Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 13, Number 9 (2017), pp. 5437-5445

http://www.ripublication.com

Aggression Monitoring In Speech Using Semantics and Pitch

S.J.Priscilla[1] and M.Vanithalakshmi[2]

[1]PG student, S.A. Engineering College, Poonamalle-Avadi Road,Veeraraghavapuram, Thiruverkadu, Avadi, Chennai, Tamil Nadu, India.

[2]Assistant Professor, S.A. Engineering College, Poonamalle-Avadi Road,Veeraraghavapuram, Thiruverkadu, Avadi, India.

Chennai, Tamil Nadu, India-600077.

Abstract

The development of Technology in digital era speaks to a modern accomplishment in transformative history of IT segments. Emotions and stress assume a vital part in the improvement of undesirable conduct. Stress happens amid the any movement or execution like while correspondence, play undertaking, dramatization, organize action or while in examination like oral or composed is called as Laboratory stress. Detecting negative emotions and stress at an early stage can prevent aggression. Due to increased difficulty in automation, predefined datasets were used for recognition in the earlier approaches. In this paper, we proposed an acoustic Model based for monitoring human stress using ARM 11 processor. Low level and high level stress points are detected by means of total harmonic distortion level. Acoustic features capture normal communication speech of human. The real time voice input is recorded and based on the pitch and intensity level of the speech, the state of the human is monitored.

Keywords: Emotion, stress, acoustic model, ARM11 processor.

I. INTRODUCTION

In searching for the transformative underlying foundations of human discourse, numerous analysts swung to the vocal signs of nonhuman primates [1] [2] rather than a "gestural birthplaces" perspective of how dialect may have developed. Notwithstanding, youngsters utilize signals for correspondence before their first talked words, and grown-up speakers ordinarily go with all their discourse with expressive manual motions [3], while human marked dialects are all out dialects that don't utilize discourse. In this manner, any hypothesis of dialect inceptions must address the way that motions shape a critical piece of the human "dialect execution framework." Hewes (1973) [4] contended that our progenitors could intentionally control signals well before discourse developed. Corballis [5] recommended that manual motions prepared for the development of handedness connected to cerebral lateralization and—misusing the "generativity" of manual activity—for the advancement of human dialect. Armstrong et.al [6] (2007) bolster an essential part for notorious signals in dialect development and recommend that marked dialects are the first and prototypical dialects.

Chronic examples of thought which impact evaluation and improve the probability that a man will encounter worry as pessimistic, (for example, low self-adequacy, or a conviction that you are unequipped for overseeing stress) can likewise improve the probability that a man will wind up plainly discouraged. Side effects of Major Depression may include: rest issues; exhaustion; craving changes; sentiments of uselessness, self-loathing, and blame; a failure to think or decide; tumult, anxiety, and peevishness; withdrawal from run of the mill pleasurable exercises; and sentiments of sadness and powerlessness. Despondency is likewise connected with an expansion in self-destructive deduction and self-destructive activities, and may make a man more helpless against creating other mental issue.

Full of feeling excitement tweaks all human open signs. Clinicians and language specialists have different assessments about the significance of various prompts (sound and visual signals in this paper) in human influence judgment. Ekman [23] found that the relative commitments of outward appearance, discourse, and body signals to influence judgment depend both on the full of feeling state and the earth where the emotional conduct happens. Then again, a few reviews [24] and [25] showed that an outward appearance in the visual channel is the most critical full of feeling sign and corresponds well with the body and voice. Many reviews have hypothetically and observationally exhibited the benefit of the joining of various modalities (vocal and visual expression) in human influence discernment over single modalities [24], [26]. Not quite the same as the conventional message judgment, in which the point is to gather what underlies a showed conduct, for example, influence or identity, another real way to deal with human conduct estimation is the sign judgment [27]. The point of sign judgment is to portray the appearance, as opposed to

the significance, of the indicated conduct, for example, facial flag, body motion, or discourse rate. While message judgment is centred on translation, sign judgment endeavours to be a goal depiction, leaving the derivation about the passed on message to abnormal state basic leadership.

