Transition from Observation To Knowledge To Intelligence (TOKI)

Editors

Dr. Victor ODUMUYIWA, Dr. Olufade ONIFADE,
Prof. Amos DAVID & Prof. Charles UWADIA

Victor ODUMUYIWA
Department of Computer Sciences,
University of Lagos
Nigeria

ISBN: 978-978-976-000-8

The responsibility for opinions expressed in articles, studies and other contributions in this proceeding rests solely with their authors.

Transition from Observation to Knowledge to Intelligence

3rd Biennial International Conference on Transition from Observation to Knowledge to Intelligence (TOKI)
15-16 August 2019
University of Lagos, Nigeria

Editors

Dr. Victor ODUMUYIWA
Dr. Olufade ONIFADE
Prof. Amos DAVID
Prof. Charles UWADIA

Design and Implementation of a Speech-to-text Converter for Integration with Customer Relationship Management Applications

ADEBISI John Adetunji[1]
ABDULSALAM Khadeejah Adebisi[1]
AKHIGBE Bernard Ijesunor[2]
AFOLABI Babajide Samuel[2]
[1]Department of Electrical Electronics and Computer Engineering
University of Lagos, Akoka, Lagos
[2]Department of Computer Science and Engineering
Obafemi Awolowo University, Ile-Ife. Nigeria.

Abstract: A proficient speech-to-text converter for integration with Customer Relationship Management (CRM) applications is presented in this work for knowledge management. The leading motive is to articulate a system which would provide leading performance in terms of complication, correctness and storage requirements for enterprise environment. Customer service agents in the enterprise make and receive calls periodically, such calls are being recorded and play back in future for reference in case of complaints, however the need for the respective text of the recorded audio in real-time cannot be over emphasized hence an integratable speech-to-text converter. The converter was designed using voice recording activity detection algorithm. The conventional recording activity detection algorithms which track only voice cannot successfully identify potential speech from input, convert to text and integrate such into the enterprise due to the unwanted parts of the speech that may appear to be speech. In this system, the speech-to-text conversion module uses sound waveform from an amplifier and changed over into a succession of fixed estimate acoustic vectors in a procedure called include extraction to differentiate noise from speech, thus improved accuracy. The system is a web application that uses Visual Studio Code for the integration and Node Package Manager (NPM) as a means for managing and installing all dependencies used. The implementation uses Node.js, JavaScript and Vue.js for the user interface. A database containing all recorded and converted speech of the user is adequate to give recognition accuracy of approximately 95% for trained users and 82% for non-trained users. Thus, this system entertains insight of real time speaker integrated into enterprise applications like CRM, self-care portals etc.

Keywords: Speech-to-text Converter, feature extraction, customer relationship management, knowledge management, Integration.

1. Introduction

Speech is the essential, normal and productive type of specialized strategy for individuals for collaboration purposes. Speech recognition is an innovation that empowers a computer to catch words spoken by people with an assistance of receiver (microphone) (Abbasi *et al.*, 2010). Today, speech advancements are usually accessible for a restricted yet fascinating scope of the undertaking. These advancements empower machines to react accurately and dependably to human voices and give helpful and significant administrations. As speaking with a computer is quicker utilizing voice instead of utilizing console or manual input, individuals will incline toward such framework (AbuZeina *et al.*, 2012). Communication amongst people is overwhelmed by spoken dialect, thusly it is normal for individuals to expect voice interfaces with Computers (Bijl, and Hyde-Thomson, 2001). Speech Recognition is between disciplinary sub-field of computational phonetics that create approaches and advances to empower the acknowledgement and interpretation of "spoken words into text" by personal computers. It fuses learning and research in the semantics, software engineering, and electrical building fields (Weber, 2003). Speech recognition includes hardware alongside software that acts together to discernably identify human speech and make an interpretation of the identified word into a series of text (Weber, 2002). Speech Recognition works by separating sounds called phonemes. Phonemes are unmistakable units of sounds.

