A Novel Approach to Identify Problematic Call Center Conversations

Meghna Abhishek Pandharipande
TCS Innovation Labs - Mumbai
Tata Consultancy Services Limited
Yantra Park, Thane (West), India - 400601
Email: meghna.pandharipande@tcs.com

Sunil Kumar Kopparapu
TCS Innovation Labs - Mumbai
Tata Consultancy Services Limited
Yantra Park, Thane (West), India - 400601
Email: sunilkumar.kopparapu@tcs.com

***Abstract*—Voice based call centers enable customers to query for information by speaking to agents in the call center. Most often these call conversations are recorded for analysis with the intent of trying to identify things that can help improve the performance of the call center to serve the customer better. Today the recorded conversations are analyzed by humans by listening to call conversations, which is both time consuming, fatigue prone and not accurate. Additionally, humans are able to analyze only a small percentage of the total calls because of economics. In this paper, we propose a visual method to identify problem calls quickly. The idea is to sieve through all the calls and identify problem calls, these calls can then be further analyzed by human. We first model call conversations as a directed graph and then identify a structure associated with a normal call. All call conversations that do not have the structure of a normal call are then classified as being abnormal. In this paper, we use the speaking rate feature to model call conversation because it makes it easy to spot potential problem calls. We have experimented on real call center conversations acquired from a call center and the results are encouraging.**

***Index Terms*—Call Conversations, Speaking Rate; Speech Analytics; Modeling; Visual analysis of Speech**

I. Introduction

Call center businesses record customer telephonic interactions to extract valuable insight into products, strategy, services, process and operational issues. Speech Analytics (SA) is the method of automatically analyzing recorded calls to extract useful and usable information. An accurate analysis of the call center conversations shed light on some very crucial usable information that would otherwise be lost [1] and [2]. Analyzing conversations is an expensive task if done manually in addition to being not comprehensive, typically, a small fraction of all the recorded call conversations are carefully heard by human supervisor; further the entire conversation is listened to determine if the call is a problematic call or not. Clearly, there is a large portion of the recorded calls that are not part of the calls listened to and hence more often than not the problematic calls might not be part of the human analysis.

However with the advent of Automatic Speech Recognition (ASR) technology, the task reduces to automatic transcription of telephone calls followed by analyzing words or phrases in the text transcriptions [3]. However, the process of converting audio conversations into transcribed text is not very accurate, typically 50-60% recognition accuracies, even if one uses the state-of-art speech recognition technology. The accuracies worsen further when there is no readily available ASR for the language spoken in the conversation [4]. In many cases transcription of the audio conversation is not sufficient to identify a problem call conversations, because the problem might not translate into meaningfully transcribed text. For example, *Thank You* spoken in a sarcastic tone might not suggest the problem in the call because the phrase "Thank You" is generally associated with a non-problematic call.

In this paper, we describe a method to enable automatic identification of problematic calls without actually transcribing the audio conversations into text. We use the speaking rate [5] feature to abstract the call conversation and use directed graph to represent a call conversation. We identify a structure of the directed graph which represents a normal call. Any call conversation which does not have the structure of a normal call is then flagged abnormal. The main contributions of this paper are (a) Use of a paralinguistic feature to represent a call conversation, (b) Modeling a call conversation as a directed graph and using this directed graph structure to identify abnormal call conversations. The rest of the paper is organized as follows. With a brief literature survey in Section II we model a call conversation in Section III as a directed graph. In Section IV we describe the speaking rate feature and capture a typical call center conversation in Section V. We present our results in Section VI and conclude in Section VII.

II. Brief Literature Survey

This is relatively a new area and to the best of our knowledge not much has been done in this area. In literature most of the work is being done on transcribed call conversation, which means we have to depend on the ASR for audio transcribed text, typically yielding 50-60% recognition accuracies. Hironori [6] describes a method to identify important segments from transcribed textual records of conversations between customers and agents. They look for changes in the accuracy of a categorizer designed to separate different business outcomes.