Computer vision is the contour points based stress identification. How many contour points are detected in the camera region, hence low contours are detected in the camera region this is low level stress and high contour points detected means high level stress. Acoustic model is some of voice modulation data's are stored in database, that voice data's are stored in various frequencies, so compared to that voice data's, so known the which state of stress in human. Finally known the human stress are compared to the acoustic model and computer vision data's. And send the stress people information to the service desk.

"Speech processing" is the investigation of discourse signs and the preparing strategies for the signs. The signs are typically prepared in an advanced portrayal, so discourse handling can be viewed as a unique instance of computerized flag handling, connected to discourse flag. Parts of discourse handling incorporates the procurement, control, stockpiling, exchange and yield of discourse signs. The information is called discourse acknowledgment and the yield is called discourse blend.

"Speech recognition(SR)" is the between disciplinary sub-field of computational semantics which joins learning and research in the phonetics, software engineering, and electrical designing fields to create philosophies and innovations that empowers the acknowledgment and interpretation of talked dialect into content by PCs and mechanized gadgets, for example, those sorted as brilliant advances and mechanical technology. It is otherwise called "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT).

"Speech synthesis" is the fake generation of human discourse. A PC framework utilized for this design is known as a discourse PC or discourse synthesizer, and can be actualized in programming or equipment items. A "text-to-speech" (TTS) framework changes over typical dialect content into discourse; different frameworks render typical etymological portrayals like phonetic interpretations into discourse. Frameworks contrast in the span of the put away discourse units; a framework that stores telephones or di-telephones gives the biggest yield go, however may need lucidity. For particular utilization spaces, the capacity of whole words or sentences takes into account great yield. On the other hand, a synthesizer can fuse a model of the vocal tract and other human voice attributes to make a totally "manufactured" voice yield.

II. REVIEW OF WORK

Ashish Panat et.al [7] concentrated the feelings and the examples of EEG signs of human mind for utility in the finding of psychosomatic issue in more basic and temperate route with the assistance of ECG flag in their Analysis of feeling issue in light of EEG signs of Human Brain. Z. Khalili et al. [8] [9] and Jerritta Selvaraj et.al [10] have chipped away at Emotion discovery utilizing EEG. Prashant Lahane et.al [11] additionally gave their commitment in Emotion identification utilizing EEG and fabricate a productive and solid feeling acknowledgment framework. Adrian et.al [12] broke down the feeling utilizing EEG signs, to screen elite competitors. In their work comes about just give data about the cozy connection between cerebrum exercises created by two feelings. Absence of benchmarks was one of the constraints in the advancement. Simina Emerich and et.al [13] built up a bimodal feeling acknowledgment framework utilizing the mix of outward appearances and discourse signals with various classifiers. They utilizes a SVM (Support Vector Machine) Naive Bayes and K-Nearest Neighbor for their examination work. Ritu D.Shah et.al [14] and Jerritta Selvaraj et.al [10] gave their commitment in Emotion recognition utilizing SVM. They demonstrated that Both RRS and FVS strategies indicated comparable arrangement precision and blend of non-straight investigation and HOS tend to catch the better enthusiastic changes. Hurst type examines the smoothness of a period arrangement and depends on self-comparability and relationship properties. It likewise assesses the nearness or nonappearance of long-range reliance and its degree on a period arrangement [10] [15] [16].

Arbib [17] contends that a capacity for complex impersonation one of a kind to the human line made conceivable the advancement of cerebrum components for emulate and subsequently proto sign, an arrangement of traditional motions used to formalize, disambiguate, and expand emulate. It was additionally estimated that, once proto sign has set up a capacity for the free making of subjective signals to bolster an open finished semantics, the ability to utilize conventionalized manual informative motions (proto sign) and the ability to utilize vocal open motions (proto discourse) developed together in a growing winding [18] to bolster proto dialect [19] [20], an open-finished multimodal open framework. Nonetheless, the correspondence frameworks of nonhuman primates need compositionality, a pivotal property of present day human dialects. This is the thought that dialect gets its energy from having an open-finished vocabulary as well as from having a sentence structure that permits words to be consolidated into expressions, with the outcomes open to further blend, additionally empowers the listener to derive the significance of the general articulation from the importance of its parts and the developments used to amass them.

