

# International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452  
Maths 2019; 4(6): 133-136  
© 2019 Stats & Maths  
www.mathsjournal.com  
Received: 12-09-2019  
Accepted: 16-10-2019

**Sudarsan VS**  
Machine Learning Intern  
Seaport AI Chennai, Tamil  
Nadu, India

**Govind Kumar**  
Director Seaport AI Chennai,  
Tamil Nadu, India

**Corresponding Author:**  
**Sudarsan VS**  
Machine Learning Intern  
Seaport AI Chennai, Tamil  
Nadu, India

## Voice call analytics using natural language processing

**Sudarsan VS and Govind Kumar**

### Abstract

The main challenge of call centres based on voice call is the quality monitoring, where a person has to listen to all the recorded audio files or a random sample of audio files, to check how the call centre representative has performed. It also helps in recording the feedback from the customer which is very useful in various fields of business analytics like marketing, sales, service etc. This process involves enormous amount of human effort and the time consumed is very high. With the rise of machine learning applications in almost every industry, in this paper we have tried to develop a software with the use of Machine Learning and Natural Language Processing, which helps in performance analytics of voice call agents.

**Keywords:** Voice call analytics, machine learning, natural language processing, artificial intelligence, deep learning

### 1. Introduction

Modern organisations are increasingly deploying voice bots to streamline various workplace processes. Chances are that if you dial your bank's helpline, you may be prompted by a pleasing formal sounding voice to input your request. That voice may be a chatbot's, designed to handhold you in reaching your desired result in the most efficient manner. The basic solution has been around for a while but thanks to advancement in AI technology, over the years it has rapidly improved in terms of the quality of interaction and the customer outcomes. There are several variations of conversational bots around us. Even our phones have chatbots in the form of Siri and Google Assistant. Amazon Alexa is another example of a verbal or voice based chatbot found commonly in households today. These bots fuel the Internet of Things (IoT) ecosystem and help users to naturally engage with all the smart devices at home. It is expected that by the year 2020, 50% of searches will be voice based. If that is where customers are heading, companies need to be on that platform as well. AI can create a paradigm shift in workplace functioning enabling by voice bots. The old UI-based legacy apps will be replaced by voice bots. There will be no need for different apps for different tasks. The communication and action will become streamlined saving organizational time and other resources. Tech giants like amazon, google, Facebook, Microsoft are giving more importance in voice-based applications, newly released flagship smart phones have speech to text features, they have released many open source API's, for transcribing speech to text, this shows that the industry 4.0 is moving towards voice based applications and voicebots. Motivated by this very fact we started this project, of all the commercially available softwares our's is one of a kind we have used *low latency speaker-independent speech recognition*, however specific users voice can be trained. We have chosen the voice call samples from [youtube.com](https://www.youtube.com) which was uploaded for educational purpose. These video files are then converted into audio format. The major challenge in development of this software is to find the best API among commercially available speech to text API's such as: api.ai

- Wit.ai
- Pocket sphinx
- Google

**2. Related work**

The first component of speech recognition is, of course, speech. Speech must be converted from physical sound to an electrical signal with a microphone, and then to digital data with an analog-to-digital converter. Once digitized, several models can be used to transcribe the audio to text. In the above mentioned models we tried google voice, pocket sphinx and wit.ai.

The Most modern speech recognition systems rely on what is known as a Hidden Markov Model (HMM). This approach works on the assumption that a speech signal, when viewed on a short enough timescale, can be reasonably approximated as a stationary process-that is, a process in which statistical properties do not change over time.. The final output of the HMM is a sequence of vectors. To decode the speech into text, groups of vectors are matched to one or more phonemes (a fundamental unit of speech). This calculation requires training, since the sound of a phoneme varies from speaker to speaker, and even varies from one utterance to another by the same speaker. A special algorithm is then applied to determine the most likely word (or words) that produce the given sequence of phonemes. One can imagine that this whole process may be computationally expensive. In many modern speech recognition systems, neural networks are used to simplify the speech signal using techniques for feature transformation and dimensionality reduction before HMM recognition.