For instance, "those" is comprised of three phonemes; the first is "the" sound, the second is the "o" sound and the third is the "s" sound (Dia *et al*., 2012). The speech Recognition programming endeavour's to coordinate the recognized phonemes with known words from a putaway lexicon. From the innovation point of view, Speech Recognition has a long history with a few influxes of real developments. As of late, the investigation of Speech Recognition has profited and progressed on an expansive scale, Artificial Intelligence (AI) and machine learning are being utilized to prepare models that could recognize client spoken words quicker and even comprehend the setting in which they are talked. Speech industry players incorporate

Google, Microsoft, International Business Machine corporation (IBM), Baidu, Apple, Amazon, SoundHound, a significant number of which have broadcasted the centre innovation in their speech recognition frameworks as being founded on deep learning (Benesty, *et al.,* 2007).

It is imperative to take note of the expressions "Speech recognition" and "Voice recognition" are here and there utilized conversely. In any case, the two terms mean distinctive things. Speech recognition is utilized to distinguish words in spoken dialect while voice recognition can be characterized as a biometric innovation used to recognize a specific person's voice or for speaker ID. (Goldberg, 2016). The capacity to change over speech signals into text has numerous applications, these applications fluctuate from the therapeutic field to even military rocket direction framework. Discourse to content change is the fundamental thought behind most present-day advancements, e.g. "Chabot's", "Home Automation System", "Route Systems", "Voice Control Systems" and "Individual Assistants".

Speech to text change in this century is generally utilized in close to home associates as they get sound information, the framework parses the sound and after that gives a content input to the administrator. The procedure of speech to text integration feature into enterprise empowers clients through administrators to understudy the capacity to specifically get notes by simply putting the application toward a path where the voice signals can be picked straightforwardly and after that changed over to text continuously.

2. Speech and Vocabulary

Insight into various studies conducted by outstanding researchers on the implementation of speech to text converter basing it on the principle of speech recognition. It presents their observed problem with a speech to text conversion and the methods used in solving these problems. A perfect circumstance during the speech recognition process is that a speech recognition motor perceives, which all words articulated by a human in any case, basically, the execution of a speech recognition motor relies upon a number of components. Vocabularies, various clients and uproarious condition are the main considerations that are considered in the depending factors for a speech recognition motor.

2.1. Speech Recognition

Speech recognition systems can be divided into the range of classes based totally on their capacity to recognize that words and listing of words they have. A few classes of speech cognizance are categorized as under:

Isolated speech: isolated words generally contain a pause between two utterances. It doesn't imply that it only accepts a single phrase however rather it requires one utterance at a time.

Connected speech: connected phrases or related speech is comparable to isolated speech but permit separate utterances with a minimal pause between them.

Continuous speech: continuous speech permits the user to speak almost naturally, it is additionally referred to as the computer dictation. *Spontaneous speech:* at a basic level, it can be thought of as speech that is natural sounding and not rehearsed.

2.2. Vocabulary

The vocabulary size of speech recognition framework influences the handling necessities, precision and multifaceted nature of the framework. In voice Recognition framework: speech-to-text the kinds of vocabularies can be named pursues: *Little vocabulary:* single letter. *Medium vocabulary:* a few letter words and *Huge vocabulary:* more letter words. The work uses little and medium vocabularies which is broadly divided into two main categories based on speaker models namely speaker dependent and speaker independent. There are three ways to deal with speaker autonomy. The primary methodology is to utilize learning building procedures to discover perceptually spurred discourse parameters that are moderately invariant between speakers. The avocation for this methodology is that if a specialist spectrogram per user can read spectrograms with high precision, it ought to be conceivable to locate the invariant parameters (Robert, 2013).

