MAT JOURNALS
Journal of Computer Science Engineering and Software Testing
e-ISSN: 2581-6969
Volume 5 Issue 2

Analysis of Speech in Human Communication

Basavaraj N Hiremath[1*], Malini M Patil[2]
[1]Research Scholar, [2]Associate Professor
[1]Department of Computer Science and Engineering,
[2]Department of Information Science and Engineering,
[1,2]J.S.S Academy of Technical Education, Bengaluru, Karnataka, India
***Email:** *basavaraj@ieee.org*
***DOI:** http://doi.org/10.5281/zenodo.3250518*

Abstract

The human communication has a vitalmode called speech which results from the voice along with knowledge of language. The voice of each person is distinct because individual-specific vocal cord anatomy, vocal cavity, and oral and nasal cavities. It forms a basic block for copious knowledge into various analysis like lexical analytics, natural language processing, text mining, sentiment and satire. Apart from linguistics analysis, the physics of voice contributes to uniquely cognize as a signal. The paper aims at understanding a computer program called 'PRAAT' to analyze and synthesize a phonetics by computer. The work carried out in the paper focuses on presenting the comparative analysis real time voice data sample and the benchmark voice data in praat for all the parameters of the voice analysis. The results are promising and make a way to build decision making solutions to patterns of voice, recognition and reproduction processes using the facts of analytics.

***Keywords:** Digital assistants, phonetics, speech recognition, voice analytics*

INTRODUCTION

Voice plays a key role in human communication, voice at its root has its physics means the production of voice through glottal space starts from passing of air with articulation. The speech forms by the base knowledge of language [5]. So, the glottal air pressure creates lot of dependency on the further flow of speech as an understandable language. Nowadays, there lies intelligent assistants in all modes to ease the burden of repetitive human intervention to various devices that human kind interacts to interact for actions, for example speech assistants like Siri, Alexa made their ample usage in day to day business especially in the domain of customer relationship management [3]. To process the speech by machine significant resources like transforming speech into text by devices with best performance paves the way to text analytics for decision making [2]. The analysis and study of semantics and syntax of a specific language has led to other proficient subjects called as natural language processing, if it happens on linguistics.

The text processing by considering linguistics alone led to understand emotions thoughts and views of people's communication and their behaviour, this analysis revolves around language dictionary. The other significant analysis bases on the characteristic features of voice i.e. roots of speech. In recognition [8] of specific speech pattern and glottal signature i.e. excitation of signal produced by glottis [9]. The study of voice as a signal accesses the spectrum of sound.

Praat is a computer program comes with public license, speech manipulation, various graphical representation of speech and its synthesis, the usability is for various domain interpretation and

recognition of sound. It was founded by the team in university of Amsterdam [1]. Recently, a package called parselmouth, a python opensource interface has been facilitated to use praat functionalities in visualization, sound manipulation and acoustic data analysis making the tool more versatile for coding in integration with python programming language [7]. This paper summarizes the significant parameters at the phase of spectrum of voice to underline the need for presenting relevance experiment. The paper is structured with, in section II a literature survey of relevant significant papers are reviewed, then in section III details of method on experiment is briefed with details used for carrying out experiment with software called 'praat' along with description of datasets and section IV with results and conclusion with description on significance of future work.

LITERATURE SURVEY

The speech in human communication is basically complex process built in stimuli of language and establishes relationships in complexity. [11] Speech discrimination and deviations in relationships of pitch has more significance in sound pathology like study auditory neuron system. The authors are conducting various experiments in praat programming software to study context of detection of sarcasm specially the role of prosody i.e. elements of speech to understand patterns of stress in language utterance for all acoustical analysis [12]. The authors [13] exclusively used features extracted from praat scripts to acquire, frequency, pulses, jitter, shimmer, pulses and harmonicity by using classification of feature sets, an emotion detection system is developed by support vector machines to detect seven emotion features sadness, anger, neutral, disgust, happiness, anxiety and panic. Pitch points are exclusively extracted by authors [14] to analyze the pitch parameters to extract emotions of happy and neutral and found the state of change point from neutral to happy emotion. It is significantly noted that the signals originated from happy emotion has a 'high' pitch in (Hz), whereas the 'less' in signals originated from neutral signals. Later its identified that pitch points manipulated output signals can create emotion change states. The research [15] other than text which is based on linguistic resources is voice (audio) and facial emotions (video). The recognition of sentiment and emotions on video datasets also deduce to extraction of voice cues. So the features of loudness in sound, probability and frequency in voicing and speech respectively, these are exclusively used in communicative strategy in the news media to express sentiment and proper delivery based on the sentences uttered. The authors [17] have described the slangs and natural language processing are important from the point of extracting analytics in speech. So, after deducing speech into language and voice, the study of voice makes it a significant factor in building blocks.

