload and places a greater demand on the crew to be vigilant.

A prescribed remedy for fatigue is sleep. It is also been studied that the effect of sleep on fatigue and determined that the quality of sleep is more important than the number of hours of sleep. This makes the task of monitoring fatigue even more challenging. There is no well-accepted non-intrusive technique to measure quality of sleep. This thesis explored the potential for using voice to perform real-time fatigue detection.

2.2 Using Automatic Speech Recognition for Fatigue Analysis

Voice has been successfully used in many applications other than simple speech to text conversion. Applications involving speaker verification, deceit detection, reading tutors, automatic language recognition and translation are actively being developed. These applications share core technology based on a statistical approach to speech recognition. The three approaches are speaker verification, word spotting, and ASR. It was determined that an ASR approach was most promising. Each of these approaches are briefly described below.

Speaker verification is the task of verifying a subject's authenticity based on his or her voice characteristics. This process entails two phases: enrollment and verification.

During enrollment a speaker model is built using a subject's speech. During verification, this speech is used as a template to verify the authenticity of the speaker. A speaker verification system was used as a primitive change detection system to determine whether systematic variations in the long-term statistics of the speech signal due to fatigue could be modeled using a classic pattern recognition paradigm. The results from this approach were not promising since a clear variation in the likelihood ratio scores between the speakers and the models as a function of fatigue was not found.

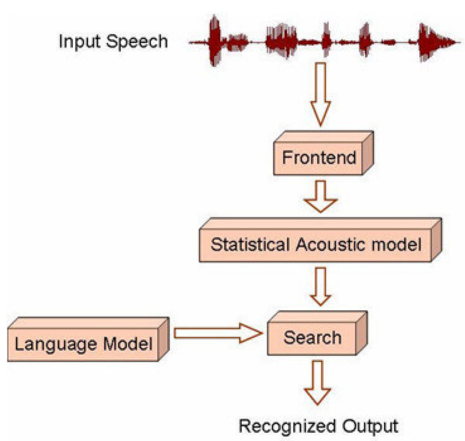


Figure 2.1: Basic components of a large vocabulary speech recognition system

2.3 Automatic Speech Recognition

The ASR system used in this thesis is a public domain LVCSR system developed by the Intelligent Electronic Systems (IES) program at Mississippi State University. A speech recognition system, shown in Figure 1, consists of the following four blocks: feature extraction, language model, acoustic model and search. Feature extraction converts the incoming signal to a stream of vectors, and typically uses an MFCC approach. The acoustic model is a statistically trained model that learns the temporal and spectral characteristics of the speech signal. A language model is used to guide the recognizer with some a priori information about the language of interest in the application. The search block typically uses a Viterbi decoding algorithm and finds the best path through the search space using language model and acoustic model probabilities. The entire speech recognition framework can be represented using Bayes Rule. The words can be divided into sub-units called phonemes. There are approximately 46 phones in English language. The acoustic models can model words or phonemes.

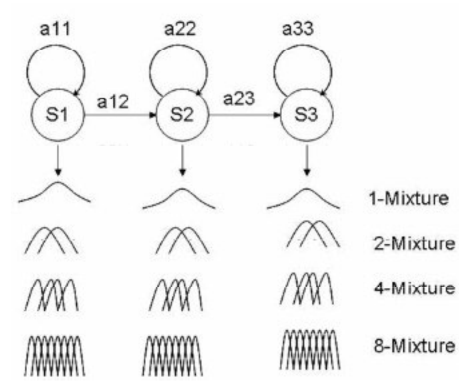


Figure 2.2: Three-state HMMs used to model phones in an ASR system

Each model is represented with an HMM that is implemented using a stochastic finite state automaton. For limited vocabulary tasks, word models achieve high performance. However, for large vocabularies that often consist of more than 100,000 words, word models are not practical and hence phone models or cross-word triphone models are used.

Phonetic models, often referred to as phone models, are typically used for large vocabulary systems. The articulation of words in human speech is context dependent, i.e. the articulation of a particular sound depends on its surrounding sounds. It has been determined that context-dependent phones that model the current phone in the context of the previous and next phone is a reasonable compromise between performance and complexity. Context-independent and context-dependent phone models are often represented by a three-state HMM as shown in Figure 2.

Each state in an HMM can be represented by a Gaussian mixture model (GMM). Recognition performance generally increases as the number of mixture components in the GMM increases because an increase in the number of mixture components improves the ability of the GMM to model arbitrary distributions.