2.3. Speech Acknowledgment System

With the assistance of mouthpiece, there are some components used for speech acknowledgement systems such as the voice input with sound as a major contribution to the framework, the PCs sound card creates the proportional advanced portrayal of got sound. Others include *Digitization***,** *Acoustic Model, Language Model* and *Speech Motor,* which are being used for input signal and other processing in the speech recognition procedure depicted in Figure 1.

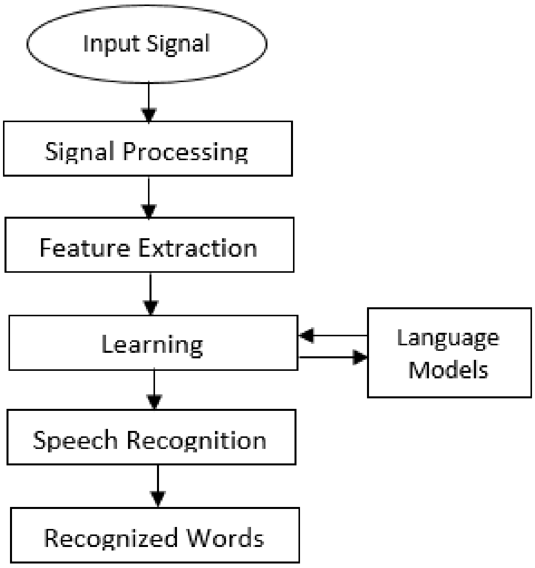


Figure 1: Speech Recognition Procedure (Source: Aşlıyan, 2008)

2.3.1. Feature Extraction

The primary objective of extraction is to process a closefisted succession of highlight vectors giving a conservative portrayal of the given information flag. The element extraction is generally performed in three phases. The main stage is known as the speech examination or the acoustic front end. It plays out some sort of range transient investigation of the flag and creates crude highlights portraying the envelope of the power range of short speech interims. This is critical to speech recognition using the Mel-recurrence Cepstrum Coefficient (MFCC).

The MFCC depend on the known variety of the human ear's basic transmission capacity frequencies with channels separated straightly at low frequencies and logarithmically at high frequencies used to catch the vital qualities of speech. It indicates a human impression of the recurrence substance of sounds for speech signals does not pursue a direct scale. In this manner for each tone with a genuine recurrence, f, estimated in Hz, an abstract pitch is estimated on a scale called the Mel scale (Abushariah, 2012). The Mel-recurrence scale is straight recurrence separating beneath 1000 Hz and a logarithmic dispersing over 1000 Hz. As a source of perspective point, the pitch of a 1 kHz tone, 40 dB over the perceptual hearing edge, is characterized as 1000 Mel's.

2.3.2. Pattern Recognition

This methodology utilizes all around figured scientific structure and sets up predictable speech design portrayals, for solid example correlation, from an arrangement of named preparing tests by means of a formal preparing calculation. Pattern recognition is of paramount importance in this work as speech design portrayal can be as a speech layout or a measurable model and can be connected to a sound (littler than a word), a word, or an expression (Furui *et al.,* 2004).

3. Review of Literature

Weber (2003) worked on a platform that utilizes regular dialect preparing and speech handling in parallel. He proceeds to clarify his innovation and product usage. This framework accept voice input, it contrasts it and two sentence structure records as he called them and a characteristic dialect lexicon with no integration feature or relation with enterprise systems.

Lyberg (1998) clarified in his production the significance of speech to text transformation and rattled off various manner by which this could be performed. Among all the strategy used in his production, he put a lot of weight on utilizing the Hidden Markov Model (HMM) and how this model could be utilized to measurably decide an arrangement of the word from a lexicon of the word. Lyberg accentuated on how the model chipped away at shifts in term of execution as utilized by different speakers, he clarifies how deciding sentence pressure was an issue amid the examination and furthermore how it was difficult to accurately comprehend words that sounded alike as on account of words like "These", "this", "their" and "there". He additionally clarifies that the checking implies check the orthography and interpretation of the words in the speech. This distribution clarifies the significance of utilizing a several strategies for testing the model manufactured and further rattled off the greater part of the basic issue handled in a speech to text change.