The utilization of manual and real signals to speak with different conspecifics has been accounted for a few types of nonhuman primates. Great reviews incorporate those of [21] [22], who gave point by point depictions of various motions

(notwithstanding other open practices) utilized by monkeys and gorillas. Later reviews concentrate on the individual fluctuation of gestural collections and the psychological instruments basic gestural correspondence. We next concentrate on primate signals and demonstrate that (1) utilization of open motions is normal crosswise over species, (2) there is impressive fluctuation in motion collections from gathering to gathering, and (3) motions are utilized flexibly in various settings, with utilize contingent upon the conduct of the beneficiary. This flexibility appears to be inferable from learning. We will think about reviews on gestural correspondence of gorillas both in imprisonment and in the wild, including every single awesome primate and siamangs (as illustrative of the little chimps or gibbons). We view practices as signals just on the off chance that they serve to achieve a repetitive social objective and are coordinated at a specific beneficiary. Manual and real motions can be bunched into three flag classifications—sound-related, material, and visual—contingent upon the perceptual framework used to get them. Sound-related signals create sound (yet not with vocal ropes) while material motions incorporate physical contact with the beneficiary and visual motions produce a predominantly visual impact with no physical contact.

III. PROPOSED SYSTEM ARCHITECTURE

ARM11 processor is interface to the microphone. Stress is mainly detected by using how human interpret through words. Microphone is used to record the human voice data's. ARM11 processor is comparing the voice behavior data with the threshold level of harmonic distortion set by the system, so as to tell which state of stress is in human.

ARM 11 is considered as a mini Personal computer more of like a Personal Digital Assistant.ARM11 BCM2836 is supplied with power supply of 5V. ARM 11 has GCC compiler used for compilation of data voice processed. ARM 11 consists of integrated Raspberry Pi 3 kit with general usage input and Output ports. ARM 11 Kit has a IC compiler and four USB ports, Display port, External memory slot, Audio Output Slot and Power Supply pin.

ARM 11 is prebuilt with the required library storage files for audio wave processing and signal analyzing. Linux OS is flashed into the external memory drive disk port. The figure 3.1 shows the block diagram of our proposed system.

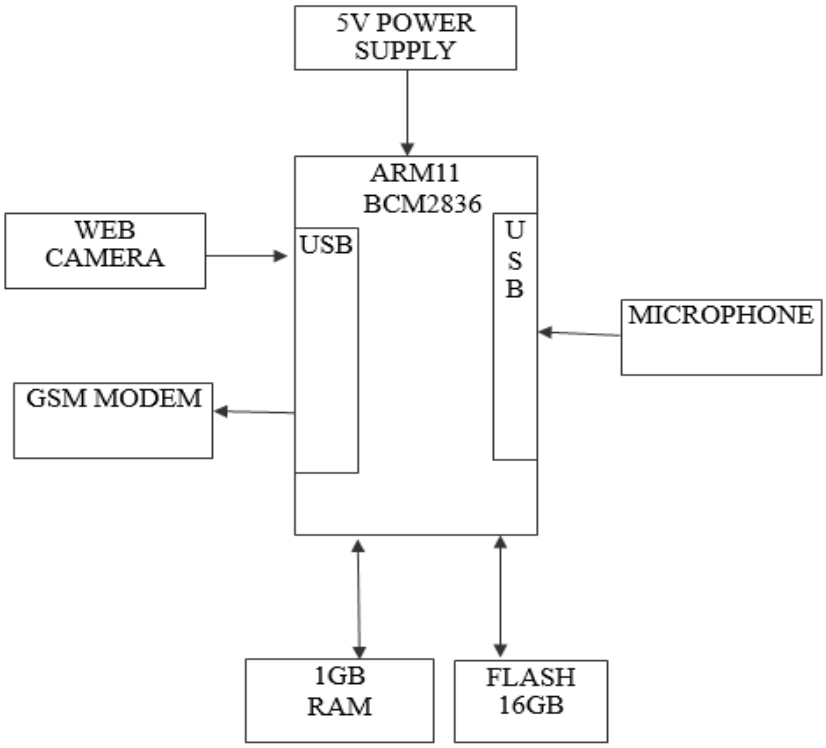

Figure 1: Block Diagram of the Proposed System

The Voice input is recorded and stored in the system. The ARM 11 processor is used for comparing the threshold harmonic distorted wave levels with the real time voice data input from the human.