We used different techniques for different API's, while using wit.ai we have to convert large audio files to small chunks as wit.ai allows only small audio files for transcribing to avoid high latency

**3. Speech recognition**

The audio calls of the particular agent is merged into one large file programmatically in order to optimise the transcribing process and in order to reduce latency (the delay before a transfer of data begins following an instruction for its transfer) we In have to chunk the file into small chunks (10secs).

We used fuzzywuzzy module to compare the manual transcribed text file of the voice calls and the text data transcribed from different API's like sphinx, google cloud, google voice etc.

The accuracy of transcribing was tested between pocket sphinx, google cloud speech to text and google for accuracy using Harvard Open Speech repository from the internet. The accuracy of wit.ai was more compared to other packages.This work only focuses on the analysis of call centre conversations and future work will include our own automated speech recognition software.The transcribed text is stored into data sets the entire code is done using python language.

The pocket sphinx module is used for both offline and online conversion of speech to text and wit.ai is used for online conversion of speech to text hence the accuracy of wit.ai is high.

**4. Call analytics metrics**

The following metrics are widely used in the call centre industry, to perform analytics.

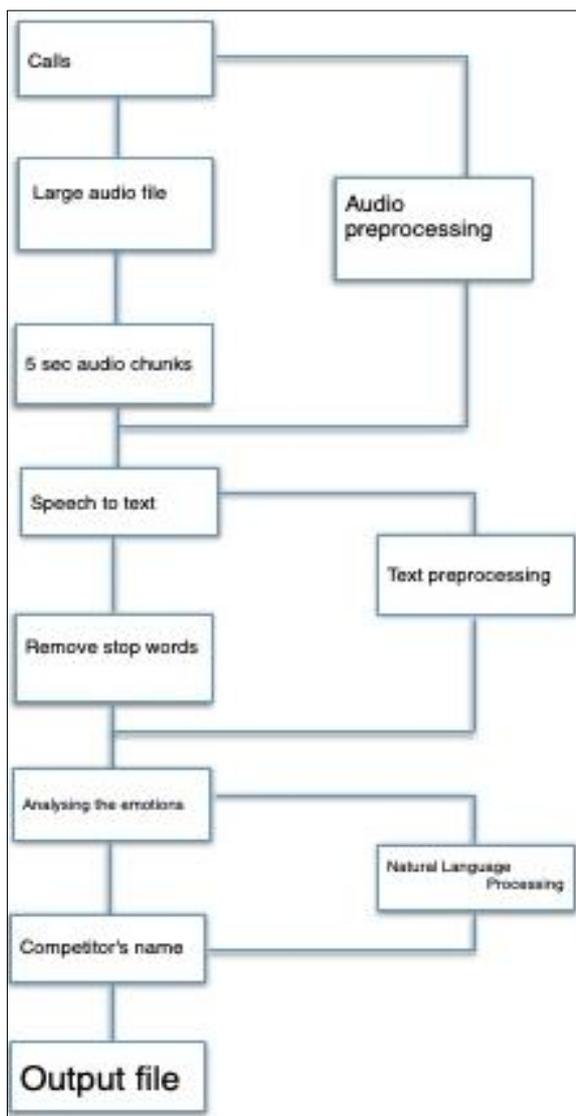


Fig 1: Flow chart for wit.ai

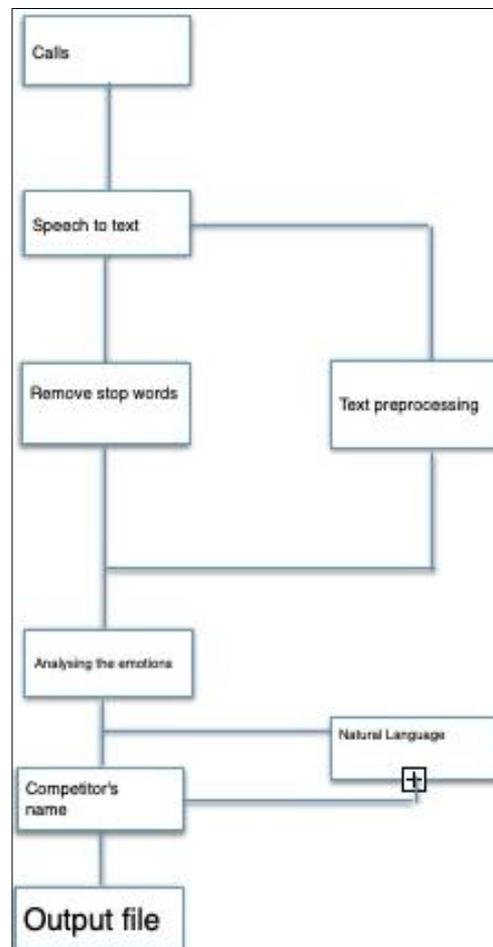


Fig 2: Flow chart for Google and sphinx

- **Customer emotions detection:** The various emotions of the customer such as whether he is satisfied or not, whether he is happy or angry with the agent is detected. This data can also be used to gather insights like review for new product launch, marketing, sales, service, etc.