The principle of speech recognition for speech to text change by Mitchell *et al.,* (1999) is basic to enhance correspondence with PCs. The work proved the ordinary method for contributing into the framework is not and cannot be disposed. However, utilizing voice input is a technique that ought to be investigated. Although, the work focus more on speech recognition frameworks; especially alluring for individuals wishing to utilize PCs who don't have console abilities. The framework can interface with a wide range of uses to permit the perceived text yield to specifically contribute to the application, e.g. a word processor. He proceeds to additionally clarify the restriction of this framework and routes by which it could be enhanced.

Ibrahim Patel *et al*., (2010), examined that recurrence ghastly data with Mel frequency is utilized to present as a methodology in the acknowledgement of speech for development of speech, in light of acknowledgement approach which is spoken to in HMM. A mix of frequenct ghostly data in the traditional Mel range which depends on the methodology of speech recognition. The methodology of Mel recurrence uses the recurrence perception in a speech inside given goals bringing about the covering of goals highlight which results in the point of confinement of recognition. In speech recognition framework which depends on HMM, goals disintegration is utilized with a mapping approach in an isolating recurrence. The consequence of the examination is that there is a change in quality measurements of speech recognition regarding the computational time and learning exactness in speech recognition framework.

Sharma and Haksar (2012) have spoken to recognition of speech in a more extensive manner. It alludes to the innovation that will perceive the speech without being focused at the single speaker. Inconstancy in speech design and speech recognition is the primary issue of concern in this work. Their speech recognition framework has the capacity of the basilar layer duplicated in the front-end of the channel bank. To acquire better acknowledgement the result was tested with band subdivision and it was nearer to the human recognition. The framework channel, which is developed for speech recognition is evaluated of clamour and clean speech but not related to the enterprise.

Reddy and Mahender (2013) in their diary on speech to text transformation utilizing android telephones to simplify how Mobile telephones have turned into a fundamental piece of our everyday life, causing higher requests for a substance that can be utilized on them. Advanced cells offer client improved strategies to associate with their telephones however, the most characteristic method for communication remains speech. Google gathered an expansive database of words from the day by day passages in the Google web indexes well as the digitalization of in excess of 10 million books in Google Book Search venture. The database storage contains over 230 billion words. In the event that we utilize this sort of speech recognizer, it is likely that our voice is put away on Google's servers. This reality gives a constant increment of information utilized for preparing, along these lines enhancing exactness of the framework.

Patel and Prasad (2013) chipped away at Speech Recognition and Verification Using MFCC and VQ. The objective of the venture was to make a speaker acknowledgement framework and apply it to a speech of an obscure speaker. By exploring the removed highlights of the obscure speech and after that contrast them with the put away extricated highlights for each unique speaker so as to recognize the obscure speaker. The element extraction is finished by utilizing Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) as order calculations. The blunder rate of the framework was around 13%. In the second frame.

Swamy and Ramakrishnan (2013) worked on an Efficient Speech Recognition System: The framework created utilizing distinctive strategies uwing Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ) and Hidden Markov Model (HMM). The outcome in the greater part of the cases returned a general productivity of 65%. The investigation additionally uncovers that the HMM calculation can distinguish the most ordinarily utilized disengaged word. Subsequently, the speech recognition framework accomplishes 68% proficiency.

Pinson *et al.*, (2010) clarified that a piece of the innovation includes extraction of speech flag spikes, or occasions, that show vital parts of the flag. This area additionally includes catching the transient connections between the occasions. In the by and by favoured innovation, a plan of weighted classifiers was utilized to extricate occasions. Another area of the development include building the plan of weighted classifiers for use in a programmed speech recognition motor, that piece of the innovation include recognizing successions of occasions rather than, or notwithstanding, identifying singular occasions. He at last, proceeds to clarify that another piece of the framework includes portioning a speech motion at perceptually essential areas. His innovation in rundown partitions the work on analysing every speech flag and after a specific word is broke down, it shouldn't be handled once more.

