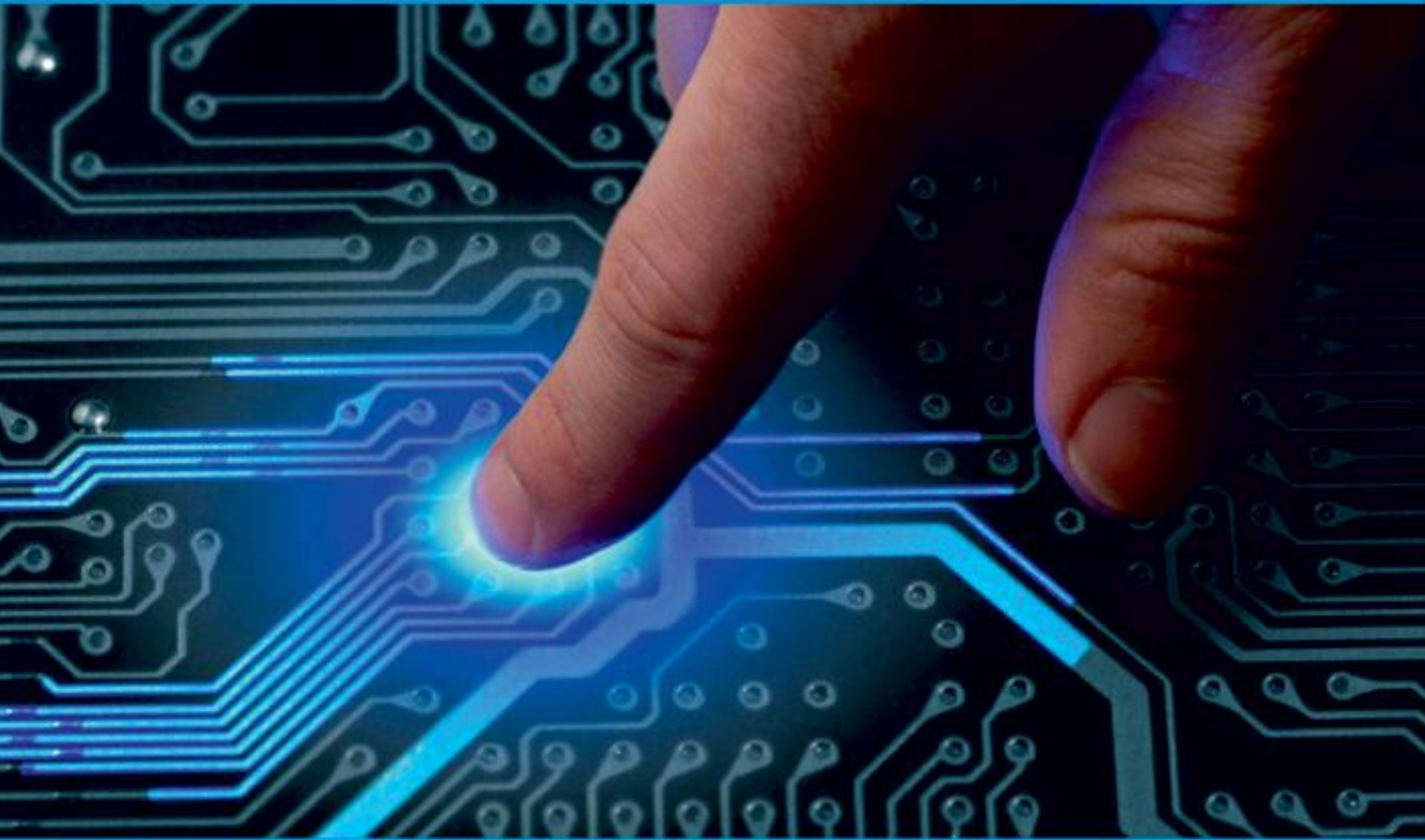
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907147 |