METHODS AND MODELS

The standard design of experiments (DOE) prompts for basic feature identification by analyzing the basic statistics of the data sample. To compare the results from two different samples for optimal process output and to understand expressive features that affect the output from identifying these features. The study of experiment revolves around understanding influence of features which affect the outcome result. Box-Behnken design is one of the standard methods that describes 3 level design, where minimum level values, maximum level values and overall mean values [18].

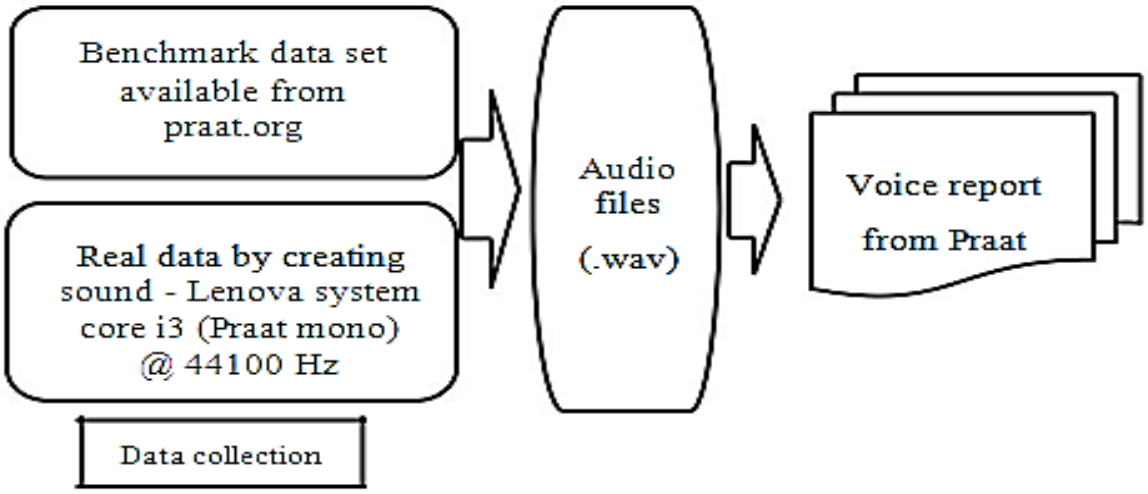

Figure 1: *Block Diagram for collection of voice features.*

The parameters to be set for the experiment is to be decided based upon the output variable. Then the significant variable in which the delta has a major impact on the outputvariable. In the controlled experiment what variable needs to be considered must be observed in the experiment. This guides the framework to identify facts related to the features and recorded for conclusion.

By observing and doing basic test on programming software called praat is used because of its characteristic feature of accepting all varieties of datasets in sound formats[16], it can be used to annotating and segmenting sound signals, praat has graphical user interface with two window driven menu, one is for object data source for input and save, another is for output in the form of text, graph or charts. Praat is a platform independent and has facilities to code and program using its own scripting language and for phonetic analysis. Praat can transform the data into graphical and numerical. The datasets for this experiment have been selected from benchmark and a real time dataset for comparison. The data flow diagram is represented in Fig. 1 in blocks for understanding.

Pitch Contour

In the present experiment the features are identified in three categories as shown in Fig. 2. In speech for natural, spontaneous and machine matching.

The observation in this experiment is to capture the pitch contour [6] which displays the maximum pitch for benchmark dataset by utterance of English word with vowel, but the other sample has a ‘noise’ so called the ‘natural cough’ in the utterance which directly affects the fluency[4] in speech. Though the cough noise is spontaneous and is not having any linguistic words for articulation has made the pitch note to go high.