4. Design Methodology

This work designed an integratable speech-to-text converter using voice recording activity detection algorithm. The information sound waveform from an amplifier is changed over into a succession of fixed estimate acoustic vectors Y1:T = y1,...,yT (Gales and Young, 2008) in a procedure called include extraction. The decoder at that point endeavours to find the arrangement of words w1:L = w1,...,wL which is well on the way to have produced Y , i.e. the decoder attempts to find:

$$\hat{W} = \arg_{ue} \max \{P(w|Y)\} \qquad (1)$$

Since P(w|Y) is difficult to show directly, Bayes' Rule is utilized to change. The probability showing p(Y |w) is controlled by an acoustic model and the earlier P (w) is dictated by a dialect show. The essential unit of sound spoken to by the acoustic model is the telephone. For instance, "bat" is made of three telephones/b//ae//t/. Around 40 such telephones are required for English. The HMM Acoustic Models (Basic-Single Component) each talked word w is decayed into a grouping of Kw fundamental sounds called base telephones. This arrangement is called its elocution qx(1:wK) = q1,...,qKw . To take into account the likelihood of various elocutions, the probability p(Y |w) can be figured over numerous articulation where the summation is over all legitimate articulation groupings for w, Q is a specific succession of elocutions (Gales and Young, 2008). These state occupation probabilities, likewise called occupation tallies, speak to a delicate arrangement of the model states to the information and it is direct to demonstrate that the new arrangement of Gaussian parameters defined by:

$$p(Y\ |w) = \Sigma_q\ p(Y\ |Q)\ P(Q|w) \qquad (2)$$

The summation is over all valid pronunciation sequences for w, Q is a sequence of pronunciations.

$$P\ (Q|w) = \Pi\ P(q^{(wl)}|w_l) \qquad (3)$$

These state occupation probabilities also called occupation counts, represent a soft alignment of the model states to the data and it is straightforward to show that the new set of Gaussian parameters to maximizes the likelihood of the data given these alignments. A similar re-estimation equation can be derived for the transition probabilities. The design methodology addressed the identified challenges in line with the speech synthesis flow in Figure 2.

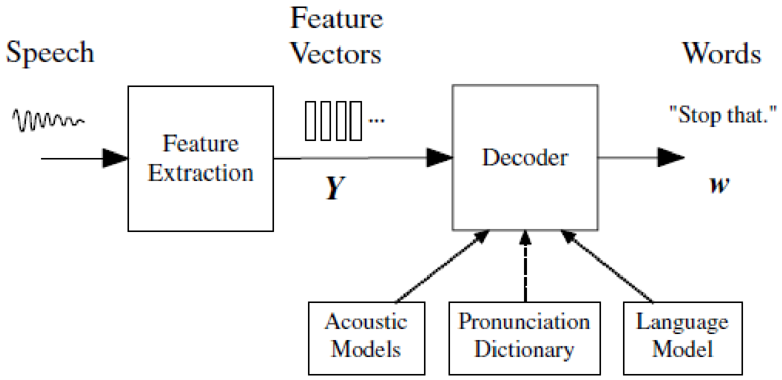


Figure 2: Architecture of a HMM-based Recogniser (Source: Gales and Young, 2008)

5. Database and Interface Creation

To simplify the training and testing of the integrated speech-to-text converter, speech Database (DB) is important. Varieties of speech mockups were obtained from diverse speakers to form the database. Two kinds of DBs were created, and the tools used include but not limited to Software Development Kits (SDKs), frameworks, JavaScript programming language and Application Programming Interfaces (APIs). The application core consists of the view model part and core functions that handles the transcribing of speech into text in a format that occur when the speech has been gotten. This section also handles the linking and integration part of the gotten textual output to the view which is being rendered with the HTML and integrated into the CRM. The tools and libraries used are;

- *Visual Studio Code:* This is an editor and was used to write the code.
- *Git*: This was used for version control.
- *SCSS*: This was the CSS preprocessor library used for the styling of the user interface.
- *Webpack*: This is a build process tool that allows be just confuse the entire application flow once and not worry about other configurations.
- *Node Package Manager (NPM):* This provides a means for managing and installing all dependencies that would be needed in the project.