An Approach to Voice based Analytics for Sentiment Analysis with FreeSwitch

Abhilash L Bhat, Veena V

Assistant Professor, Dept of ISE, R.R. Institute of Technology, Bengaluru, India

Assistant Professor, Dept of CSE, R.R. Institute of Technology, Bengaluru, India

ABSTRACT: For most retail businesses, customer interaction via call centers is a very significant communication channel. Organizations typically receive thousands of customer calls every day. If the audio data thus collected can be aggregated and analyzed, it can yield quality insights into customer expectations, preferences, service issues & product usage. While speech analytics is not a new technology to the market, most of the business executives are still sceptical about the value it can add. Speech analytics is a powerful tool for analyzing recorded calls, structuring customer interactions and gaining insight in the hidden information. It can be used for audio mining, speech categorization, intelligence extraction, decision making, monitoring agent performance. In our product we are making simple and advanced analytics by converting voice to text and using that converted text we can analyse the customer responses towards the company which helps the companies to serve their customers better.

KEY WORDS: FreeSwitch, Speech Analysis, Speech to Text Conversion, Sentiment Analysis

I. INTRODUCTION

For most of the retail organizations, client interaction through call centers is an extremely huge communication channel. Associations commonly get a huge number of client calls each day. In the event that the sound information in this manner gathered can be collected and examined, it can yield quality bits of knowledge into client desires, satisfaction, benefit issues and product use. While voice based analytics is not another innovation to the market, the majority of the business administrators are as yet suspicious about the esteem it can include. Speech analytics is a powerful tool for analyzing recorded calls, structuring client interactions and gaining insight within the hidden info. It is often used for audio mining, speech categorization, intelligence extraction, to make decisions and controlling agent performance.

If applied properly and used effectively speech analytics will facilitate improve service quality, scale back operative expenses, boost revenue, and scale back client attrition. If integrated well with overall strategy it will facilitate businesses drive product and method innovation resulting in vital market differentiation. However there are vital challenges in remodelling speech knowledge to a structured type which might be subjected to additional analysis.
Speech analytics is becoming to be a revolutionary approach in the measurement of the customers emotions, context and therefore the intent. Business success depends heavily on the client expertise then enhancing this expertise is crucial for fulfilment of any business. Hearing the voice of the client presents a challenge to even the foremost refined contact center. Many various measurements are proposed to verify and appraise the service quality of client interactions like disconnection rates, holding times or reaction times. However these measurements tell regarding events inside client interactions rather than the explanation why they occurred.

With the arrival of technology, quantity of data keeps increasing up which is hard to research large amount of information to provide some purposeful information. Several analytical techniques and tools are developed by organizations to research and acquire business insights from information. Voice analytics is one such branch of analytics that helps varied organizations by analyzing spoken words or speech between two or additional individuals. It's one of the quickest growing technologies as a result of it helps in analyzing the emotions of individuals throughout the decision which may be helpful in characteristic their satisfaction towards a product or service.

Speech Analytics uses different analytic methods and mathematical algorithms to recorded audio conversations of call center interactions. There are different steps in Speech analytics. First, all the recorded conversation is converted to text

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907147 |

by the method of speech to text technology. Then various analytics is performed on the converted text in order to discover the root causes and hidden insights in the information.

The results of Speech analytics can be used in graphics, reports and alerts. Users also can run queries associated with their business wants so as to capture insights concerning many contact center problems like agent performance, client satisfaction, perennial calls, cancellation requests etc.

Speech analytics is a subtle software system technology designed to reinforce contact center potency and profit by extracting important data from each incoming decision. All speech analytics tools realize key words and phrases in conversations between agents and customers. Real time speech analytics solutions that analyze phone calls as they are on-going will alert the supervisor to intervene or prompt the agent to do a special approach, switch to a special script, or try a particular up-sell.

