

# *BirdEview: Advance version of call monitoring system by using mining techniques.*

Shivam Mehta

Department of Computer Engineering  
Watumull Institute of Electronics Engineering  
and Computer Technology  
Thane, India.  
Iamshivam27@gmail.com

Aarohi Mahajan

Department of Computer Engineering  
Watumull Institute of Electronics Engineering  
and Computer Technology  
Thane, India.  
Meaarohi99@gmail.com

Sneha Chitale

Department of Computer Engineering  
Watumull Institute of Electronics Engineering  
and Computer Technology  
Thane, India.  
Snehachitale26@gmail.com

Prof. Dhananjay Raut

Department of Computer Engineering  
Watumull Institute of Electronics Engineering  
and Computer Technology  
Thane, India.  
Dhananjay\_raut@yahoo.com

**Abstract**— The main objective is to ensure the standards of every call center with respect to its performance, services and get insights into customer's emotions involved with the company using different data mining techniques [1]. As the call center is the second most important thing just after the actual product it becomes very important to deal with the agent's performance and different methodologies to improve services and decrease the volume of the call flow by improving services. In order to achieve this, we do have to make sure that the agent is performing his/her task appropriately and in order to validate we will use Text mining to be able to monitor each and every call made or received by an agent. Another thing we are going to focus upon is the ability to get customer's emotional insights for every issue raised by customers and escalated by the agents, and these issues will be termed as Focus points [2]. Interaction mining will be the technology used for extracting emotional insights. So firstly when agents make an outbound call to the customer or will receive an inbound call by the customer, We will use Speaker diarisation to solve the problem of who spoke when during the call in order to perform data mining techniques live by converting speech to text for text mining and use emotion analysis for interaction mining[3][4][5][6][8]. After converting speech to the text we will compare to different key-points and customer's emotional input to generate a report for each and every call.

**Keywords**—interaction mining, text mining, machine learning, speaker diarisation, emotion analysis, agent analysis.

## I. INTRODUCTION

The paper is going to present a way to build software for the call center. Today Customer Service is considered the second most important sector after the actual product or service, generating billions of dollars each and every year. So generally in every call center, they have a team that is responsible for maintaining the quality of every agent associated with the call center [2]. They are referred to as a Quality Analyst in a team of Quality Assurance. Now the

process of quality assurance team is to randomly select a few numbers of calls and then listen to those calls and give them a score and generate a report depending on the performance of the agent. So the problem associated with this procedure is that each and every call is not been monitored and only a few calls are been randomly selected and quality analyst gives a score and generate a report based only on those calls. This makes it an unfair procedure to generate score and report. So to avoid this every calls of agents in the call center should be monitored [1]. Now that is practically impossible because of the ratio to number of agents and the number of the quality analyst in the call center are very less so monitoring every call is not possible because of every agent may attend up to 50 to 60 calls a day in small scale call center which may vary depending on the size of every call center. So, if we try to monitor every call made in just one day it may take up to a week just to evaluate every call by the quality analyst. An easy solution to this problem is to ask the customer's feedback to the calls which actually makes the task easy. When the customer would give feedback then the feedback will be evaluated by the quality analyst and scores and reports can be generated based on customer's feedback. But another problem with this is customer fails to give feedback every time which makes the whole procedure uncertain. The call center was facing this issue for a very long time and many types of research have come up to solve the problem of call monitoring. Which gave rise to a new method where calls were converted from speech to text, and then calls were diarised with an intention to know who spoke when and to apply key points in order to check whether proper protocol was followed [8]. On general terms, key indicators can be the proper call opening, Acknowledging the customer, To listen and to probe at the right time, Always show empathy, Give resolution, Close the call and dispose the call which included

feedback from an agent mentioning the reason of call. This solves the maximum problem but has many loopholes.

There always are going to be loopholes and a room for innovation. This makes us think about how we can minimize loopholes and make it a better version of itself. BirdEview came up with one solution, a solution that could do everything, which can minimize maximum loopholes, which can prove its robustness, it will be one software and one solution to everything. This software will cover all aspects having a full 360-degree view. The software will use Fully supervised speaker diarisation with an intention to know who spoke when. For speaker diarisation, the approach used will be called unbounded interleaved-state recurrent neural networks used in order to get the minimum error rate. Because of RNN, there will be a very minimum error rate of 7.6%. As it will be a supervised technique, it could use stochastic gradient descent for training the model. It will be able to decode both online and offline which will suit better for latency-sensitive software like for BirdEview or for any other call center [8].

