

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

Diagnosing Mental Health Disorder Associated with Substance Abuse Using Speech/Voice Recognition

Elliot Kizzy Nkem¹, Kabari Ledisi Giok²

Department of Computer Science, Ignatius Ajuru University of Education, Rivers, Nigeria^{1,2}

ABSTRACT: One of the major ways through which humans communicate with one another is through speech. The possibility of converting human speech or spoken words into alphanumeric text and control commands is achieved through speech recognition. The technology allows humans to talk to a machine and give it some set of instructions which it will follow. This paper presents how speech/voice recognition technology can be used to detect and diagnose mental health disorder associated with substance abuse. We obtained data sets for both normal and abnormal person and use it to plot their audio waveforms for comparison. Python programming language was used for the implementation and serious difference was observed which can be used for detection of substance abuse.

KEYWORDS: Mental health disorder, Substance Abuse, Artificial intelligence, Speech Recognition, Signal Classification.

I. INTRODUCTION

Making machines more like human beings has always been a strong desire of man especially the AI scientists. This desire has led to the emergence of disciplines like artificial intelligence (AI) which emulate human behaviour in machines. Due to the popularity of Artificial Intelligence in science fiction movies, its idea or concept is well established which portrays humans communicating with computers as they would with other beings. The principles used in AI include the ideas defined by Man Machine Interfacing (MMI), which allows the development of more human-usable robots. Speech/voice as well as actions since the dawn of civilization have been the basic, common and efficient forms through which we humans communicate and interact with one another. The advent and universal use of mobile phones, radio, and television etc. has made voice communication and amplification even more significant. Today, voice technologies are commonly available for a limited but interesting range of task. These technologies enable humans to communicate with the computer through voice commands and computers from their own end, answers precisely and reliably to these voice commands and provide the necessary services that are required of them. Communication among humans is dominated by speech, therefore, it is only natural for people to want voice interfaces that will enable them communicate with the computer directly as well.

Nowadays, as giving instructions to the computer is faster using voice commands, people now prefer it to that of the computer keyboard commands. Voice recognition technology is beginning to gain much grounds and is applied today mostly in access control systems to verify an individual's true identity. It uses the biometric technology as a means of validating or confirming someone's true identity and this is strictly based on the person's physiological characteristics, in this case, voice. With the level technology has gone today, there are also biometric systems that allows other physiological characteristics such as fingerprints or facial features etc to be used to identify and verify the true identity of a person but the purpose of this paper, we are only focusing on the voice recognition which is sometimes called speech recognition.

Voice recognition which is one of the few biometric approaches that have the advantages of high precision as well as low intrusiveness is basically all about talking to a computer or a machine and having it recognize what we are saying as well as it doing exactly what we want it to do in real time. Recently, voice or speech recognition has received significant attention as it has potential for a wide range of application. Google has applied it in their google speech-to-

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

text app and even Cortana has incorporated it as we can see in modern computer operating systems. We even have modern doors that uses this technology in granting access into a building only to the authorized person or persons. In this paper, therefore, we are looking at how we can use the speech or voice recognition as popularly called to detect and diagnose patient with mental health disorder which comes as a result of abuse of drugs or other substances in the society. The voice or speech recognition system will be that which will be able to detect signs of anxiety and depression in the speech patterns of the patients with mental health conditions and potentially providing a faster and easy way of diagnosing these conditions which could be difficult for the human expert to spot and which often are overlooked in people.

Mental health is an urgent and important world issue. About 450,000 people worldwide are suffering from mental illness. Significant mental illnesses are among the leading causes of disability, according to the World Health Organization (WHO). Mentally ill people have a mortality rate 2.22 times higher than the general population and mental disorder attributes approximately 8 million deaths each year. Among the top 10 things that causes death in our world today, suicide has been identified as one of them. It is on the record that in the United States alone, approximately 44,193 persons committed suicide as at 2015. Some of the current medical approaches for tracking the progression of disease, however, are insufficient, especially in Nigeria and Africa as a whole. Clinicians traditionally use face-to-face interactions for mental disorder assessment and diagnosis. This approach may not provide 100% accuracy on that regard which is why we need an alternative means of detecting and diagnosing mental health disorders. This study aims to develop a voice/speech recognition system that will be able to detect and diagnose patients with mental health disorders which has to do with drug abuse in the society. We investigate different cases of mental health disorders associated with mental abuse in hospitals, develop a system that will serve as a more precise alternative means of detecting and diagnosing mental health disorders and build a graphical User Interface for the system.