Challenges in analyzing speech data

- The speakers speaks differently due to gender, age, dialects, physical attributes (such as vocal tract). Any speech recognition system must take care of these options into thought. As an example "service officer" is also recognized as "serve his provide her".
- Humans along with speech also interacts via facial expressions, emotions, postures and eye movements, and these are not recognized by an automatic recognition system (ASR).
- While interacting in the real time environment humans encounter lot of unwanted sounds called noise and these need to be removed from the speech signals.
- Homophones(i.e. words that are pronounced same but have different meaning. For e.g. two and too) and word boundary ambiguities causes major problem to speech recognition systems.
- Grammatically spoken languages are completely different from written languages.

A. FREESWITCH

FreeSWITCH is an open source, scalable, cross-platform telephonic platform which is developed to route and interconnect the major communication protocols by using audio, video, text or any other type of media. It was designed0 to fill the space left by proprietary commercial solutions. By using FreeSwitch, many applications can be built by using wide range of freely available tools as FreeSwitch provides0 a stable telephonic platform.
In FreeSWITCH, call control and IVR functionality can be monitored by Application Programming Interfaces0 and these applications can be written by using some of the programming languages such as C language, C++, Python, Perl, Lua, JavaScript, Java and Microsoft .NET via Microsoft's CLR or via Mono.
The Call control applications of FreeSwitch uses the Event Socket, which is an Internet socket-based communications facility in FreeSwitch which provides a language independent interface. The Event Socket Library0 (ESL) and the ESL-wrappers of FreeSWITCH are available in Erlang, JavaScript, Lua, Perl, PHP, Python and Ruby.

B. Steps in Speech Analysis
Below diagram depicts the common steps in Speech Analysis process.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907147 |

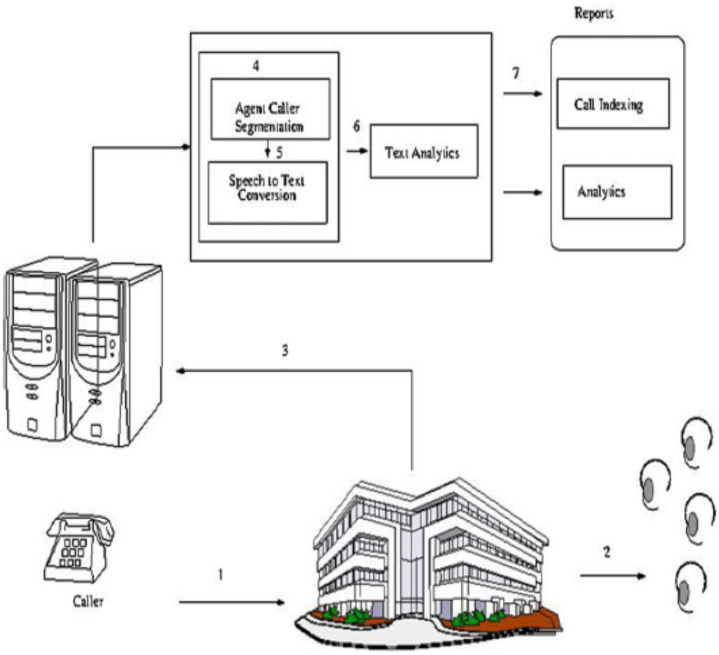


Fig 1. Steps in Speech Analysis

Speech Analysis has the following steps.

Step 1: A user makes the call which lands on the private branch exchange (PBX) of the call center. Make sure that the customers can call from any of the gadget such as landline, mobile, VoIP) and they can call from any place (office, home, street).
Step 2: After waiting in the queue for some time, the call connects on the executives desk and the customer starts interacting with the executive.
Step 3: While the interaction is on, the call between the executive and the customer is being recorded along with some information such as the number from where the0 call is connected, the time, etc. on the server. These recorded calls are need to be analyzed to gather the details.
Step 4.1: The recorded conversation is first processed to differentiate the hold music and voice of the customers.
Step 4.2: The voice portion of the customer and the executive is then separated into speech spoken by the executive and that speech spoken by the customer.
Thus before converting the audio into text, the two essential steps are carried namely that of distinguishing the hold music from voice and also the separation of the customer and the agent spoken speech.
Step 5: This speech segments are passed separately through a voice to text converter, which is an automatic speech recognition engine.
Step 6: The interaction between the customer and the executive which is converted text is processed by using text analysis tools to derive the required information from the converted text.
Step 7: Gather the information present in the conversation which can be used for later search or to generate the reports.