After diarisation, the software will do two different things simultaneously. The first thing it would do during the call is to convert the whole conversation from speech to text. The model used for speech to text conversion will be HMM also known as Hidden Markov Model and in this model feature extraction will be achieved using MFCC (Mel Frequency Cepstral Coefficients)[6]. Vector quantization will be used to train the data set and Viterbi algorithm will be used to test the dataset after training. The second thing it will do emotional analysis using MFCC, conversion to the spectrogram and neural networks. Ryerson Audio-Visual Database of Emotional Speech and Song which is also known as RAVDESS will be used to train the dataset and test the same dataset.

After extracting the text it will be compared with the key performance indicators based on standards of the call center [3]. And at the end of the call, the software will complete its analysis and give output based on the emotion recognition. Finally generating a score based on both KPI and emotional analysis. It will also extract focus points which will be current issues faced by the customers. So finally when it will generate the report it will contain every small detail of a call like minutes of a call, focus point, score and other small important aspects based on each call.

## II. MOTIVATION

Today in call centers the quality control team is definitely giving their fullest but still, quality control is not accurate. Calls are randomly selected and based on randomly selected calls scores are allotted and reports are generated with no transparency to mistakes. Agents are directly called up by their quality analyst only if there are fatal calls more than the threshold. Presented software is responsible for monitoring each and every call and generating score and reports on the basis of every call made or received by agent with full transparency where agents get to know their mistake in real-

time and it compares with previous mistakes done by an agent which works as a reminder.

## III. RELATED SURVEY

Today call centers use different methodologies for maintaining the quality standard of the call centers. Well this approach for verification has a lot of loopholes which can also be termed as partial approach because it leads to inaccurate results. Today in call centers, each and every call is not monitored instead only few calls are randomly selected and based on those calls the score is generated for an agent. Which brings to a technology which can help call center to monitor each and every call in the call center [6]. By converting the speech to text and then applying text mining with intention to extract exact focus points [5]. Text mining can be done with different methods like Information extraction, Information Retrieval, Natural Language Processing, Clustering, and Text Summarization [6]. BirdEview will set standard KPI's in order to compare it with actual conversation between agent and the customer. Now based on the comparison between the actual conversation and KPI's the score can be generated. Now here each and every call is taken into consideration. BirdsEview will perform this on the live call instead of operating on the recorded call [3].

Emotional analysis used by BirdsEview is done considering both audio and text instead of text or audio individually. It can also be termed as hybrid approach where software uses both text and audio. Application of audio either text for emotion analysis leads to low accuracy that is 51% and 62% respectively and when combined and applied for analysis has an accuracy of 81%[7].

Now all this is only possible because of diarisation of live audio. The software will use Fully supervised speaker diarisation with an intention to know who spoke when. For speaker diarisation, the approach used will be called unbounded interleaved-state recurrent neural networks used in order to get the minimum error rate. Because of RNN, diarisation error rate or DER is of 7.6% [8]. As it will be a supervised technique, it will use stochastic gradient descent for training purposes.

Agent analysis's objective will be to increase agent performance which can lead to less dissatisfaction from the customer end.

#### IV. PROPOSED METHODOLOGY

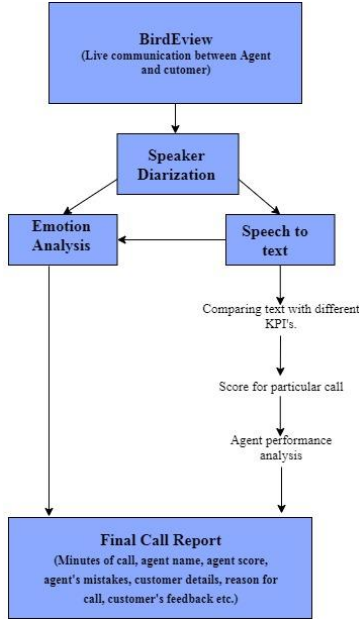


Figure 1: Proposed system

As shown in the above Figure 1, Software will barge in the call, monitoring the insights. Speaker diarisation will be operated on the live call with an intention to obtain speaker components which refer to different character component on the call.