II. RELATED WORK

1. Mental Health and Substance Abuse

Psychological well-being is more than the simple absence of mental issue. The useful element of psychological well-being is highlighted in the definition of health by World Health Organization (WHO) which they included in their constitution. The definition quoting WHO goes as thus: "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." Ideas of psychological wellness incorporate personal well-being, independence, presumed self-efficacy, skill, intergenerational attachment and the ability to identify a person's emotional and intellectual capabilities. Mental health was as well described as a condition by which a person or set of persons realize their potentials, and with that, are able adapt to challenges of life and in return, make a good and sustainable impact in their respective communities. Mental health is primarily concerned with inspiring the competencies of every individual together with their communities and giving them all that is needed to reach their aspirations. Current research on substance abuse shows meaningful differences in gender in epidemiology, societal factors, and substance abuse characteristics. Furthermore, there are gender differences in the medical and mental health and treatment challenges associated with substance abuse [1][2].

Mental health presently has become an important global challenge. Approximately 450 million persons globally are faced with the challenge of severe mental illness, culminating in catastrophic individual consequences and enormous social pressure. The causes of these mental disorders are most often tied to the abuse of drugs, especially among the youths. Whether administered or not, every drug taken by an individual has an impact on the central nervous system (CNS) of his or her body. It affects people's physical and mental well-being when they take these drugs. Table 1 gives examples of the broad effects of substances on the way people think and feel.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

Table 1: Substance groups and their possible side effects

Substance Group	Possible Effects
Depressants (alcohol, solvents, sleeping tablets, benzodiazepines, e.g. diazepam)	Depress the Central Nervous System. Relief of tension and anxiety, Feeling of well-being and calm, Possible drowsiness and loss of concentration.
Stimulants (cocaine, ecstasy, amphetamine, steroids, khat, GHB)	Stimulate the CNS. Feeling livelier and more awake. Increased energy and heightened mood. Increased stamina. Nervousness, anxiety and sleep loss.
Alter perceptual function (cannabis, LSD, ketamine)	Change people’s perceptions. Heightened senses and mood. Visual distortions and hallucinations. Euphoria and feelings of floating
Reduce pain (heroin, codeine, methadone etc)	Reduce sensitivity to pain. Feelings of warmth and contentment. Sedation and sleep. Can mask symptoms of mental distress.

Emotional well-being ought to be a problem of everyone, not just for those who suffer from it. Problems associated with mental health or disorder affects the society massively in one way or the other. Mental health by this reason, consequently, is a major challenge for global development. No community is resistant to mental disorders. As a matter of fact, the risks are higher among the poor and homeless individuals, refugees, victims of violence, jobless persons etc. To everyone, psychological, physical and social wellbeing is closely intertwined with essential aspects of life. As time goes on, people will realize that mental health is vital to individuals, societies and countries ' overall well-being. Sadly, mental health and mental illnesses are not given the same priority as physical health in most parts of the world. Instead, people disregard it and see it as something that is not important.

2. Speech/Voice Recognition

Speech/Voice recognition are two of the several technologies that have evolved overtime. It is important to note that speech recognition and voice recognition are two different technologies with completely different objectives even though people use them synonymously to mean the same thing. According to Lawrence [3], “while speech recognition is aimed at converting speech signals to digital data, voice recognition is aimed at identifying the speaker.” Speech recognition is what makes human to computer interaction possible. You hear a version of it every day. We most times call our bank for balances, or call the local theatre for movie times and pricing. Again, most times when we call telecommunication companies like MTN, GLO or Airtel, we observe that we are often guided by an already programmed voice. That voice is a type of speech recognition program and it is called telephone-speech recognizers. Speech recognition is therefore the process of capturing spoken words using a microphone or telephone and converting them into a digitally stored set of words Lawrence [3].