Gilad [7] describes a system that automatically transcribes calls using a speech recognition engine. The domain specific importance of the conversation fragments is identified based on the divergence of corpus statistics. This is used to analyze the content of the call conversation. They further use information retrieval approaches on the transcribed text to provide knowledge mining tools for both call-center agent and for administrators of the center. The system developed in [7] helps in gaining insight hidden in the recorded calls, which can help reduce cost of operation and improve products, processes. This enables making quality monitoring more effective by routing calls about key business issues to supervisor for review affecting the overall customer experience.

Vincenzo [8] provides a solution for pragmatic analysis of call center conversations in order to provide useful insights for enhancing Call Center Analytics to a level that will enable new metrics and Key Performance Indicators (KPIs) beyond the standard approach. These metrics rely on understanding the dynamics of conversations by highlighting the way participants discuss about topics. By this, they claim, one can detect situations that are simply impossible to detect with standard approaches such as controversial topics, customer-oriented behaviors and also predict customer ratings.

III. Modelling Call Conversations

A typical call center conversation between the agent in the call center and the customer is a sequence of speech segments spoken by the agent and the customer. In its most coarse form it can be represented as a directed graph with nodes and edges as shown in Fig.1. Here the red arrow shows the beginning of the conversation, while the blue arrow shows the end of the conversation. Typically, when the call starts the agent starts off with a welcome message and also ends the call with a greeting or a good bye or thank you message. The rest of the conversation is due to the customer (Customer Speak) or the agent (Agent Speak) and is interlaced, this is shown by continuous black lines in Fig 1. The dashed lines in Fig 1 represent the agent and the customer speaking with a pause. The edge weights P_{cc}, P_{ca}, P_{ac} and P_{aa} are the probabilities and represent the probability that the customer speaks after the customer speaks, agent speaks after the customer, customer speaks after the agent and agent speaks after the agent respectively. In some sense Fig. 1 captures the conversation in a typical call center conversation. Clearly each of the edges in the graph comes with a probability that determines and captures the nature of the conversation. For example if $P_{cc} > P_{ca}$ then the customer is speaking more and not allowing the agent to speak indicating customer being angry or unsatisfied. During a typical normal interaction between the customer and the agent one can expect where $P_{ac} > P_{aa}$ and $P_{ca} > P_{cc}$ where the agent speaks and also allows the customer to speak and so on, a typical conversation.

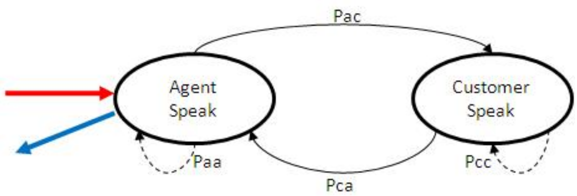


Fig. 1. Level 0 Model of a Normal Call Center Conversation

However a realistic call conversation has hold time, the time when none of them (agent or the customer) is speaking because the agent is doing something so as to facilitate fetching the information sought by the customer, so there is another node associated with *hold music* state. In a conversation the hold music state is entered from the agent speak state and exits to agent speak state. Clearly if the conversation stays in the hold music state (large P_{mm}, see Fig. 2.), then the agent is performing poorly because he is putting the customer on wait or the customer has a complex query that is not being addressed by the agent. In a similar vein if P_{am} is larger than P_{aa} or P_{ac} then the agent seems to be putting the customer on hold a large number of times indicating agent unable to resolve customer query.

Clearly this mode of modeling a call center conversation enables one to understand certain aspects of the call in terms of performance of the agent. Understandably, the probabilities P_{**} capture the complete call conversation and gives an overall picture of the call.

Note 1: In order to identify the probabilities P_{**} the call conversation needs to be segmented into sections that have been spoken by agent and the customer [9].

Note 2: The probabilities P_{**} do not in themselves help in identifying portions of the call that might be not normal.