##### A) Speaker diarisation



Figure 2: Speaker Diarisation

The diagram shown in Figure 2 shows a hypothetical situation, where it will show the working of speaker diarisation where the call is partitioned into different segments in order to extract components. Hereby applying speaker diarisation on live call where it will help to identify who spoke when in the conversation in order to perform agent analysis and emotion analysis. Speaker diarisation first monitors the whole conversation and then helps to identify components. Speaker diarisation follows a two-step procedure where at first it checks for speaker change points, which can be termed as speaker segmentation. Then after performing speaker segmentation, it will perform clustering on the bases of its characteristics [8]. In this case, as shown in Figure 2 where the audio stream which is live call is diarised and call center gets

to know who spoke at what particular point that is when agent spoke and when customer spoke in conversation.

##### B) Speech to text conversion



Figure 3: Speech to text conversion

The next step is to convert the diarised audio into text. For converting the speech into text as shown in figure 3, software will use conversion of speech to text approach where diarized audio will be considered as an input and exact texts will be formed. It is a three-step method where at first it recognizes the audio segments then it recognizes the language spoken and finally converts speech to text with its timecode [6].

##### C) Agent analysis

Attribute	Sub Attribute	Weightage
Call opening	Opened the call with the greeting. The agent did mention his name and company's name. Good energy displayed with a smile.	10%
Ask probing questions	Agent asked questions that help the customer clarify the problem or request. Agent asked relevant questions to understand the issue. Cannot be interrogative and be polite.	10%
Correctness of sentence formation, grammar and punctuation	Avoid lengthy sentences. Grammatical errors Miss pronunciation of words	5%
Call waiting Seconds.	Seconds taken by Agent to answer call by Customer	5%
Building Rapport	Showing Empathy. A high rate of speech. An agent cannot sound robotic. Have a positive attitude. Create a delightful experience that makes customers return feeling satisfied with the assistance provided. Proper hold procedure used.	20%
Call Closure.	Was the customer asked for further assistance? . Opportunity to educate the customer and give meaningful information for the future use.	10%
Completion of the call in a reasonable amount of time	The agent has not taken an unreasonable amount of time to resolve the issue.	10%
Responsibility	Taking responsibility correctly and escalate issue when needed.	10%
Manage delays with grace	Inform customers when there is a need to step out of call conversation for research by keeping call on hold. The expected call hold cannot exceed more than two minutes or else informing the customer for more time.	15%
Matching the customer's level of technical sophistication.	Avoid jargon, acronyms, or instructions that the customer may not understand.	5%
Call barging, Improvements, Call whispering.	Call barging only when needed. Call whispering should not be present. Did agent improve on mistakes done before	10%

Figure 4: Agent Analysis Score weightage

The agent will be judged on the following criteria where the text will be compared to the different KPI's as shown in figure 4. The score weightage for agents has been calculated out of 100%, where the different attribute has different weightage, let us see the first attribute in figure 4 is Call opening meaning if the agent greeted the customer well with his and company's name which is explained in the sub-attribute field, a 10% weightage has been given to this attribute, the second attribute is Ask probing question meaning if the agent asked the relevant question and didn't make the customer feel interrogated, weightage for this is 10%, similarly for other attributes we have distributed the weightage accordingly and in the sub-attribute, the points to be noted for the following attribute has been described.

The text will be compared and scores will be generated accordingly which will store on mistakes done by agents previously and compare next time to verify was improvement reflected or not. Well, the software has one more important thing which is known as the critical errors which are shown in the figure 5 [2]. Critical errors see the overall analysis of the agent if the agent answered all the customer's questions or not and whether the customer problem was solved or not.

If Agents does do not answer appropriately or if agents misguide customers or make any error for the escalating issues than that call be considered as a fatal call and the generated score will be zero.

Attribute	Sub Attribute
Answering to all the questions asked by customer appropriately	Right Attitude for communicating is most important.
Solving Customers problem	Raising proper complaints, escalating the proper complaints, following proper procedures and actions.