The science of speech recognition has countless applications. This technology is widely used mostly for automated translations, dictation, hands-free computing, medical transcription, robotics, and much more. You've benefitted greatly from the speech recognition application if you've ever made payment through your mobile phone using an automated system. On the other hand, voice recognition functions by analyzing speech characteristics that vary among people. Different persons have different speech pattern that stems from their physiology (the shape and size of the throat and mouth) and behavioural patterns (the tone of their voice, their speaking style, accent, etc.). Voice recognition systems are significantly different from speech recognition systems. Most often, voice recognition technology is used to check the validity of a person speaking or to ascertain the identity of an unidentified speaker. Speaker verification and speaker identification are the categories of voice recognition identified by Lawrence [3].

According to [4], “Both voice and speech recognition as artificial intelligence areas have seen significant advances in signal processing, algorithms, computer architecture and hardware for the past five to ten years. Such developments include the rapid adoption of a methodology of statistical pattern recognition, an information-driven approach that makes use of a rich collection of speech expressions from a diverse population of speakers, the use of probabilistic

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

acoustic and language modelling, and the use of dynamic software-based query strategies[4]. There exist lots of voice recognition systems that are available nowadays some of which can recognize just a few words while others can recognize more.

Typically, because of their constraints and increased cost, voice recognition systems can only be used in a few specialized cases. Particularly when users are unable to use a keyboard to enter data, they are used as an alternative to keyboards. This may be due to the occupation or weakness of their feet. Nevertheless, users can probably talk to a headset instead of entering instructions using the computer keyboard Nasir et al. [5].

3. Speech production and Signal Classification

Speech is a sound produced by the human voice mechanisms. Voice speeches is however introduced to the medium through vibration and transmit to listener. The frequency of vibration of these voices in air medium is called the fundamental frequency Heng [6]. Different person would have different speech pitch. The vocal tract is the main physical defining factor in speech. The vocal tract is commonly considered beyond the vocal folds as the organs of speech production. Adult human vocal tract systems consist of this five which are laryngeal pharynx, oral pharynx, oral cavity, nasal pharynx and nasal cavity Nasir et al. [5]. People pronounce approximately four syllables, ten to twelve phonemes and two or three words out of their vocabulary (which can include 10 to 100 thousand words) per second in ordinary fluent speech[7][8]. Errors in the production of speech are relatively rare in spontaneous speech at a rate of about once in every 900 words. Words usually spoken or heard in the early days of life or easily imagined are quicker to say than ones that are rarely said, learnt later in life, or are abstract[9][10].

Typically, adult female voice average fundamental frequency is higher compare to male. [6] opines that a typical adult male speech average pitch from 85 to 180 Hz while that of the female speech average pitch is from 165 to 255 Hz. For an unified voice-frequency distribution stream, the bandwidth reserved is typically 4 kHz for humans. Heng [6] argued that the required sampling frequency for the system is at least twice the maximum range of fundamental frequency based on Nyquist-Shannon sampling theorem.

Speech recognizers can be put into different categories by defining the kind of utterances they can recognize which include isolated words, connected words and continuous words. Isolated word is referring to single word. Whilst the Networked word system is identical to segregated words, it allows a lowest break between them to run different speech together. Eventually, constant speech recognizers enable users to talk quite spontaneously whereas the content is determined by the machine. Recognizer with constant speech abilities are among the hardest to construct because they use different methods to determine limits of speech Heng [6].

In [11], his work was on Stress Evaluation by Voice: From Prevention to Treatment in Mental Health Care, he believes that stress-related mental illnesses alter emotional expression and lead to changes in some vocal qualities. Here, through a computer algorithm, the author tried to measure stress frequency using vocal emotion recognition technology. Comparing its proposed program with evaluations through self-administered surveys (GHQ-30) and interview evaluations, the envisaged technology obtained almost the same susceptibility to detection as GHQ-30. According to the researcher, "the ease of deployment and low cost are the key benefits of the current technology. This is why it is encouraging to apply it in preventive medicine and, if used in tandem with the multiple biomarker-based testing techniques currently under development, it will be possible to further persuade health care specialists."