- *Figma***:** This is the design tool used for the application design and prototyping.

6. System Implementation

The system implementation uses various technologies ranging from Node.js which was used as a server-side environment, written with JavaScript while Vue.js was used as a dynamic structure for the user interface. The effectiveness of the methodology used for the speech to text converter is proven during implementation. It involves all processes and effort that led to the success of this work. It is affiliated with all modules right from the naming convention to the system architecture and then the development stage to the completion of the work. The system consists of modules and components working together. The implementation based on the design such that each component supports a well-defined abstraction. The basic starter code base is as aspect of implementation where codes are built upon initializing the project with Git (Version Control Tool), it also keeps track of all changes made to the codebase and handle reversals to different version of the application. Figure 3 depicts implementation architecture/flow.

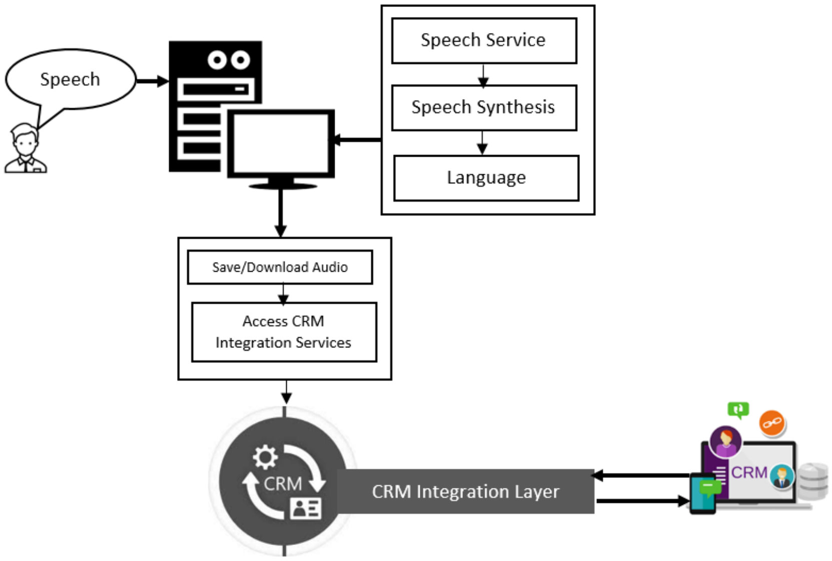
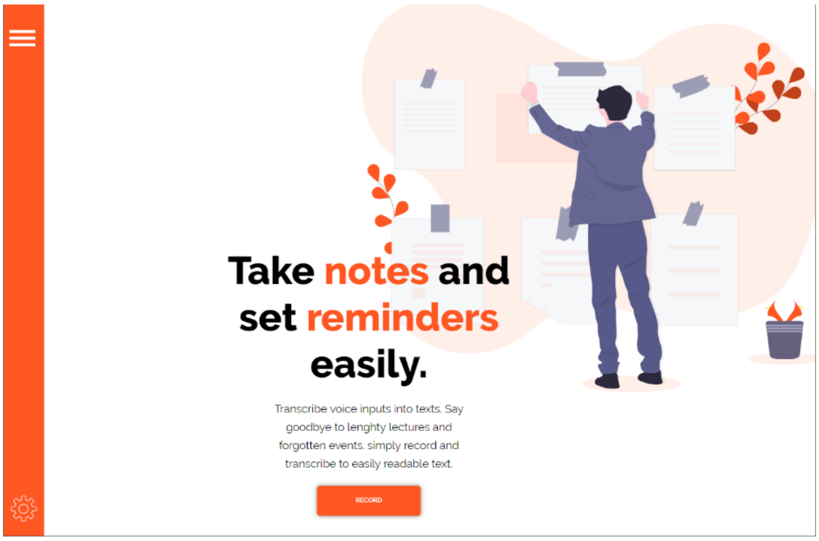