Figure 3: Screenshot of the Kit

IV. EXPERIMENTAL RESULT AND FUTURE SCOPE

The proposed work is simulated using Linux Operating System for detection and monitoring of the Aggression level in Human.

The waveform analyzer master audio file stored is processed using Python. The library files are stored in the external RAM to get and record the audio voice. The Recorded voice file is stored in ".cpp" extension and buffer time for recording is primarily set to "5 seconds". Audio waveform analyzes the speech signal. The Total Harmonic Distortion level for normal voice data is set to 95. The Voice input for above the threshold level is stated as "Relatively Aggressive" with the corresponding frequency level of the speech.

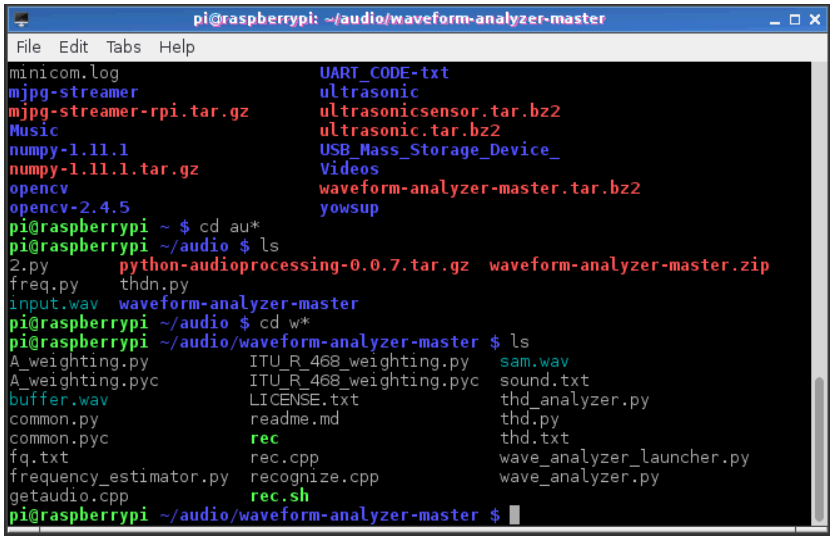

Figure 3: Audio Execution

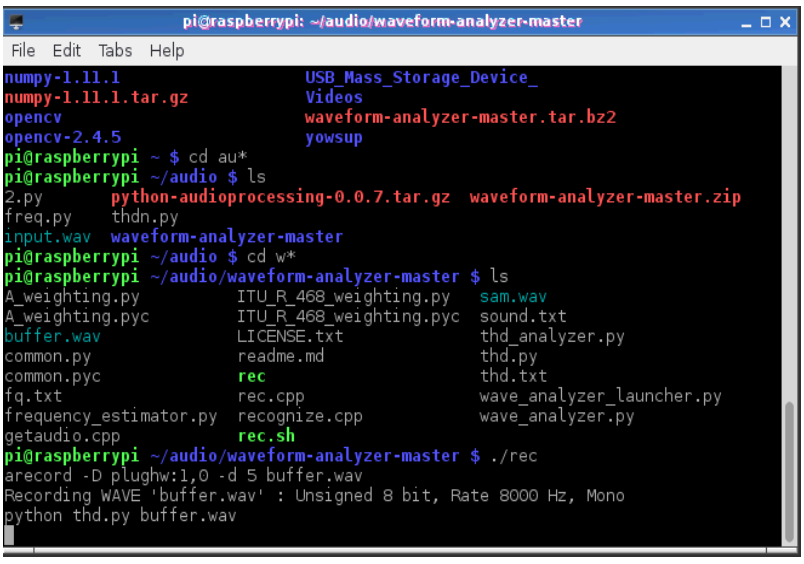

Figure 3: Recording of the Voice Input

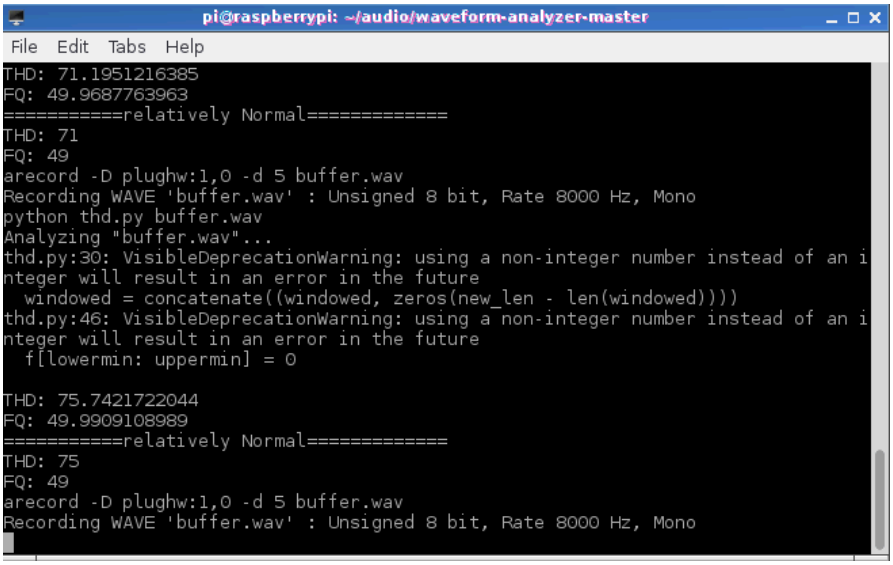

Figure 3: Aggression monitored Data

V. CONCLUSION

This paper thus includes a module to detect the aggression detected by means of predefined fundamental frequency level for stress monitoring in human. The detection of high level stress is by voice data's are stored in various frequencies, that is compared to the total harmonic distortion level so known the state of stress in human.

FUTURE WORK

One of the fundamental difficulties is the multifaceted nature of human conduct and the expansive fluctuation of signs which ought to be contemplated. Feelings and stress assume a critical part in the advancement of undesirable conduct. Our future work includes a module to detect the aggression detected in the camera region, detected means by high level stress that voice data's are stored in various frequencies, so compared to that image captured by the Web camera. Finally known the human stresses are compared to the acoustic model and computer vision data's. And send the stress people information to the service desk. We use computer vision based to identify the human stress using gesture and speech processing.