Adarsh et al. [12], as related to a voice-based biometric system, introduced a simple but efficient and effective automated speaker recognition system. This was done by first performing a comprehensive analysis of the MFCC strategy for recognition with the Time Domain approach by modelling both of these strategies using MATLAB 7.0 and analyzing the uniformity of recognition using several methods. This is accompanied by the time domain method being applied on the TMS 320 C6713 DSK. The biometric system which is voice-based is premised on the recognition of insignificant or single words. Once during the training session, a specific speaker will utter the password to train and store the entry word features. The client utters the password again later in the testing session in order to obtain acknowledgement if there is a win. During the training phase, the feature vectors peculiar to that speaker are collected and this is afterward used to give authentication to the same speaker who utters the same word again during the testing stage. At this point, an attacker also may check the device by invoking the same word to check the underlying security feature.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

III. METHODOLOGY

Object-Oriented Analysis and Development Methodology was used. We obtained data sets for both normal and abnormal person and use it to plot their audio waveforms for comparison. Python programming language was used for the implementation and serious difference was observed which can be used for detection of substance abuse.

IV. EXPERIMENTAL RESULTS

Figure 1 and figure 2 shows the visual illustrations of the sample outputs for both a normal person and abnormal person.

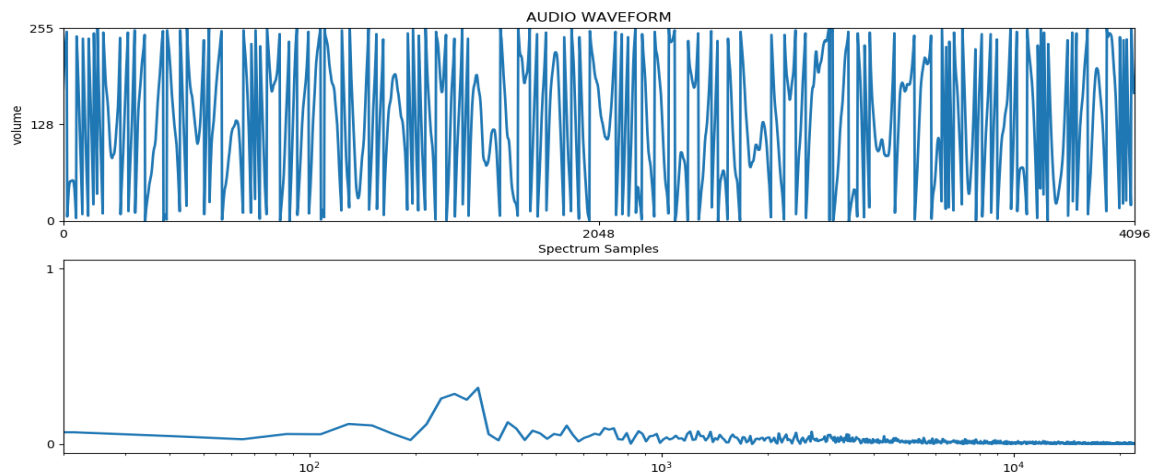


Fig. 1. Sample of a normal situation

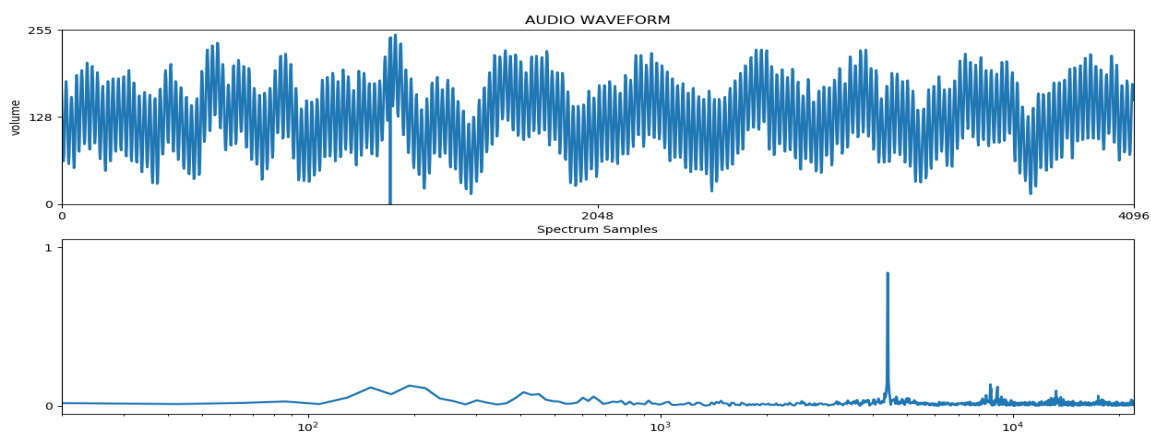


Fig. 2. Sample of an abnormal situation

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

V. CONCLUSION

Speech/Voice Recognition Systems (SRS) are growing day by day and has unlimited applications. Different application systems were created using dictation technology and spoken dialog technology based on their achievements. This research paper examined how to attribute drug abuse to mental health problem as well as how to identify and diagnose serious mental condition by using speech / voice recognition. At the end of the day, an intelligent system was designed which will help in detecting and distinguishing a normal voice or speech from an abnormal voice / speech.