Figure 3: Implementation Architecture

The major implementation page is the recording, this page reliance on the flexible nature of vue came in handy as it provided the opportunity to use node commands within the renderer process therefore giving me access to important API. The audio file which was being recorded by the user and when this audio file has been gotten, would then be sent over to the speech synthesis service created to transcribe the gotten speech into text. This process involved (i) Successfully getting the audio from the user, (ii) Saving the audio file and (iii) Sending the saved audio file into the speech synthesizer. After which, the speech service/synthesizer returns a textual output of what it perceives the spoken words to be and passed same through the CRM discovery service for backup and retrieval purpose in future.

Figure 4 shows the user interface just before the start button is pressed. When the start button is presses, it triggers the switch State function and this function controls certain activities such as controlling and toggling of state of the record button, in addition checks for the current state of the button then determines whether or not to call the recordAudio() function or stopRecording() function. Figure 5 shows how the recording screen would be displayed just before recording. After the recording has been performed. The user gets the ability to download the audio to his machine, get the transcribed textual output in another window as well as successful integration the CRM Enterprise Application. Following the implementation of the core logic, unit testing was carried out on the system to ensure that all modules and features are working adequately.



Take notes and set reminders easily.

Transcribe voice inputs into texts. Say goodbye to lenghty lectures and forgotten events. simply record and transcribe to easily readable text.

Figure 4: Implementation welcome screen

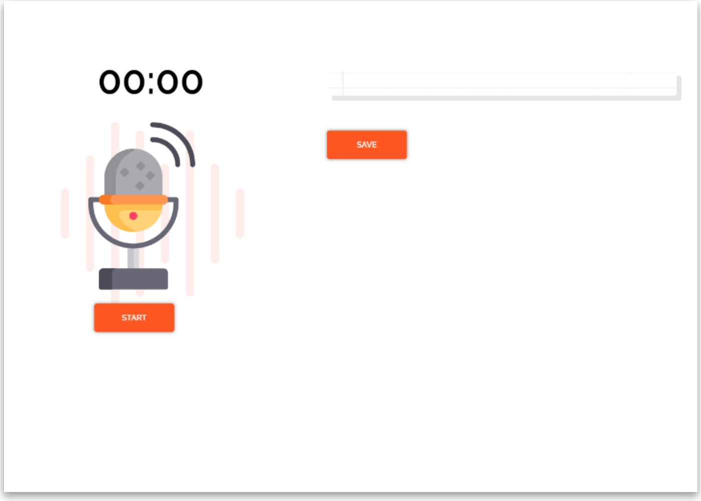


Figure 5: Implementation interface before recording

7. Discussion of Results

To perform the speech-to-text conversion implementation, speech samples in English language were obtained from 20 speakers (male and female). Common customer service interaction statements, namely, “good morning, my name is Sarah”, “how can I help you”, “what is your name?", "where are you calling from", "what can we do for you" were nominated. Four examples were taken for each statement. The sum of words in each statement is copious. Preparing the system for each word requires memory. Although memory requirement could be reduced by exploring language structure. The assessment of the spoken words and the recorded text was evaluated based on accuracy as shown in Table 7.0.