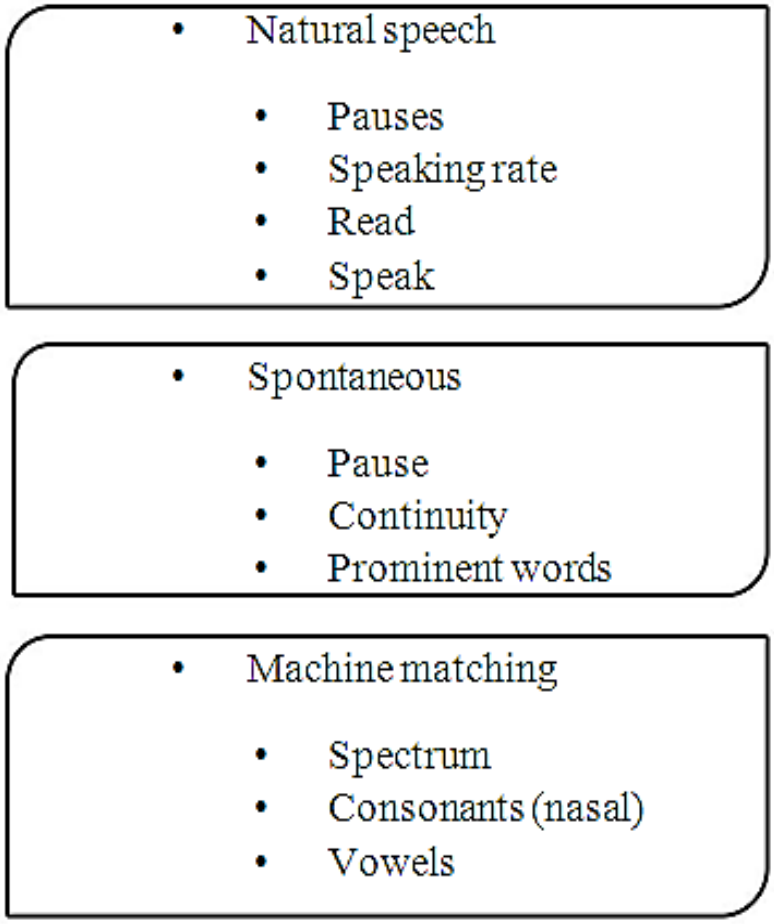

Figure 2: *Deducing voice analytics.*

Table 1 displays various parameters of voice in both the samples. The significant is the pitch measured in Hz. The benchmark dataset sample used for extraction feature relatedto pitch. The Fig. 3 and Fig. 4 are spectrum of sample 1 created by utterance of word in English as 'dampskunk' and its pitch contour respectively which highlights the spectrograph with natural word utterance. The Fig. 3 has the visible part has 1.2287 seconds and significant visible waves of signals, correspondingly in Fig. 4 two significant contour lines of pitch are observed.

Table 1: *Details about pitch and voice report.*

Sample	**Dampskunk with no cough Sample 1**	**Dampskunk Sample 2**
Maximum pitch (Hz)	160.3499	233.9472
Time range of selection	0.007694 to 1.224397 seconds	0.172790 to 1.890563 seconds
Shimmer	Shimmer (local, dB): 1.114 dB	Shimmer (local, dB): 1.050 dB

