ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Review Article

Open Access

Revolutionizing Call Centers through ASR and Advance Speech Analytics

Ashish Bansal

USA

ABSTRACT

Technological improvement and innovation have led to the development of the classic contact call center into an omnichannel call center. The progress has occurred because customers are exposed to the use of SMS but also social media, chats, and other websites. Therefore, customers can contact a company through diverse ways creating a crucial analytics case for most call center companies.

In this paper, we will briefly outline speech recognition and analytics and discuss their benefits. As well as we will talk about how a hybrid system can help not just converting audio to text but also how downstream NLP tasks can be utilized to get more advance call analytics. These NLP components analyze the transcribed text, extracting the customer's intent, context, and specific requests. They comprehend the customer's query and gather the necessary information to generate relevant and accurate responses. Subsequently, a text-to-speech model is employed to convert the generated response back into voice format for delivery to the customer.

***Corresponding author**

Ashish Bansal, USA.

Received: January 01, 2022; **Accepted:** January 08, 2022; **Published:** January 29, 2022

Keywords: Speech to Text, NLP, Call Analytics, Classification, Call Center Technology, ASR

Introduction

In the bustling world of enterprise call centers, handling anywhere from hundreds to hundreds of thousands of calls daily is the norm. These interactions generate vast amounts of real-time data, providing invaluable insights into customer behavior, employee performance, and overall business operations. However, the sheer volume and complexity of this data make it challenging to access, analyze, and extract meaningful information.

Speech-to-text analytics, also known as voice or speech analytics, has evolved beyond simple transcription to include a deep understanding of the meanings, emotions, and contexts conveyed in customer interactions. This technology converts spoken words into written text, which can then be analyzed using advanced AI and natural language processing (NLP) algorithms. These systems not only capture the content of conversations but also identify key metrics such as sentiment, trends, and anomalies, providing a comprehensive view of the customer experience.

In order for speech analytics to work in a call center, there needs to be some automatic speech recognition software that is used to convert speech into text (either in real time or in a batch). Currently, automatic speech recognition techniques has advanced in leaps and bounds the past 10 years. There are a large number of ASR solutions that work well in a call center environment.

These solutions are designed on simple acoustic model to advance deep learning and generative AI models ranging from Kaldi to Whisper. Some other ASR's are wav2letter++, openseq2seq, vosk, Speech Brain, Nvidia Nemo, and Fairseq.

Powered by Artificial Intelligence, an ASR system uses Natural Language Processing (NLP), acoustic and language models, and Machine Learning (ML) technologies to accurately understand and transcribe spoken language into written text, which can then be used for different purposes. Mod- ern sophisticated ASR systems are even capable of understanding jargon, accents, and different speech patterns. As well as able to detect multiple language used on a call and has capability to transcribe them in multiple languages. Some of the ASR also has translation capabilities translating from one language to another for the ease of implementing such technology across geographical domains.

Let's look at a specific example of how ASR works in Generative AI voice bots to enable human- like conversations and provide self-service support to customers without involving live call center agents. First, a customer initiates a voice interaction through a phone call or a voice-enabled de- vice. Call center speech recognition software uses ASR to transcribe the customer's spoken words into written text to process and accurately understand the customer's intent.

Once the voice input is converted into text, an ASR system employs NLP to analyze the transcribed text and extract valuable information from it using special models trained on extensive amounts of data. Based on that, an ASR system pulls together the necessary information and gen erates a relevant response, which is then converted back into voice format through text-to-speech technology.

Automatic Speech Recognition is a valuable technology that is currently being actively implemented and used in call centers and contact centers for a variety of purposes. ASR helps stream- line call center operations, accurately route customer calls, deliver self-service customer support through human-like conversations, maintain compliance, gain valuable insights from voice inter-actions, and reduce overall customer support costs.

Automatic Speech Recognition (ASR) has been a key tool for Contact Center as a Service (CCaaS) companies in their quest to automate and improve customer query processing. By using ASR solutions, companies can offer more flexible and satisfactory customer service, and have access to advanced technologies and analytics based on industry best practices.