II. LITERATURE SURVEY

In [1] Automatic voice-controlled systems have changed the way humans interact with a computer. Voice or speech recognition systems allow a user to make a hands-free request to the computer, which in turn processes the request and serves the user with appropriate responses. After years of research and developments in machine learning and artificial intelligence, today voice-controlled technologies have become more efficient and are widely applied in many domains to enable and improve human-to-human and human-to-computer interactions. The state-of-the-art e-commerce applications with the help of web technologies offer interactive and user-friendly interfaces. However, there are some instances where people, especially with visual disabilities, are not able to fully experience the serviceability of such applications. A voice-controlled system embedded in a web application can enhance user experience and can provide voice as a means to control the functionality of e-commerce websites. In this paper, we propose a taxonomy of speech recognition systems (SRS) and present a voice-controlled commodity purchase e-commerce application using IBM

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907147 |

Watson speech-to-text to demonstrate its usability. The prototype can be extended to other application scenarios such as government service kiosks and enable analytics of the converted text data for scenarios such as medical diagnosis at the clinics.
In [2] Speech analytics utilizes speech recognition, predictive analytics, and authentication of the data streams while assessing customers' complaints in real-time. Assessment occurs through the collection and analysis of current data mixed with historical facts to determine patterns and to predict trends. In the current research, the authors have chosen to focus primarily on speech analytics, serving as an umbrella term encompassing speech analytics, audio-mining technologies. The use of speech analytics typically refers to a broader range of speech products, such as analyzing voice identification, emotion detection, and phonetics/speech analysis. Speech Analytics for Actionable Insights proceeds with the discussion of an overview of enterprise needs for speech analytics, a brief history of the speech recognition, the infrastructure of phonetic versus transcription approaches and real-time versus post-call solutions, major speech analytics vendors and their features, applications found within case studies, and recommendations and guidance. The primary goal of this monograph is to help business decision-makers educate themselves on the burgeoning field of speech analytics as well as to understand how it impacts the broader enterprise landscape.

In [3] Voice conversion (VC) is a task that alters the voice of a person to suit different styles while conserving the linguistic content. Previous state-of-the-art technology used in VC was based on the sequence-to-sequence (seq2seq) model, which could lose linguistic information. There was an attempt to overcome this problem using textual supervision; however, this required explicit alignment, and therefore the benefit of using seq2seq model was lost. In this study, a voice converter that utilizes multitask learning with text-to-speech (TTS) is presented. By using multitask learning, VC is expected to capture linguistic information and preserve the training stability. This method does not require explicit alignment for capturing abundant text information. Experiments on VC were performed on a male-Korean-emotional-text-speech dataset to convert the neutral voice to emotional voice. It was shown that multitask learning helps to preserve the linguistic content.

In [4] propose two approaches for speech recognition via supervised and unsupervised learning. Speech signals are non-stationary signals. Treating speech in computing domain falls under sequential learning task i.e. if we want to make a sense of current statement, we may need to go through the context in which it was spoken. Recurrent Neural Networks (RNN) is been used in speech recognition problems because of its powerful sequence modeling capacity. In this paper we have proposed Bi-directional Recurrent Neural Network with Long Short Term Memory model (LSTM), so that speech signal reconstruction can be done in a proper way without performance loss. For unsupervised learning, model is designed on the basis of Restricted Boltzmann Machine (RBM) which generates a reconstruction based output and helps in conversion of voice into text, each letter by letter.

In [5] Accessing information during a call without having to open up a web browser in a phone is not only easier but also efficient in time-constraint situations. To achieve this, we have implemented an Interactive Voice Response System (IVRS) for SIP-based phones. Telecommunication services plays a vital role in exchange of information. Although this system is limited to SIP-based phones, we propose a general IVRS that can be implemented for a regular network operator.