REFERENCES

- [1] Tuchman, E., "Women and addiction": the importance of gender issues insubstance abuse research, *J Addict Dis*, Vol. 29, pp. 127–38, 2010.
- [2] Whitehorne-Smith, P., Morgan, K., De La Haye, W., and Abel, W. D., "Genderdifferences in psychosocial risk and protective factors for adolescent alcohol use and misuse in Jamaica", *WIMJ Open*, Vol. 2, pp. 18–21, 2015.
- [3] Lawrence, T., "Key Differences between Speech Recognition and Voice Recognition", Retrieved on the 2nd of October, 2019 from: https://www.streetdirectory.com/travel_guide/139545/technology/key_differences_between_speech_recognition_and_voice_recognition.html
- [4] Sadaoki, F., "Automatic Speech Recognition and Its Application to Information Extraction", Department of Computer Science Tokyo institute of Technology, Japan, 2015.
- [5] Nasir, U., Rashid, M., Mostafa, G., Belayet, H., Salam, S., Nithe, N., Rahman, W., and Halder, S., "Development of Voice Recognition for Student Attendance", *Global Journal of Human-Social Science: G Linguistics & Education*, Vol. 16, No. 1, 2016.
- [6] Heng, Y., "Development of Speech Recognition System for Forensics Application", A Report Submitted in Partial Fulfilment of Requirements for the Award of Bachelor Degree of Electronic Engineering (Wireless Communication) Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia, 2015.
- [7] Levelt, W. J., "Models of word production", *Trends in Cognitive Sciences*, Vol. 3, No. 6, pp. 223–232, 1999.
- [8] Garnham, A., Shillcock R. C., Brown, G. D., Mill, A. I., and Culter, A., "Slips of the tongue in the London–Lund corpus of spontaneous conversation", *Linguistics*, Vol. 19, pp. 7–8, 1981.
- [9] Oldfield, R. C. and Wingfield, A., "Response latencies in naming objects", *Quarterly Journal of Experimental Psychology*, Vol. 17, No. 4, pp. 273–281, 1965.
- [10] Bird, H., Franklin, S., and Howard, D., (2001). "Age of acquisition and imageability ratings for a large set of words, including verbs and function words", *Behavior Research Methods, Instruments, and Computers*, Vol. 33, No. 1, pp. 73–9, 2001.
- [11] Shinichi, T., "Stress Evaluation by Voice: From Prevention to Treatment in Mental Health Care", *Global Journal of Artificial Intelligence*, Vol. 16, No. 1, pp. 224 – 229, 2015.
- [12] Adarsh, K., Deepak, A., Diwakar, R., and Karthik, R., "Implementation of a Voice – based biometric system.", A project work submitted to the Department of Electronics and Communication Engineering, Visveswaraya Technological University, Belgaum, 2007.