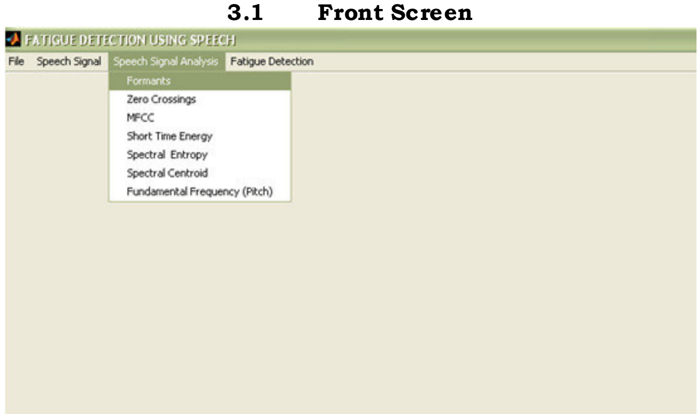
2.4 Acoustic Correlates of Fatigue in the Speech Signal

Studies on military aircrews operating B1 bombers showed that voice had a similar pattern to that of other cognitive measures of fatigue. An automatic voice- based fatigue detection system should use fatigue cues present in the speech signal. In order to determine these fatigue cues, one needs to understand the changes that occur in human speech as a person becomes fatigued. Literature suggests that there is a spectral and temporal variation in the speech pattern as humans become increasingly fatigued.

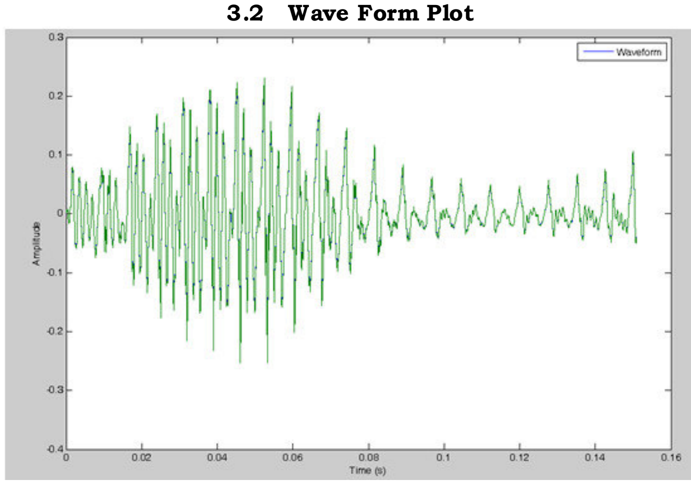
Sounds produced by humans can be represented as a convolution of the excitation signal and the vocal tract characteristics. The excitation can be modeled as either a periodic signal or noise. For voiced speech sounds, the excitation can be modeled as a periodic signal whose fundamental frequency is determined by the vibration of the vocal cords. The vocal cords close temporarily to increase the air pressure generated from the lungs, and they open when the pressure exceeds the resistance of the vocal cords. The vocal cords vibrate due to a combination of factors, including their elasticity, laryngeal muscle tension, and the Bernoulli effect. The opening and closing continues as long as the lungs pump air through the vocal cords and into the oral cavity.