Figure 5: Agent analysis with respect to critical errors

#### D) Emotion Recognition

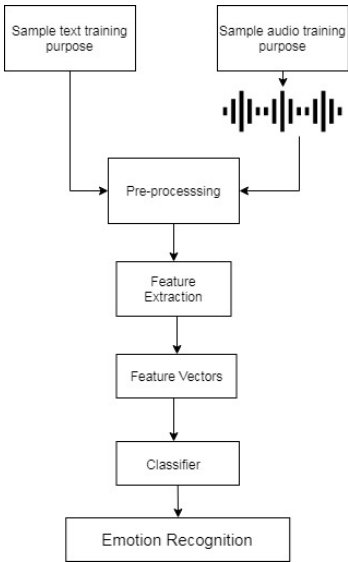


Figure 6: Emotion recognition for testing purpose

Emotion recognition is a process by which software will be able to recognize the emotion of a customer. The software will have a dataset for training purposes where different speech and text will be saved. Speech and Text will be pre-processed and then features will be extracted from both and finally, vectors will be obtained which will be used by the classifier. This software uses the Hybrid approach for emotion recognition which uses both audio and text for recognition. The hybrid approach is used with one motto which is to obtain more accuracy. According to one study if the software uses audio for emotion recognition then accuracy will be 51.1%, if the software uses text for emotion recognition then accuracy will be 62% and finally, if the software uses a Hybrid approach for emotion recognition than accuracy will be 81%[7]. Emotions can be characterized as happy, sad, fear, disgust, surprise, and anger. Emotion recognition can be used for business insights and can also be considered as feedback by customers as shown in figure 6.

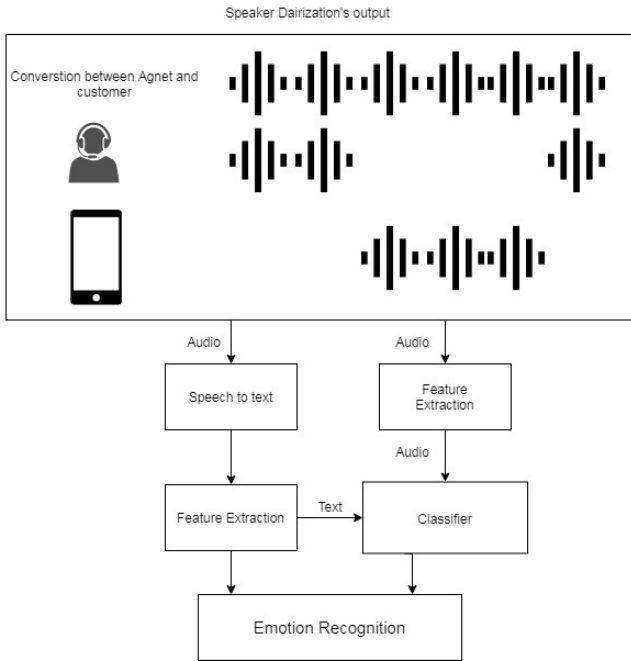


Figure 7: Emotion Recognition for testing purpose

#### E) The final report

Now finally after using different methodology the most important module of software is the final report. The final report consist of

- Agent's score obtained
- Agent's Name
- Agent's Mistake
- Length of call
- Customer id
- Customer detail
- Reason for call
- Emotion
- Resolution
- Feedback

The final report will be originated for each and every call in the call center [3].

## V. RESULT AND DISCUSSION

In this section we present the results of our proposed model on which we applied all above mentioned techniques. The database used for emotional analysis for testing and training is RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song). The accuracy of the model is 91.86% and the loss incurred by the model is 8.78% shown in figure 8.

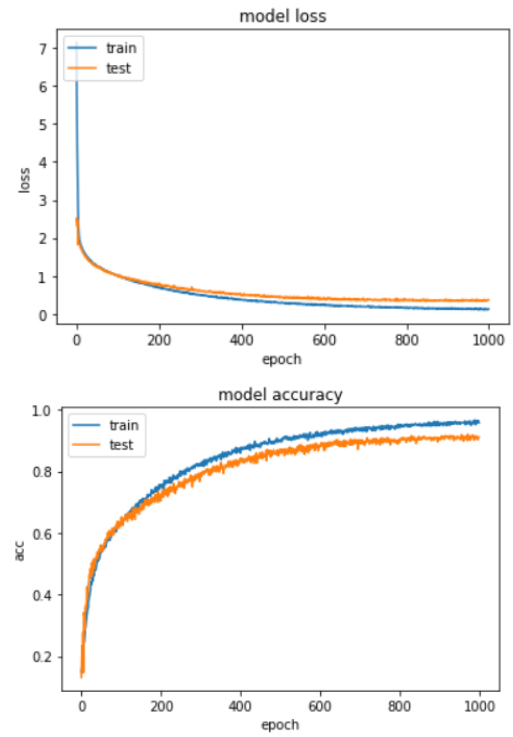


Figure 8: Accuracy and loss model

After the speech to text transcription is completed, shown in figure 9 and then the agent analysis is given in an .xml file i.e. the score according to the analysis is displayed in the final report, shown in figure 10.