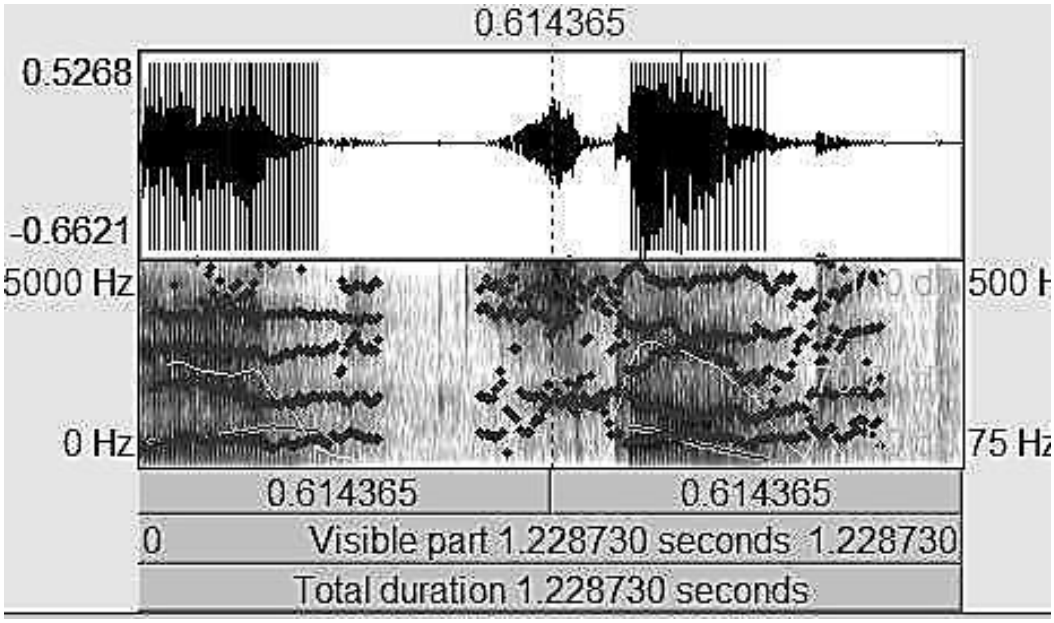

Figure 3: *Spectrum for sample 1.*

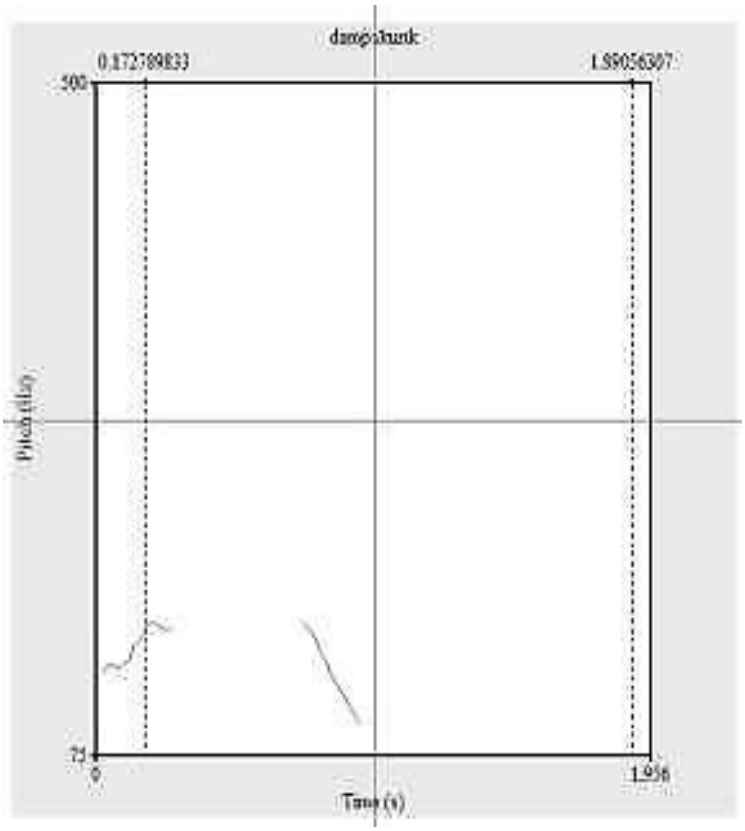

Figure 4: *Pitch contour for sample 1.*

The Fig. 5 and Fig. 6 are the spectrograph and its pitch contour for sample 2. The sample has the word utterance and sound 'dampskunk with cough' where a curved line of three numbers of distinct contours can be observed along with voice pause. The spectrograph has visible part of wave signals as 1.9563 seconds which includes 'word-pause-cough'. In Fig. 6, the third significant contour relates to pitch which has high measured more than 200 Hz basically glottal air pressure created by 'cough'.