3. DESIGN & IMPLEMENTATION

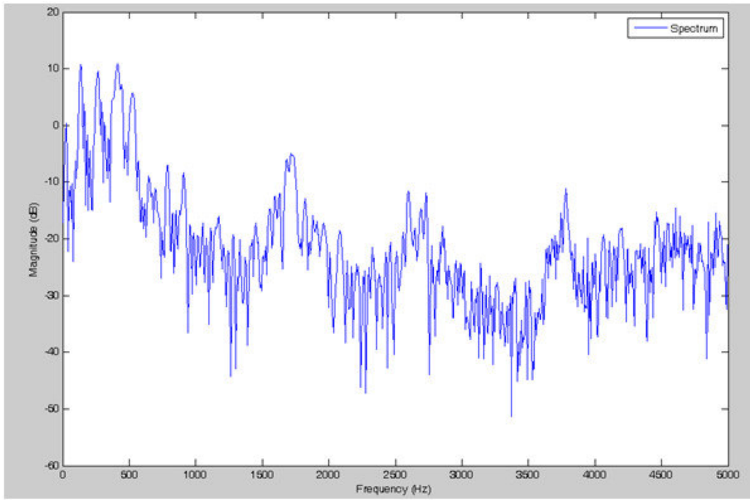
3.1 Front Screen



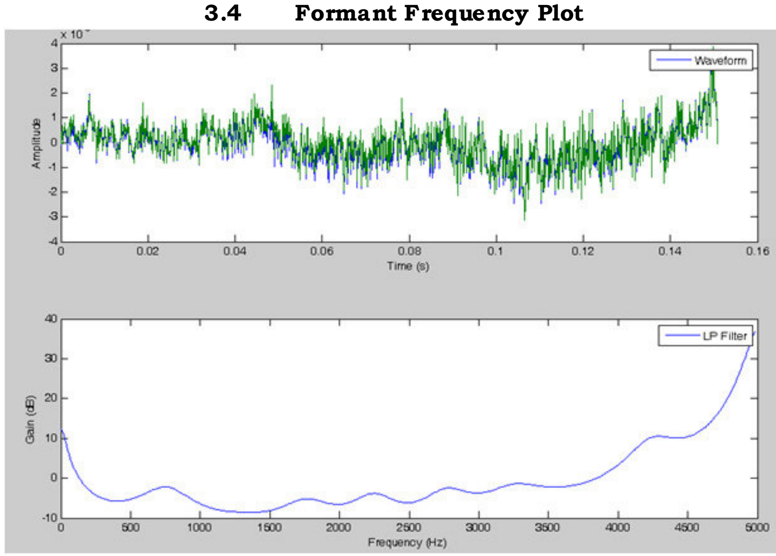
3.2 Wave Form Plot



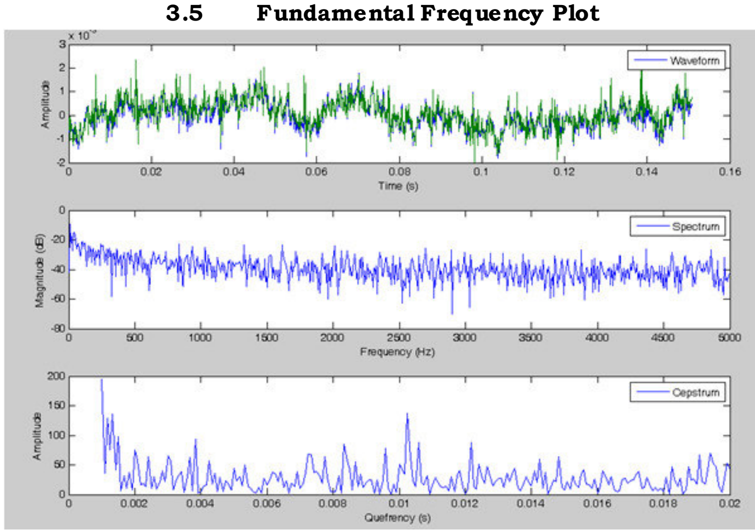
3.3 Spectrogram Plot



3.4 Formant Frequency Plot



3.5 Fundamental Frequency Plot



4. RESULT

- Feature Vector for Morning Speech of subject

0.1005 0.2156 0.0269 0.0408

0.0078 0.1901 0.0333 0.0195 42.0177

0.3159 0.0048 0.2903

- Degree of Energy of subject in Morning =4.75406
- Pitch in Morning =1000Hz
- Degree of Fatigue is Low

Feature Vector for Evening Speech of subject
0.1105 0.1682 0.0279 0.0450 0.0076 0.2188 0.0350 0.0251 35.9439 0.4316 0.0028 0.2399

- Degree of Energy of subject in Evening =2.83346
- Pitch of subject in Evening =1000Hz
- Degree of Fatigue is More

5. CONCLUSION

Non-intrusive fatigue assessment systems are needed to successfully monitor the level of alertness of all personnel during critical mission or life-threatening activities. This thesis explores the first attempt at detecting fatigue from voice using an ASR system. Various approaches such as speaker verification, word spotting and LVCSR techniques were analyzed in this thesis, and the LVCSR approach was found to be superior for this particular task. The LVCSR approach did not require fatigue-dependent data for training and it used a fixed grammar. LVCSR approach was relatively more effective when dealing with OOVs, as compared to the word spotting approach. The OOVs caused insertion and substitution errors in the final output. The fatigue detection system treated the insertions and deletions generated by the LVCSR system as false alarms. The problem of false alarms was tackled by implementing a confidence measure algorithm. The LVCSR system output was annotated with a word posterior-based confidence measure. The confidence measure was used to filter out false alarms. Use of the confidence measure improved the robustness of the fatigue detection system to OOVs by 20%.

REFERENCES

1. J. Whitmore and S. Fisher, “Speech During Sustained Operations,” *Speech Communication*, vol. 20, no. 2, pp. 55-70, November 1996.
2. H. Greeley, J. Berg, E. Friets, J. Wilson, G. Greenough, J. Piconeand J. Whitmore, “Fatigue Prediction Using Voice Analysis,” submitted to *Behavorial Research Methods*, February 2006.
3. W.M. Campbell, J.R. Campbell, D.A. Reynolds, D.A. Jones, T.R. Leek, “High-level Speaker Verification with Support Vector Machines,” *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 73-76, Montreal, Quebec, Canada, May 2004.
4. J.E. Beck, K. Chang, J. Mostow, and A. Corbett, “Using a Student Model to Improve a Computer Tutor's Speech Recognition,” *Proceedings of International Conference on Artificial Intelligence in Education*, vol. 7, pp. 2-11, Las Vegas, Nevada, USA, June 2005.