- **Banned Words:** Use of banned words can be detected. Words in conversation texts have been compared with the list of banned words and checked whether agent or customer has used any banned word or not. The number of matching banned words has recorded as number of banned words in database.
- **Greeting words:** In industries like call centre, where there is direct interaction with the customer, it is very much important to for the call centre agent to greet the customers. But many times because of the hectic schedule, the representative might miss greeting the customers. Words in conversation texts have been compared with the list of greeting words and checked whether agent or customer has used any greeting word or not. The number of matching greeting words has recorded as number of greeting words in database.
- **Usage of competitor's name:** The various cases under which the competitors name is used by the customer can be recorded. This can give deep insights in marketing.
- **Performance score:** A performance score based on the above mentioned parameters is devised, to evaluate the call centre representative's performance.

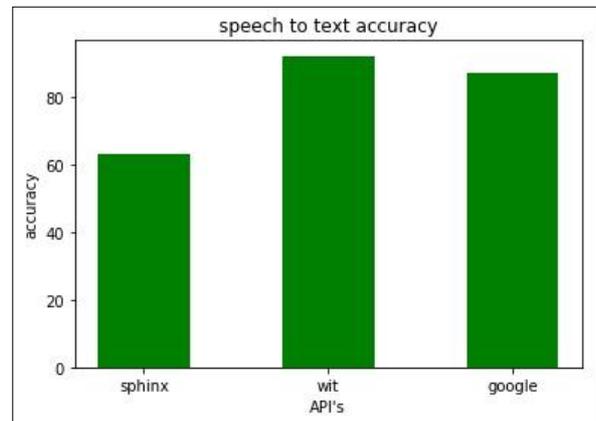


Fig 3: Speech to text accuracy

**5. Experimental results**

**5.1 Speech to text accuracy:** It can be seen that the speech to text accuracy for wit API is nearly 90% when compared to the other two API's. Google speech to text API has an Accuracy of 87%.

**5.2 Detection of emotions in a call**

**5.2.1 Accuracy of angry calls detected:** The accuracy of angry calls detected is compared with the actual numbers. The accuracy of wit and google is 100% whereas of sphinx is 37%.

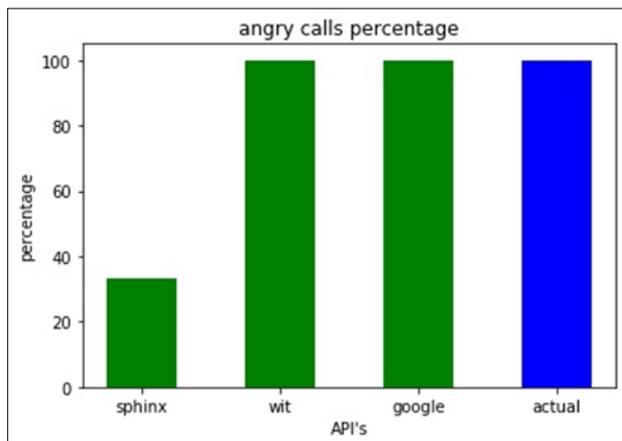


Fig 4: Angry call percentage.