REFERENCES:

[1]. Seyfarth, R. M. 1987. Vocal communication and its relation to language. In Primate societies, ed. B. Smuts, D. L. Cheney, R. Seyfarth, R. Wrangham, and T. Struhsaker, 440–51. Chicago: University of Chicago Press.

[2]. Snowdon, C. T., C. H. Brown, and M. R. Petersen. 1982. Primate

communication. Cambridge: Cambridge University Press.

[3]. McNeill, D. 1992. Hand and mind. Chicago: University of Chicago Press.

[4]. Hewes, G. W. 1973. Primate communication and the gestural origin of language. Current Anthropology Vol. 12, pp. 5–24.

[5]. Corballis, M. C. 1991. The lopsided ape: Evolution of the generative mind. New York: Oxford University Press.

[6]. Armstrong, D. F., and S. E. Wilcox. 2007. The gestural origin of language. Oxford: Oxford University Press.

[7]. Ashish Panat and Anita Patil „Analysis of emotion disorders based on EEG signals of Human Brain‟ International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.4, August 2012.

[8]. Prashant Lahane, Shrutika Lokannavar, Apurva Gangurde, Poonam Bhosale, Pooja Chidre, "EEG Based Emotion Recognition System", International Journal of Computer Science and Information Technologies, Vol. 5, Issue. 6, 2014, pp.7656-7658.

[9]. Adrian R. Aguiñaga, Miguel Lopez Ramirez, Arnulfo Alanis Garza, Rosario Baltazar, Víctor M. Zamudio, "Emotional analysis thru EEG signals, to monitor high performance athletes", The Journal of Innovation Impact: ISSN 2051-6002, Special Edition on Innovation in Medicine and Healthcare : Vol. 6. No. 1, pp.16-23.

[10]. Z. Khalili1, M. H. Moradi, "Emotion detection using brain and peripheral signals", Proceedings of the CIBEC'08 978-1-4244-2695-9/08/$25.00 ©2008 IEEE.