In [6] This paper deals with Voice over Internet Protocol (VoIP) in the environment of University of Žilina. There was VoIP solution based on SIP proxy Kamailio in addition with FreeSWITCH and RTPProxy. This solution was qite hard to maintain and administer. After longer discussion we choose Asterisk PBX as replacement for all three softwares. We will show current state and our path to new solution.

In [7] Customer service in e-commerce businesses is one of the most essential aspects for business continuity, and a good communication service framework not only has a great influence in customer service staff's work efficiency, but can also reduce the operational cost of e-commerce enterprises. This paper proposes a communication service framework, called ACS-Communicator, to facilitate the communications in customer service. ACS-Communicator is a set of Internet distributed cluster architecture. It combines communication systems and customer service workbenches together, providing an integrated production solution for customer service. Based on the LVS technology, we separate the framework into media flow and signaling management, so that the system networking becomes more simple in capacity expansion. Our proposed framework is a software defined architecture, hence reducing the overall cost for businesses. We evaluate ACS-Communicator in a large call center within Alibaba Group, which is one of the largest call centers in the world. The experimental results show that ACS-Communicator can cover 99.99% communication

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907147 |

service groups within Alibaba, can support over 500 enterprise users to provide communication services, can achieve a total call time of over 400,000 hours per day, and can achieve a high overload of 450 calls per second, when using the same computational machine resources as the existing framework (FreeSWITCH).

III. METHODOLOGY

A. System Architecture

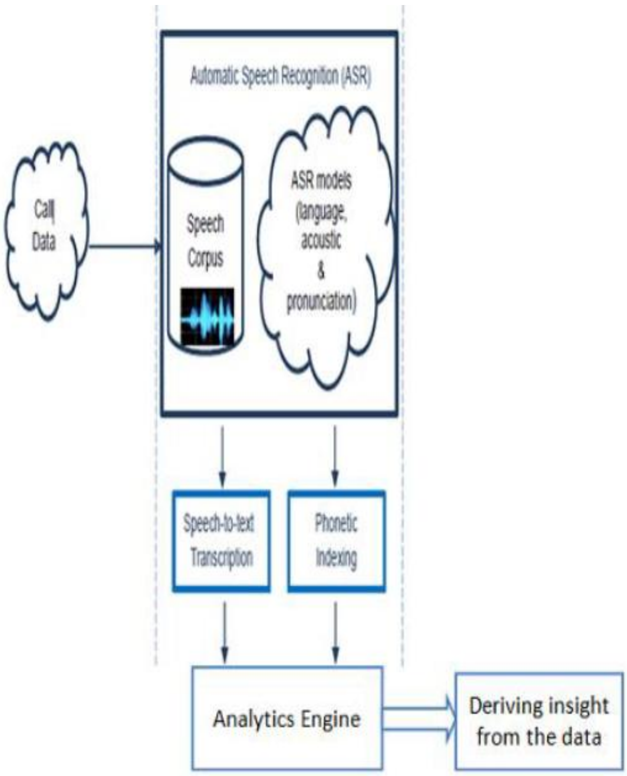


Fig 2. System Architecture

Speech Corpus: A speech corpusis a database that contains speech audio files and text transcriptions.
ASR models: The three models of Automatic Speech Recognition are Acoustic model, Language model and Pronunciation model.
Speech to Text Conversion process: Speech to Text is a method which converts the audios or recordings or the recorded conversations into textual format in order to analyze those recordings or conversations in order to derive hidden insights in those data.
By analysing those converted textual data, it helps the companies or the organizations to understand the satisfaction level of their customers about their service or the products which allows them to serve their customers better.

Analytics Engine: Analytics Engine is an engine which takes text as input and performs analysis on that textual data in order to determine the satisfaction of the customers towards the organization.