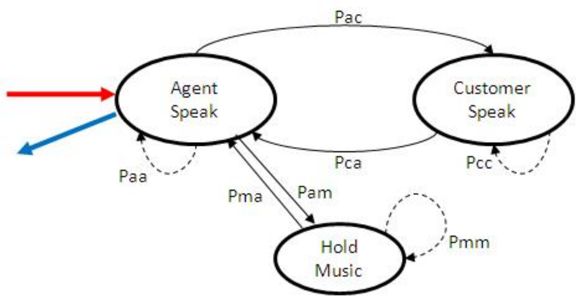
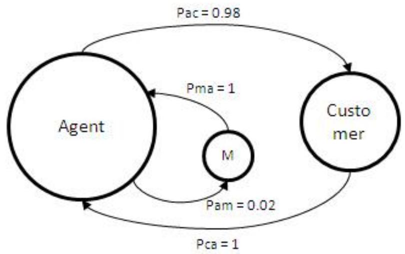


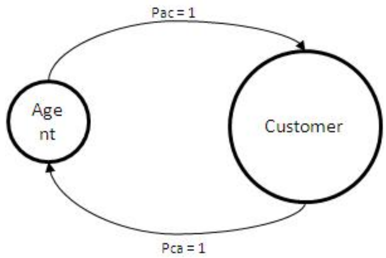
Fig. 2. Level 1 Model of a Realistic Call Center Conversation

To check the use of the directed graph to identify abnormal calls we randomly selected some 75 call center conversations. As a first step the calls were segmented into sections spoken by agent and customer using an automated method [9]. Subsequently, we found the number of transitions from one node (here node refers to Agent, Customer and Hold) to another and also calculated the duration of speech in a particular node. It was found that in a typical normal call the probability of agent talking to customer P_{ac} is much greater than the probability of customer talking to an agent P_{ca} almost double. Also the probability of agent putting the customer to hold $P_{am} << P_{ac}$ Fig. 3(a) shows a typical directed graph of a normal call while Fig. 4(a) shows the directed graph of a abnormal or problematic call. Note that the size of the node of the customer speak is bigger than the node corresponding to agent speak, meaning the customer is talking more than the agent. Which



(a) Example

Fig. 3. Directed Graph Models of Normal Call Conversations



(a) Example

Fig. 4. Directed Graph Models of Abnormal Call Conversations

signifies a problematic call, while in a normal call the size of the node corresponding to the agent speak is bigger than the size of the node corresponding to customer speak (see Fig. 3(a)). However, this gross analysis based on directed graph is not enough to identify the actual location of the abnormality. We need to model the call at a better resolution, we discuss this in the next section where we use the speaking rate as the feature to represent the call conversation.

IV. Speaking Rate Feature

In general speaking rate is measured as number of spoken words per minute (WPM). While there are several approaches to identify the speaking rate, one of the well known method of measuring speaking rate involves identifying the syllables in the spoken speech to compute the speaking rate. An algorithm to detect a syllable reliably in a spoken speech has been described in [10]. The syllable detection is based on the intensity (loudness) of the spoken speech and also the voicedness of speech. The pauses in the spoken speech are also identified so as that the number of occurrences of the syllables per unit time gives an accurate estimate of the speaking rate. In all our experiments we used the algorithm in [10] to determine the speaking rate. Once the speaking rate is computed in terms of number of syllables per second (sps), we can compute the speaking rate in words per minute (wpm) using a conversion factor of $\gamma = 1.5$ as suggested by Yaruss [11], namely,

$$SR_{WPM} = \gamma \times SR_{sps} \times 60 \quad (1)$$

where SR_{WPM} is the speaking rate in words per minute, SR_{sps} is the number of syllables per second and γ is the conversion factor between syllable and word.

Note 3: The factor γ is dependent on the language and this choice of $\gamma = 1.5$ is applicable to conversational English.

For analysis of call conversation we used the speaking rate feature. The speaking rate was computed on a segment of the call conversation spoken by the same person (agent or customer) in one stretch. As seen in Fig. 5, a typical call conversation can be segmented into portions of speech spoken by the agent, that spoken by the customer and the hold music. Each of this segment which is not hold music, can be analyzed to estimate the speaking rate, the speaking rate during hold music segment is assumed to be zero. The speaking rate due to the agent is shown on the positive y-axis and that of the customer is shown on the negative y-axis. The amplitude of this represents the speaking rate; higher amplitude represents a higher speaking rate than lower amplitude. A typical speaking rate pattern of a complete call conversation of 598.7 seconds duration is shown in Fig. 5.