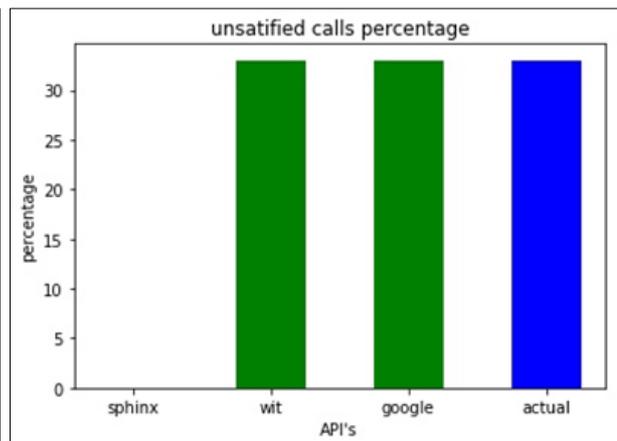


Fig 5: Unsatisfied call percentage.

**5.2.2 Accuracy of unsatisfied calls detected**

The actual percentage of unsatisfied calls is 33% which is perfectly detected by wit.ai and Google, whereas sphinx couldn't identify unsatisfied calls.

**5.2.3 Accuracy of happy calls detected:** The actual percentage of happy calls is 33% which is perfectly detected by wit.ai and google, whereas sphinx couldn't identify happy calls. *Accuracy of banned words detected:*

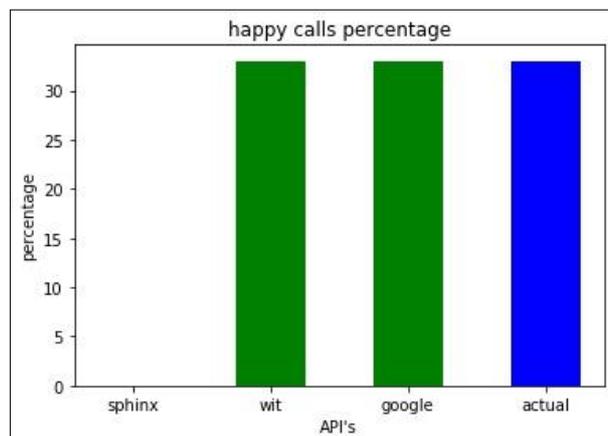


Fig 6: Happy calls percentage

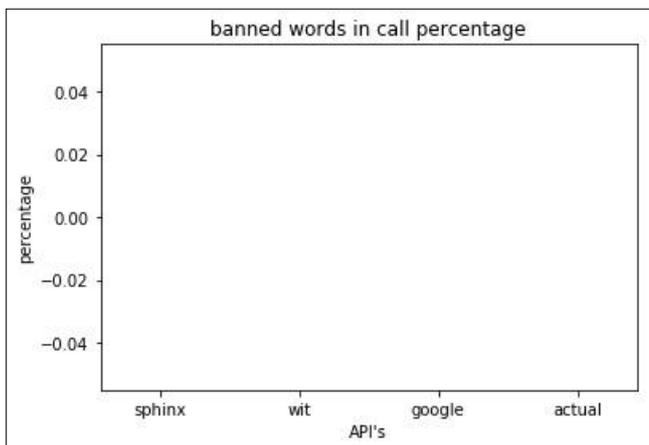


Fig 7: Banned words percentage

**5.4 There was no banned words in the calls, so it was not detected**

**5.4.1 Accuracy of competitors name detected:** Competitors name being a noun is very hard for API'S to detect in which wit.ai performed a very good job.