Text analysis which is also referred as text mining is the method of deriving the high quality information from the texts. Text analysis usually involves structuring the input text which is followed by deriving the patterns within the structured data and the evaluation and interpretation of the result.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907147 |

B. System Design

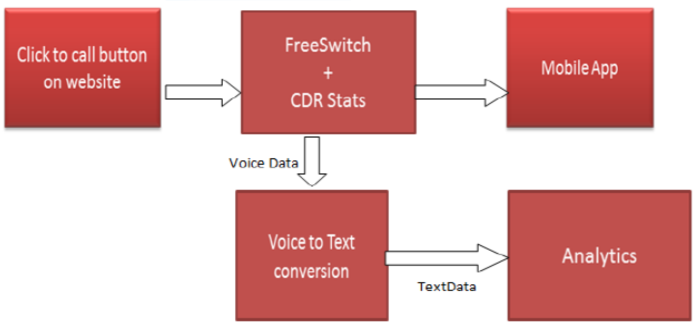


Fig 3. System Design

- An Online Call customizable button will be present in the Website.
- All Customer support executives/ Sales executives or other concerned persons will be having a mobile app.
- Whenever Customer calls from the Website, Call will be routed and distributed to the executives.
- Calls can be received in Mobile App by the executives and the customer can interact with executives.
- The conversation between the customer and the executives is recorded.
- The audio or speech is converted into text format in order to do analytics.
- Analytics is performed on the data to derive the insights from the data which helps to understand the customer interaction towards the company.

IV. TESTING AND RESULTS

As explained above, Testing can be performed by collecting the voice data from the customers and converting that voice into text and then any Machine Learning algorithm like Naïve Bayes can be applied in order to find whether the customer satisfaction is either positive, negative or neutral.

V. CONCLUSION AND FUTURE ENHANCEMENT

Speech analytics is the method of analyzing the recorded conversations in order to collect the customer information in order to improve communication and future interactions. The process is mainly used by customer contact centers to extract the information which is buried in client interactions with an organization.

In our product we are making simple and advanced analytics by converting voice to text and using that converted text we can analyse the customer responses towards the company which helps the companies to serve their customers better. Speech analytics can provide a solution by collecting the voice data and provide insights on the interactions that happens between the customer and the executives. Speech analytics can prove to be a revolutionary approach in measuring the customer‟s emotions, context and the intent. Business success depends heavily on the customer experience and so enhancing this experience is critical for the success of any business.

As a part of future enhancements, we are mainly concentrating on the following features.

Speaker Separation: It's not enough to know what was said during customer calls. It is also important to know who said it. Separation of two speakers on a single audio channel into two virtual channels allows their speech to be analyzed discretely.
Emotion Detection: Identification of the emotional state of customers by analyzing their voices for variations in pitch or tone.
Talk-Over Analysis: Highlighting those moments when customers and the executive of the organization are talking simultaneously, which indicates the customer dissatisfaction. When neither of them are talking for long duration of time, indicates that the executive is not having the sufficient knowledge about the details of the product or the customer is talking about.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907147 |

REFERENCES

1. Mandeep Singh Kandhari; Farhana Zulkemine; HarunaIsah, "A Voice Controlled E-Commerce Web Application",2018
2. J. P. Shim; Aaron M. French; BikeshUpreti, "Speech Analytics for Actionable Insights: Current Status, Recommendation, and Guidance"
3. Tae-HoKim;SungjaeCho;ShinkookChoi; SejikPark;Soo-Young Lee, "Emotional Voice Conversion Using Multitask Learning with Text-To-Speech ",2020
4. Arpita Gupta;Akshay Joshi, " Speech Recognition Using Artificial Neural Network", 2018
5. Abdullah Mohammad Ansari; Md. Faisal Nehal; Mohammed Abdul Qadeer, "SIP-based Interactive Voice Response System using FreeSwitch EPBX",2013
6. Marek Moravcik; Martin Kontsek "Proposal of VoIP infrastructure and services for academia - case study", 2019
7. ShuliWang; ZhaorongNi; GuobaoYou; JunHuang;XidongRen;Min Fu, "ACS-Communicator :a Low-Cost Customer Service Oriented Communication Service Framework for E-Commerce Businesses", 2020

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details