Although old speech recognition technology was inaccurate due to industry-specific jargon and poor call quality, end-to-end deep learning has enabled the creation of accurate models with new data. ASR solutions are divided into two categories: speech recognition and speech comprehension. Both are particularly relevant to the call center, as it helps improve voice recognition and understanding the meaning behind what is being said. Converting speech to text is an important first step in speech analytics where accuracy of ASR drives the down stream advance call analytics solutions such as speaker segmention, customer intent, and various NLP tasks. Better the quality of call transcripts would have a direct impact on the down stream NLP tasks.

When it comes to Automatic Speech Recognition, examples of applications of ASR technology within today's call centers can be found in many call center features and solutions. Below are the most common ones:

- **Perform Automated Quality Assurance ("AQA"):** You can use speech analytics to analyze some or all of your calls and automatically apply quality assurance rules to identify calls that violate your rules. AQA is commonly done on stored call recordings to transcribe calls for text analysis.
- **Perform Sentiment Analysis:** You can use speech analytics to predict the sentiment of a caller. Real time speech analytics can be used to signal the sentiment of a caller to the agent (or to a supervisor). Sentiment analysis can also be used to trigger special scripts or messaging an agent may use to respond to a caller's sentiment.
- **Perform Content Moderation:** You can currently use speech analytics in your call center to bleep or cut out sensitive or offensive content. For example, you can establish rules to redact curse words, or to redact sensitive data (e.g., such as PII data).
- **Automatically Identify Trigger Words:** Real time speech analytic processing can be used to identify trigger words (or words that require special handling). For example, a sales call cen ter may use speech analytics to monitor for words like "returns" or "attorney" or "attorney general". When those words are identified, a supervisor may be alerted or a special script may be presented to the agent to handle the situation properly.
- **Translate Calls:** Speech processing can be used in a call center to perform real time translation of callers. Sometimes, call center agents are faced with callers who speak a different language. Real time speech processing can translate the caller (and the agent) language to allow the call to be handled or transferred to the appropriate agent or destination.
- **Enable Self-Service and IVR:** With the adoption of ASR, IVR (Interactive Voice Response) systems have become more intelligent. ASR-powered Interactive Voice Response systems can capture information about customer inquiries and better identify customer needs and intentions. IVR self-service menus can help customers perform a wide range of basic tasks, such as requesting information, placing orders, making appointments and reservations, or completing transactions using voice commands. That can ultimately reduce the workload on your agents and improve customer satisfaction. ASR also facilitates call routing and helps better direct incoming calls to the most appropriate departments, teams, or individual agents based on customer intent, improving your FCR (First Call Resolution) rates and reducing the Average Handle Times.
- **Assist Agents:** Speech analytic tools are often used to provide some agent assistance (e.g., such as by prompting agents with information during a call using call whispering).
- **Conversational AI Voice Bots:** Along with AI chatbots, AI voice bots are becoming increasingly popular with consumers as they provide a more natural interaction experience and are faster to communicate with. By using the power of ASR, NLP, natural language understanding (NLU), and machine learning technologies, conversational AI voice bots and virtual assistants can automate customer service interactions, allowing you to deliver self-service customer support, which is highly expected by today's consumers. Voice bots can provide information, answer common questions, guide customers through processes, facilitate transactions, and resolve issues. The speed and 24/7 availability of voice bots and virtual assistants can help call centers improve customer experience.
- **Voice Biometrics** Call center speech recognition solutions can be used in call centers to identify and verify incoming callers through voice biometrics by analyzing the unique characteristics of the caller's voice, such as tone, pitch, and speech patterns. That not only helps improve security and streamline the caller authentication process but can also dramatically improve the experience for both callers and agents. On top of that, empowering your agents to authenticate customers without having to request additional information from them saves them a great deal of time and effort.