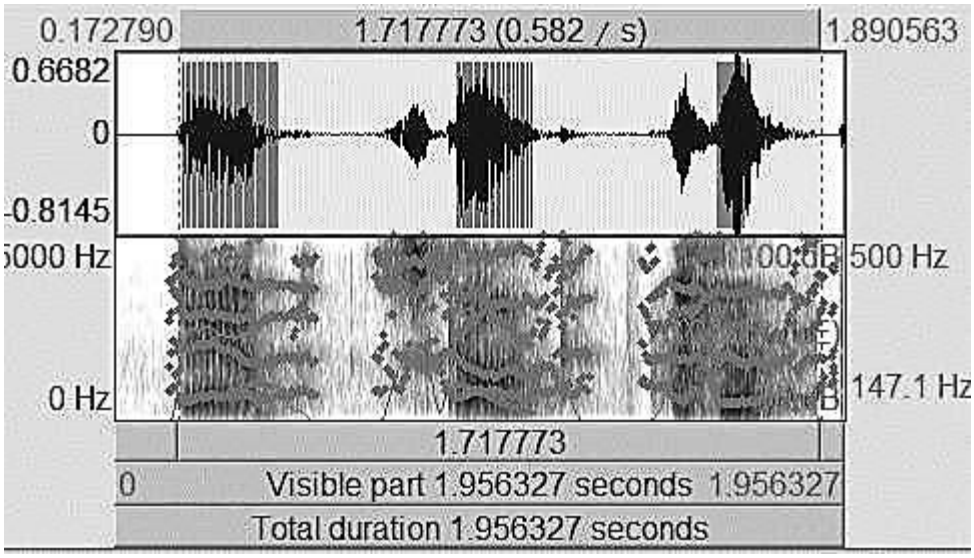

Figure 5: *Spectrum for sample 2.*

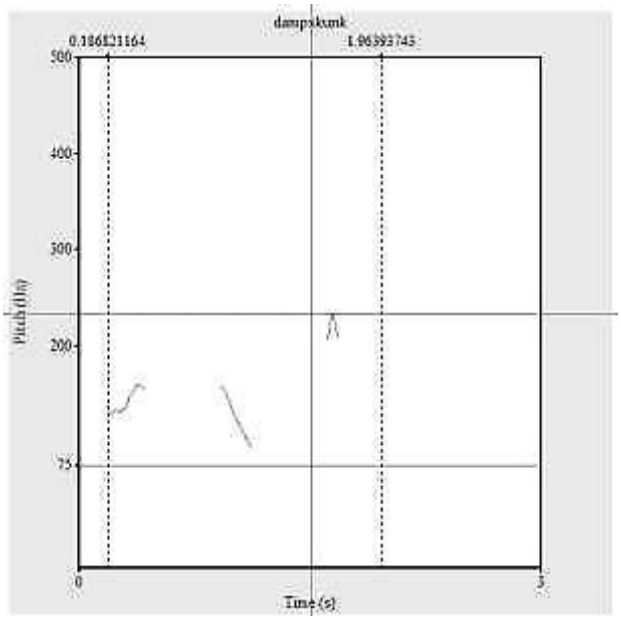

Figure 6: *Pitch contour for sample 2.*

The spectrograph basically demonstrates the following significant factors as,

- Degree of measurement of pause or breaks like voiced parts and signal breaks.
- Jitter is a measure of glottal closure which plays a very important role in measuring the glottal signature specially for
- Pathological or glottal signature in match making.
- Shimmer, a parameter for measuring sustained vowels.
- Pitch is a measure of frequency.
- Spectrograph analysis helps to manipulate the voice signal for analysis.

Read and Speak

In human communication, the impact of speakers on the listeners make to build standards as ability to speak and read. The identification of disfluency in news readers and spontaneous speakers will depend upon pauses, word fillers, interjections and repetitions. This analogy

has contributed ample knowledge in design of robotic voice and intelligent assistants to elimination of noise and recognition algorithms [7].

The data source (Recorded) used are for; 1) READ ExperimentThe recorded audio signal is in normal enclosure with data set used is 'bnh_recorded_voice_READ' the environment of creating this sample ('Google is not a conventional company. We do not intend to become one') [10] in a standard reading room with a sampling frequency of 44100 Hz in mono microphone. Fig. 7 has the observed values in the spectrograph.