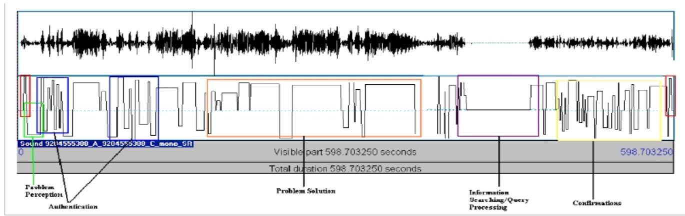


Fig. 5. Typical Speaking Rate Pattern of a Call Conversation

Note 4: Clearly, since we are representing the speaking rate on the negative y-axis for the customer, one needs to consider the absolute value of the speaking rate so a higher absolute value would represent faster speaking rate.

V. Speaking Rate Patterns in Call Conversations

There are several patterns that one can observe in the call conversations. These have been captured and appropriate portion in the conversation has been highlighted in figures to articulate the pattern observed. Speaking Rate of agent is high at the start and at the end of a call conversation Fig.6. This is to be expected as very often the agent is either reading a written script or has uttered the paragraph so many times that it becomes his second nature. For example: *Very Good Evening, welcome to* $< company/product >$ *care this is Sujit how may I assist you* at the beginning of the call or *Thank you for calling* $< company/product >$ *Care have a nice day* at the end of the call. A customer describing the problem shows

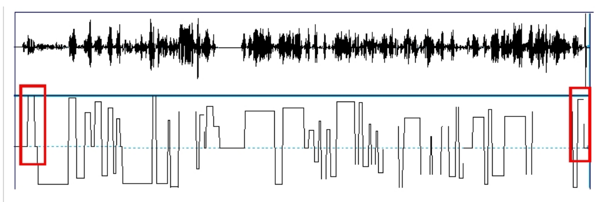


Fig. 6. Speaking Rate is high at the start and end of the call

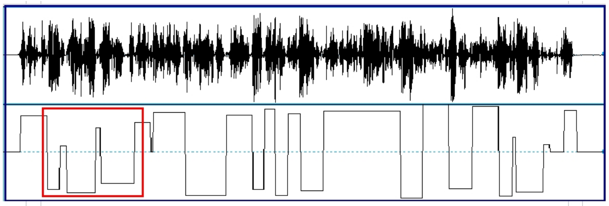


Fig. 7. Problem Description by Customer

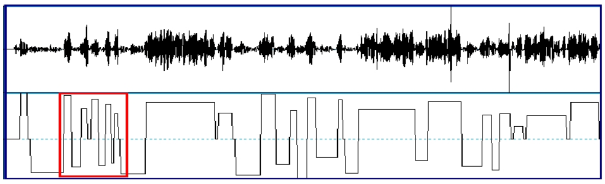


Fig. 8. Authentication Process

up as a pattern shown in Fig. 7. The pattern has more or less a constant speaking rate with intermittent agent speaks of very short duration. Fig. 8. shows the authentication procedure. Typically, the agent makes sure that he is indeed speaking to the person whom s/he is supposed to speak. Asking for name/contact number/card details. Agent providing solution (to customer query) has a pattern shown in Fig. 9. Typically the agent speaks for quite some time with small pauses indicates, agent providing solution to customer query. Agent searching for some information has a pattern shown in Fig. 10.

Clearly the speaking rate shows patterns which displays certain aspects of the call as mentioned above. Some of the other aspects that can also be identified are regions where the customer is speaking with a very high speaking rate and continuously Fig. 11 which probably indicates the customer is not happy for some reason. Additionally call ending abruptly with no "Thank you" message from the agent (Fig. 12) indicates the customer being unsatisfied or upset.

We plan to use the observed patterns in the speaking rate feature to identify the nature of transaction in the call conversation. For example, the nature of transactions can be