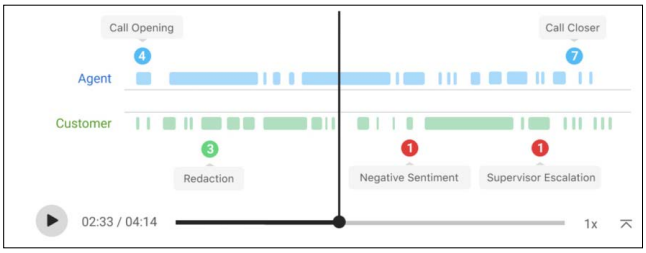

Figure 1: A Sample of Speech Analytics on A Running Call

Beyond call centers, ASR technology is also integrated into our daily lives. Virtual assistants (like Siri, Amazon Alexa, Google Assistant, and Microsoft Cortana), transcription services, dictation software solutions, language learning applications, and in-car speech recognition systems are just a few examples where ASR technology is used across industries. All that accounts for the rapid growth of speech recognition technology solutions, with the global speech and voice recognition market projected to grow from $12.62 billion in 2023 to $59.62 billion by 2030.

Methodology

Speech-to-text models to transcribe vast amounts of audio data, extracting valuable insights about customer queries, feedback, and complaints. By utilizing AI, call centers can not only improve

their data capture and analysis to elevate customer service, but also reduce the time and cost associated with manual call reviews. In this section we will outline the mechanics of speech analytics, illustrating its significance and offering guidance on its application in the call center environment.

- Speaker Diarization Speaker diarization is the process of distinguishing and segmenting individual voices within a multi-speaker audio recording. It involves steps like speech detection, segmentation, and clustering, turning a jumble of voices into distinct, labeled segments. Speaker diarization aims to answer the question of "Who spoke when". In short: diariziation algorithms break down an audio stream of multiple speakers into segments corresponding to the individual speakers.

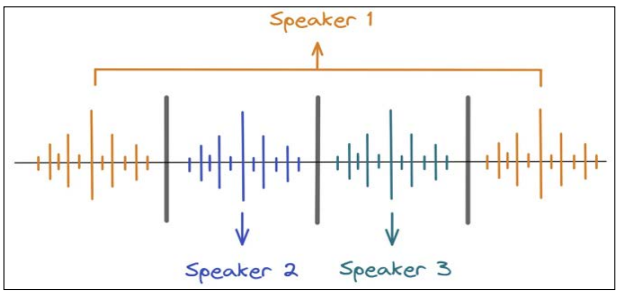

Figure 2: Speaker Diarization

By combining the information that we get from diarization with ASR transcriptions, we can transform the generated transcript into a format which is more readable and interpretable for humans and that can be used for other downstream NLP tasks.

- Word-Level Timestamps Word-level timestamps provide a precise time marker for each word within an audio transcription. This means that alongside the transcribed text, there's an exact record of when each word was spoken in the audio file. Whether it's for quality assurance, training, or dispute resolution, word-level timestamps ensure that pinpointing key moments in customer-agent interactions is hassle-free.
- Translation After a call concludes, it can be transcribed and then translated into 100+ languages spoken in Asia, Europe and the Middle Eastern. This is particularly useful for review, training, or when sharing call details with teams or departments that operate in different languages.
- Summarization AI summarization streamlines the workflow in contact centers by simplifying complex or lengthy calls into a concise summary. This technology empowers agents by providing them with the essential information from a conversation without the need to re- view the entire interaction. With this concise overview, agents can respond more efficiently, leading to quicker resolutions, improved customer satisfaction, and heightened productivity.
- Sentiment Analysis Using sentiment analysis, call centers can gauge the mood from customers' audio feedback, labeling it as positive, neutral, or negative. This insight helps match customers with the right agents for their needs, making operations smoother and improving customer interactions. Additionally, this data highlights top-performing agents, guiding best practices and pinpointing areas for coaching.
- Number Formatting The number formatting feature automates the process of converting numbers into a consistent format within text. It can automatically change numbers into either their written word form (e.g., "five" instead of "5") or their numerical form (e.g., "5" instead of "five"). This automation ensures that all numbers in the text follow the same format, making subsequent data retrieval and analysis more straightforward.
- Language Detection Operating in a call center using multiple languages your audio file can contain different languages. VoiceAI's automatic language detection eliminates the process of manually tagging languages for each audio file.
- Noise Cancellation Noise cancellation filters out background noises from an audio signal, ensuring only the primary voice or speech is captured and transcribed. Call centers often operate in bustling environments. Without noise cancellation, the background chatter, ringing phones, or even typing sounds can interfere with the transcription's accuracy. Removing these noises ensures the transcription software captures only the relevant conversation be- tween the agent and the customer.
- Punctuation As agents engage in myriad conversations daily, the absence of appropriate punctuation can make transcribed text difficult to follow, potentially leading to misinterpretations of the customer's intent or sentiment. Accurate punctuation not only demarcates the end of one thought and the beginning of another but also aids in capturing the true emotion and emphasis behind a speaker's words. For instance, the difference between a statement and a question — denoted simply by a period or a question mark — can change the entire context of a customer's query or feedback.