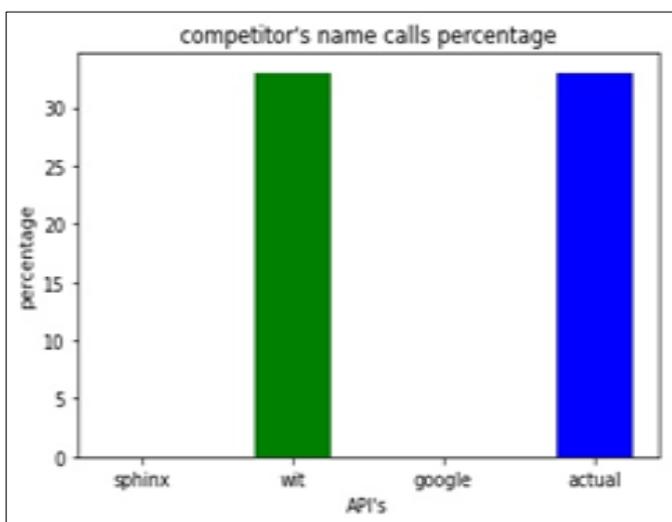


Fig 8: Competitor's name percentage

**6. Performance evaluation**

**6.1 Overall accuracy of human emotions detected:** From the above mentioned data, we found a cumulative overall accuracy of all human emotions detected in which wit.ai has 100% accuracy Google has 80%

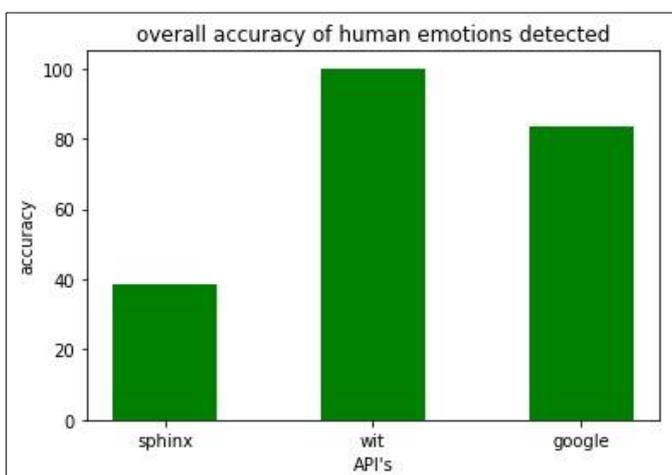


Fig 8: Overall accuracy of emotions

**Table 1:** Comparison of different API's

	Accuracy of human emotions	Accuracy of speech to text	suitable for long audio files	Voice can be trained
Sphinx	38%	63%	Yes	No
Wit.ai	100%	90%	No	Yes
Google	83%	87%	Yes	No

**7. Conclusions**

In this paper we have developed a software to measure performance of call centre representative and also to get useful insights for business analytics. We have compared different API's based on which we can use the API's for different conditions We are working on a speech to text convertor using fuzzy mathematics, and deep learning models, to increase the accuracy of the existing methods. We are also working on voice pitch based customer emotions.

**8. References**

1. Kopparapu SK. Non-Linguistic Analysis of Call Enter Conversations: Springer, 2015.
2. Natural Language Processing with Python by Steven Bird, Ewan Klein and Edward Loper.
3. Foundations of Statistical Natural Language Processing by Christopher Manning and Hinrich Schütze.
4. A note on Zadeh's extension principle by Masamichi Kon.
5. Chakraborty G, Pagolu M, Garla S. Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS: SAS Institute, 2014.
6. Takeuchi H, Subramaniam LV, Nasukawa T, Roy S. Automatic Identification of Important Segments and Expressions for Mining of Business-Oriented Conversations at Contact Centers, in EMNLP-CoNLL, 2007, 458-467.
7. Mishne G, Carmel G, Hoory R, Roytman A, Soffer A. Automatic analysis of call-center conversations, in Proceedings of the 14th ACM international conference on Information and knowledge management, 2005, 453-459.
8. Garnier-Rizet M, Adda G, Cailliau F, Gauvain JL, Guillemin-Lanne S, Lamel L *et al.* CallSurf: Automatic Transcription, Indexing and Structuration of Call Center Conversational Speech for Knowledge Extraction and Query by Content, in LREC, 2008.