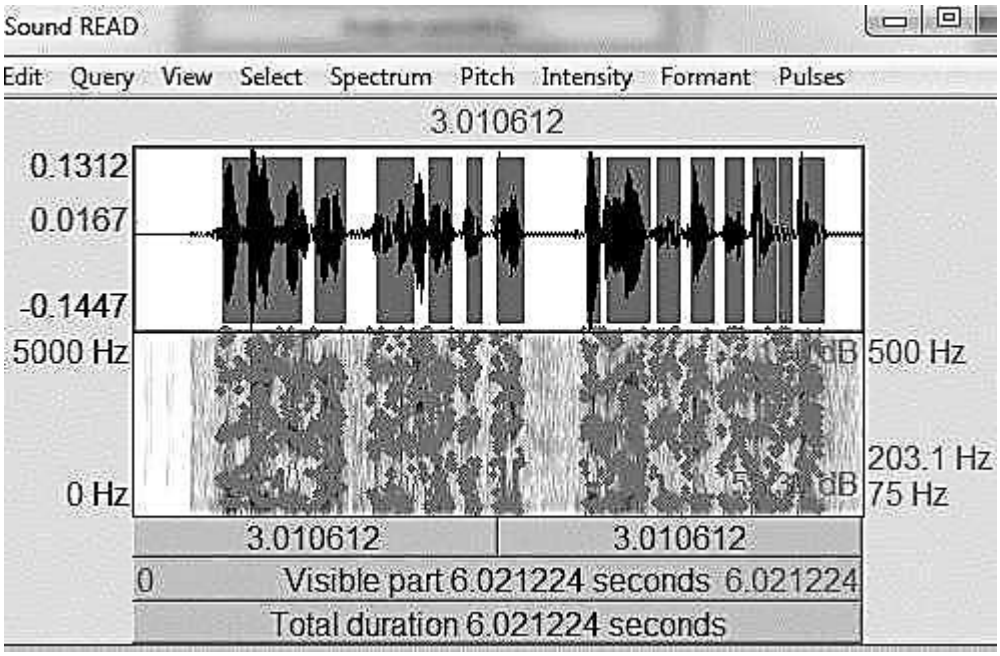

Figure 7: *Spectrograph of READ sentence.*

Speak Experiment

The recorded data set used is 'bnh_recorded_voice_SPEAK', the environment of creating this sample is a standard reading room with a sampling frequency of 44100 Hz in mono microphone. Fig. 8 has the observed values listed in the graph.

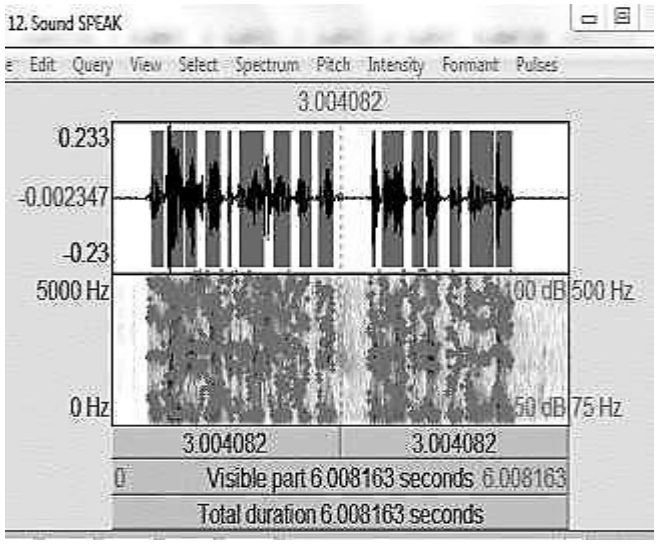

Figure 8: *Spectrograph of SPEAK sentence.*

Current method is to create constant settings in processing input voice signal bywav file by keeping frequency at standard settings. These observations have a direct impact on the consciousness and confidence level of utterance of sentence in each test.

CONCLUSION AND FUTURE WORK

The basic experiments have given us an understanding of voice cues without relating to linguistic analysis in identifying the analytics. As the speech has inclination of language and their specific variables play a vital role, but the glottal signature [9] which identifies the unique feature of human voice in arriving at the sentimental and emotion factors where we can think of accuracy and performance at a higher rate. Currently research is on to design a voice fluency [8] a modern intelligent assistant must speak.

Future work on the similar experiments can extended to for match making of speech pattern and recognition of speech by using dictionary resource. For validation of test, all the features can be done by using analytics tools like IBM's SPSS and SAS for better categorization and drill down, various detailed attributes to arrive at comparison by percentage optimization. By these experiments 'glottal signature' a nasal stress behaviour is analysed to go with parameter analysis for voice utterance.

ACKNOWLEDGEMENT

The authors wish to thank JSS Academy of Technical Education Bengaluru for having for having given the opportunity to conduct the present work. We thank the team of All India institute of speech and Hearing Mysore for having given resourceful information knowledge about relevant software and publication materials for conducting research work.