Conclusion

The application of ASR in VoIP call center solutions is vast, truly revolutionizing the way call centers process voice interactions, analyze and leverage voice data, and deliver customer support. By transforming audio data into meaningful information with ASR contact center solutions, you can streamline your call center operations, reduce costs by automating tasks, improve the experience for both customers and agents, monitor agent performance, automate quality assurance, and maintain compliance [1-14].

References

1. J Allan (2002) Perspectives on information retrieval and speech. In Information Retrieval Tech niques for Speech Applications 1-10.
2. S Busemann, S Schmeier, RG Arens (2000) Message classification in the call center. In Proceedings of the sixth conference on Applied natural language processing 158-165.
3. D Carmel, E Amitay, M Herscovici, YS Maarek, Y Petruschka, et al. (2001) Juru at trec 10 - experiments with index pruning. In TREC.
4. D Carmel, M Shtalhaim, A Soffer (2000) eResponder: Electronic Question Responder. In CooplS '00: Proceedings of the 7th International Conference on Cooperative Information Systems. Springer-Verlag 150-161.
5. J Chu-Carroll, B Carpenter (1999) Vector-based natural language call routing. Comput. Linguist 25: 361-388.
6. D Ferrucci, A Lally (2004) UIMA: an architectural approach to unstructured information processing in the corporate research environment. Natural Language Engineering 10: 476–489.
7. A Kilgarriff (2001) Comparing corpora. International Journal of Corpus Linguistics 6: 1-37.
8. B Kingsbury, L Mangu, G Saon, G Zweig, S Axelrod, et al. (2003) Towards domain independent conversational speech recognition. In Eurospeech, Geneva, Switzerland.
9. J Kleinberg (2002) Bursty and hierarchical structure in streams. In KDD '02: Proceedings of the eighth ACM

SIGKDD international conference on Knowledge discovery and data mining. ACM Press 91-101.

10. L Kosseim, S Beauregard, G Lapalme (2001) Using information extraction and natural language generation to answer e-mail. Data & Knowledge Engineering 38: 85-100.
11. G Leech, P Rayson, A Wilson (2001) Word Frequencies in Written and Spoken English: based on the British National Corpus. Longman.
12. G Riccardi, A Gorin, A Ljolje, M Riley (1997) A spoken language system for automated call routing. In Proc. ICASSP '97, Munich, Germany 1143-1146.
13. B Schiffman (2002) Building a Resource for Evaluating the Importance of Sentences. In LREC02, Las Palmas, Spain http: //www.ibm.com/sof tware/pervasive/voice server.
14. Ashish Bansal (2019) Punctuation and Capitalization Restoration using Bi-LSTM Network. In ternational Journal of Science and Research (IJSR) 8: 2020-2025.