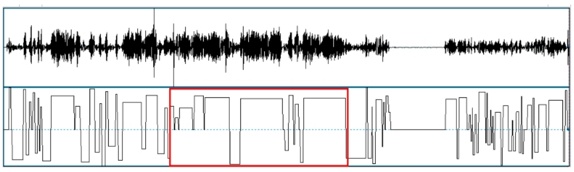


Fig. 9. Solution being provided by agent

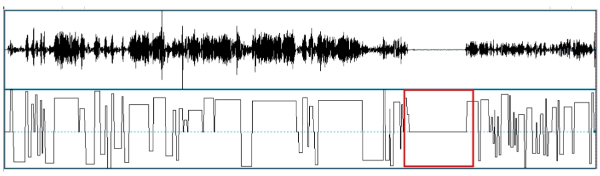


Fig. 10. Agent searching for Information

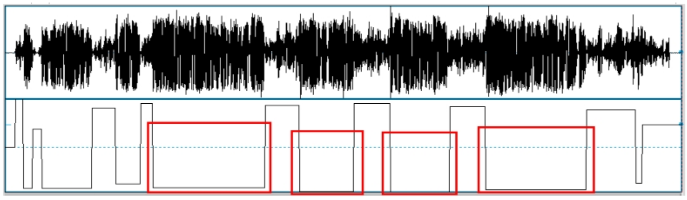


Fig. 11. Abnormal Call Pattern: Customer not allowing agent speak (Customer upset)

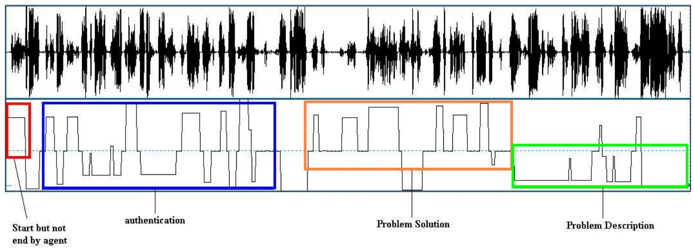


Fig. 12. Abnormal Call Pattern: Call ends abruptly no Thanks message from agent (Customer upset)

enumerated as follows:

1) $[WM]$ Welcome message by the agent
2) $[PD]$ Problem Description by Customer
3) $[CA]$ Customer authentication,
4) $[AS]$ Agent providing a solution
5) $[AC]$ Agent confirming the customers understanding of the solution
6) $[OP]$ Pauses, Music
7) $[EM]$ End message by the agent.

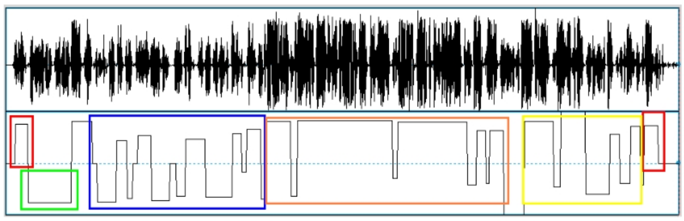


Fig. 13. Typical Speaking Rate Pattern of a Call Conversation

Fig. 13 Captures a typical call center conversation. The red box at the beginning is the $[WM]$ component and indicates the start and the red box at the end of call indicates the component $[EM]$. The green box indicates description of problem by customer which is the $[PD]$ component while the blue box indicates the authentication process $[CA]$, where the agent makes sure that he is indeed speaking to the person he should be speaking to. The orange box indicates a typical solution being provided by the agent to customer $[AS]$. The yellow box indicates confirmation of the solution provided by the agent and affirmation by the customer to the solution $[AC]$. Table I captures this in a nutshell.

Note 5: A normal call can be defined as one of having a very specific pattern or transactions components in the entire conversation.

Transaction	Pattern (SR)	Who
Welcome Message (WM)	High	Agent
Problem Description (PD)	Uniform	Customer
Authentication (CA)	Low	Agent-Cust
Agent Solution (AS)	Uniform	Agent
Solution Confirmation (SC)	Uniform	Agent-Cust
End Message (EM)	High	Agent

TABLE I
PATTERNS IN A NORMAL CALL

Note 6: A normal call has typical components which are in a particular time sequence.
In our experiments, we used the pattern of a normal call (call it a normal call model, Fig. 13) and then compared a test call with this reference normal call model to identify how close the call was to a normal call.

VI. EXPERIMENTAL RESULTS

For the purpose of experiments we analyzed a set of 100 call center conversations obtained from three different call center catering to Insurance and Telecom domain. In this 100 calls, 10 calls were abnormal, for example, in one of the calls the customer had put down the phone before the call had ended. The rest of the 90 calls were normal as described in previous section. For a given conversation, we segmented the conversation into smaller segments first into voice and music [12] and then all the segments that were marked as voice were further segmented into spoken by agent or spoken by customer [9]. For each of the segment, we computed the speaking rate of the segment.