REFERENCES

1. P. Boersma, D. Weenink (2013), "Praat, a system for doing phonetics by computer",*Glot Int.*,Volume 5, Issue 9–10, pp. 341–345.
2. B. N. Hiremath, M. M. Patil (2017), "A Comprehensive Study of Text Analytics",*Int. J. Artif. Intell. Syst. Mach. Learn.*,Volume 9, Issue 4, pp. 70–76.
3. Z. S. C. (India) Basavaraj N. Hiremath, Solution Architect, K. Private limited, Bengaluru, Malini M. Patil, Associate Prof., Dept. of Info. Sc. and Eng., J.S.S. Academy of Technical Education, and K. Bangalore (2016), "CustomerRelationship Management Through Natural LanguageProcessing Using Text Analytics",*CSI Commun.*,Volume 40, Issue 1, pp. 11–13.
4. V. van Heuven, P. Boersma (2001), "Speak and unSpeak with PRAAT",*Glot International*,Volume 5, Issue 9–10, pp. 341–347.
5. V. L., Y. V. G. Liji Antony (2015), "Comparison of FluencyCharacteristics in News-readers and Controls",*JAIISH*, Volume 34, pp. 48–54.
6. Dr. Will Styler (2017), "Praat: doing Phonetics by Computer", pp. 85.
7. Y. Jadoul, B. Thompson, B. de Boer (2018), "Introducing Parselmouth: A Python interface to Praat",*J. Phon.*,Volume 71, pp. 1–15.
8. C. Weller, "http://www.businessinsider.in/IBM-speech-recognition-is-on-the-verge-of-super-human-accuracy/articleshow/57562260.cms." accessed on20032019
9. K. Ramesh, S. R. M. Prasanna, R. K. Das (2014),"Significance of glottal activity detection and glottal signature for text dependent speaker verification",*Int.Conf. Signal Process. Commun. SPCOM 2014*

10. Larry page, Sergey Brin,"An Owner's Manual" for Google's Shareholders" https://abc.xyz/investor/founders-letters/2004-ipo-letter/ accessed on 20032019.
11. P. Virtala, E. Partanen, M. Tervaniemi, T. Kujala (2018),"Neural discrimination of speech sound changes in a variable context occurs irrespective of attention and explicit awareness," Biol. Psychol., Volume 132, Issue October 2017, pp.217–227.
12. T. Matsui et al. (2016), "The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence," *Neuropsychologia*,
13. D. Tomar, D. Ojha, S. Agarwal (2014), "An Emotion Detection System Based on Multi Least Squares Twin Support Vector Machine",*Adv. Artif. Intell.*,Volume 2014, pp. 1–11.
14. M. S. Suri, D. Setia, A. Jain (2010), "PRAAT Implementation for Prosody Conversion", pp. 1–4.
15. M. H. R. Pereira, F. L. C. Pádua, A. C. M. Pereira, F. Benevenuto, D. H. Dalip (2016), "Fusing Audio, Textual andVisual Features for Sentiment Analysis of News Videos," Issue 2015.
16. J. Harrington, S. Cassidy (2003), "Building an interface between EMU and Praat: a modular approach to speech database analysis", *Phonetic Sci.*, pp. 355–358.
17. D'Andrea, F. Ferri, P. Grifoni, T. Guzzo (2015), "Approaches, tools and applications for sentiment analysis implementation",*Int. J. Comput. Appl.*,Volume 125, Issue 3, pp. 26–33.
18. S. Jun, J. Irudayaraj, A. Demirci, and D. Geiser (2003), "Pulsed UV-light treatment of corn meal for inactivation of Aspergillusniger spores", Int. J. Food Sci. Technol., Volume 38, Issue 8, pp. 883–888.

Basavaraj N Hiremath is qualified as post graduate in Computer Cognition Technology, currently a research scholar in the department of computer science and engineering J S S Academy of Technical Education Bengaluru, his areas of interests are Artificial Intelligence, Data warehouse, Business Intelligence and Analytics in domains of Airlines, Retail, Logistics and FMCG, contact email: basavaraj@ieee.org

Dr. Malini M. Patil is presently working as Associate Professor in the Department of Information Science and Engineering at J.S.S. Academy of Technical Education, Bangalore, Karnataka, India. Her research interests are big data analytics, bioinformatics, cloud computing. She has published her research papers in many reputed international journals. Contact email: patilmalini31@gmail.com

Cite this article as: *Basavaraj N Hiremath, & Malini M Patil. (2019). Analysis of Speech in Human Communication. Journal of Computer Science Engineering and Software Testing, 5(2), 8–16. http://doi.org/10.5281/zenodo.3250518*