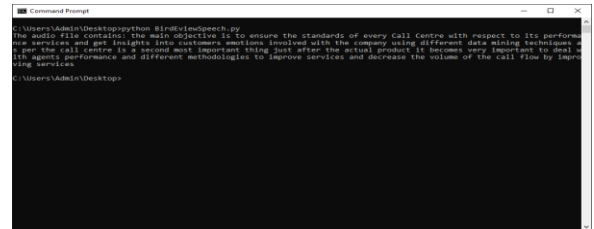


Figure 9: Speech to text transcription

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<root>
  <Agent_Name="str">John</Agent_Name>
  <Agent_score="int">100</Agent_score>
  <Call_length="str">00:43</Call_length>
  <Customer_id="int">1</Customer_id>
  <Customer_name="str">abc</Customer_name>
  <Call_reason="str">call center standards</Call_reason>
  <Emotion="str">neutral</Emotion>
  <Resolution="str">clear</Resolution>
  <Feedback="str">good</Feedback>
</root>

```

Figure 10: Final Report

## VI. CONCLUSION AND FUTURE SCOPE

With the help of our system, Customer-Agent call analysis can be done which will help call centers to present a reliable monitoring system resulting in accurate performance measurements by analyzing all the incoming and outgoing

calls, reduce total call time, correctly route the customers to respective agents to solve their needs and increase customer satisfaction.

The individual modules will be unit tested according to criteria set by the call center. The unit tests include and are not restricted to test sets for all the modules. The integrated system will evaluate scripted calls to provide results. These results will be evaluated against customer feedback. The customer feedback will help validate against the performance of the system. Another evaluation parameter will include an independent team to individually go through each call recording and evaluate it by themselves. This will be measured against the real time evaluation done by the system to validate its performance.

The future scope of our project can be enhanced by including voice activity detection which will help negate all the parts of the speech where there is no activity and also talk over analysis that will help to mine information directly from the speech. And future scope for BirdEview is applying Artificial intelligence to software and allow agent to get real time assistance from software itself instead of team leaders or senior employees. By using AI concepts software will automatically raise the issue, escalate the issue, give agent real time solution right in front of agent. This is the future of call center.

## VII. REFERENCES

- [1] <https://aircall.io/blog/customer-happiness/call-center-quality-assurance/>.
- [2] <https://info.scorebuddyqa.com/ultimate-guide-call-center-qa>.
- [3] Pallotta, Vincenzo & Delmonte, Rodolfo & Vrieling, Lammert & Walker, David. (2011). Interaction Mining: The new frontier of Call Center Analytics. CEUR Workshop Proceedings. 771.
- [4] Paprzycki, Marcin & Abraham, Ajith & Guo, Ruiyuan & Mukkamala, Srinivas. (2004). Data Mining Approach for Analyzing Call Center Performance. Computing Research Repository - CORR. 1092-1101. 10.1007/978-3-540-24677-0\_112.
- [5] Yiğit, İbrahim & Güvercin, Mehmet & Ferhatosmanoğlu, Hakan & Gedik, Buğra & Ates, Ahmet. (2017). Call center text mining approach. 10.1109/SIU.2017.7960138.
- [6] Das, Prerana & Acharjee, Kakali & Das, Pranab & Prasad, Vijay. (2015). VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT. Journal of Applied and Fundamental Sciences. 1. 2395-5562.
- [7] Bhaskar, Jasmine & Sruthi, K. & Nedungadi, Prema. (2015). Hybrid Approach for Emotion Classification of Audio Conversation Based on Text and Speech Mining. Procedia Computer Science. 46. 635-643. 10.1016/j.procs.2015.02.112.
- [8] A. Zhang, Q. Wang, Z. Zhu, J. Paisley and C. Wang, "Fully Supervised Speaker Diarization," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 2019, pp. 6301-6305.
- [9] Price, David & Nallasivan, Srinivasan. (2014). Call center quality management.
- [10] M. Garnier-Rizet, G. Adda, F. Cailliau, J.-L. Gauvain, S. Guillemin-Lanne, L. Lamel, et al., "CallSurf: Automatic Transcription, Indexing and Structuration of Call Center Conversational Speech for Knowledge Extraction and Query by Content," in *LREC*, 2008.
- [11] Redesigning the Service Quality Measurement Factors for the Airline Call Centers in Jong-In Choi Passenger Service Center, Koreanair
- [12] Ismail, S., & Ali Bou, N. (2018). Utilizing Third-Order Hidden Markov Models for Emotional Talking Condition Recognition. 2018 14th IEEE International Conference on Signal Processing (ICSP). doi:10.1109/icsp.2018.8652352