We marked the conversations using the speaking rate feature to determine the type of transaction, namely, one of $[WM]$, $[PD]$, $[CA]$, $[AS]$, $[AC]$, $[OP]$, $[EM]$ without actually hearing to the audio conversation (using Table I). We selected a few samples at random to check the correctness of the marking of the transaction by manually listening to the audio segments, and found that we were able to mark with more than 90 % accuracy. Now each call conversation to be analyzed as normal or abnormal was first segmented into voice and music [12] and then all the segments that were marked as voice we segmented into spoken by agent or spoken by customer [9]. We then tested the call using the directed graph method as level 1 and then we marked the segments with one of the labels ($[WM]$, $[PD]$, $[CA]$, $[AS]$, $[AC]$, $[OP]$, $[EM]$) along with the duration of the label. Any call that (a) missed one or more labels, or (b) or a certain label that had an unusual duration was marked as being abnormal. This process yielded 90 % results in the sense that we were unable to mark 1 abnormal call in the 10 abnormal calls in our dataset. However, as many as 7 normal calls were marked as being abnormal because we had falsely mislabled the conversation with transaction labels based on the observed pattern.

Note 7: We are in the process of identifying pattern matching techniques that can be used to reliably mark the portions of conversation as one of the transaction labels, namely,$[WM]$, $[PD]$, $[CA]$, $[AS]$, $[AC]$, $[OP]$, $[EM]$.

VII. CONCLUSION

Contact centers today are either counting on managers to manually evaluates recorded call conversation. Call centers either rely on individuals to analyze calls, or aren't performing any analysis at all. The adoption of speech analytics, the process of converting audio to text using speech recognition techniques, is being used but is both expensive and has problems in determining a thank you said with sarcasm versus a genuine thank you. This poses problem in determining a abnormal or a problematic call from a normal call conversation. In this paper we have proposed a novel way of visually identifying an abnormal call from a normal call center conversation without actually transcribing or manually hearing the call. The method used speaking rate feature to represent the call conversation and observation of patterns in this speaking rate space and mapping them to transactions in the call conversation. This is one of the contributions of this paper. The experimental results show that it is possible to indeed identify abnormal call center conversations using this methodology with very high accuracies.

REFERENCES

[1] C. Bailor, "The why factor in speech analytics," *CRM Magazine*, vol. August, 2006.

[2] wikipedia. [Online]. Available: http://www.wikipedia.com/

[3] VERINT, "Speech analytics essentials for audiolog," *IMPACT360*.

[4] S. K. Kopparapu and I. Ahmed, "Enabling rapid prototyping of an existing speech solution into another language," in *14th Oriental COCOSDA Conference*, Oct 2011.

[5] M. Pandharipande and S. Kopparapu, "Real time speaking rate monitoring system," in *Signal Processing, Communications and Computing (ICSPCC), 2011 IEEE International Conference on*, sept. 2011, pp. 1 –4.

[6] H. Takeuchi, L. V. Subramaniam, T. Nasukawa, and S. Roy, "Automatic identification of important segments and expressions for mining of business-oriented conversations at contact centers," in *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 2007.

[7] G. Mishne, D. Carmel, and R. Hoory, "Automatic analysis of call center conversations," in *ACM international conference on Information and knowledge management*, 2005.

[8] V. Pallotta, R. Delmonte, L. Vrieling, and D. Walker, "Interaction mining: the new frontier of call center analytics," in *5th International Workshop on New Challenges in Distributed Information Filtering and Retrieval*, 2011.

[9] S. Kopparapu, A. Imran, and G. Sita, "A two pass algorithm for speaker change detection," in *TENCON 2010 - 2010 IEEE Region 10 Conference*, nov. 2010, pp. 755 –758.

[10] N. H. D. Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior Research Methods*, vol. 41, pp. 385–390, 2009.

[11] J. S. Yaruss, "Converting between word and syllable counts in children's conversational speech samples," *Journal of Fluency Disorders*, vol. 25, no. 4, pp. 305 – 316, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0094730X00000887

[12] S. Kopparapu, M. Pandharipande, and G. Sita, "Music and vocal separation using multiband modulation based features," in *Industrial Electronics Applications (ISIEA), 2010 IEEE Symposium on*, oct. 2010, pp. 733 –737.