Table 1: Showing the evaluation results

English Speaker details	**Accuracy***
Trained users	95%
Non-trained users	82%

* Accuracy = Number of statements correctly converted to text

8. Conclusion

In this work, a speech to text conversion system integratable into CRM application is been designed using voice recording activity detection algorithm and implemented using visual studio code for the integration and Node Package Manager (NPM) as a means for managing and installing all dependencies used. The implementation uses various technologies ranging from Node.js used as a server-side environment, written with JavaScript while Vue.js was used as a dynamic structure for the user interface. The system has several beneficial features such as fast learning, malleable infrastructure size and robustness to speaker unpredictability (recognition of same words pronounced in different ways). The approach used promises to be a successful and powerful substitute to the straight speech-to-text converters.

The system designed in this work with the above outstanding structure is evidently a plus to enterprise applications especially the customer relationship management system. This implementation illustrates the budding of optimum configurations of enterprise application components. In theory, the accuracy improves with the expansion in training data. In light of this, improved memory is required. In future a more robust database could be formed carefully to help reduce memory requirements and increases conversion accuracy. The HMM language modeling can also be employed in other implementation to build an effective speech to text conversion system including better modeling accuracy, better context understanding, more natural insight and a more cost-effective use of parameters.

List of References

Abbasi J., Hussain M. and Ahmed S. (2010): An Implementation of Speech Recognition for Desktop Application. Unpublished B.Sc. thesis submitted to department of software engineering, Mehran University of Engineering and Technology, Jamshoro. Retrieved from https://www.scribd.com/doc/37504924/Speech-Recognition-MY-Final-Year-Project?. February 12, 2019; 11:30Pm.

AbuZeina, D., Al-Muhtaseb, H., and Elshafei, M. (2012). Cross-Word

Arabic Pronunciation Variation Modeling Using Part of Speech Tagging. In *Modern Speech Recognition Approaches with Case Studies*. IntechOpen.

Abushariah, M. A. A. M., Ainon, R. N., Zainuddin, R., Elshafei, M., and

Khalifa, O. O. (2012). Arabic speaker-independent continuous automatic speech recognition based on a phonetically rich and balanced speech corpus. Int. Arab J. Inf. Technol., 9(1), 84-93.

Aşlıyan, R. (2008). Design and implementation of Turkish speech

recognition engine (Doctoral dissertation, DEÜ Fen Bilimleri Enstitüsü).

Bennett, J. D., and Jarvis, L. M. (1999). U.S. Patent No. 5,884,256. Washington, DC: U.S. Patent and Trademark Office.

Benesty, J., Sondhi, M. M., and Huang, Y. (Eds.). (2007). Springer handbook of speech processing. Springer.

Bijl, D., & Hyde-Thomson, H. (2001). U.S. Patent No. 6,173,259. Washington, DC: U.S. Patent and Trademark Office.

Furui, S., Kikuchi, T., Shinnaka, Y., & Hori, C. (2004). Speech-to-text

and speech-to-speech summarization of spontaneous speech. *IEEE Transactions on Speech and Audio Processing*, *12*(4), 401-408.

Gales, M., & Young, S. (2008). The application of hidden Markov models in speech recognition. Foundations and Trends® in Signal Processing, 1(3), 195-304.

Lyberg, B. (1998). *U.S. Patent No. 5,826,234*. Washington, DC: U.S. Patent and Trademark Office.

Mitchell, J. C., Heard, A. J., Corbett, S. N., & Daniel, N. J. (1999). *U.S. Patent No. 5,857,099*. Washington, DC: U.S. Patent and Trademark Office.

Ksenia Shalonova (2007, May 9) "Automatic Speech recognition" Retrieved from

http://www.cs.bris.ac.uk/Teaching/Resources/COMS12303/lectures/Ksenia_Shalonova- Speech_Recognition.pdf

Pinson, M., Pinson, D., Flanagan, M., & Makanvand, S. (2010). *U.S. Patent Application No. 12/616,723*.

Patel, K., & Prasad, R. K. (2013). Speech recognition and verification using MFCC & VQ. *Int. J. Emerg. Sci. Eng.(IJESE)*, *1*(7).

Reddy, B. R., & Mahender, E. (2013). Speech to text conversion using android platform. *International Journal of Engineering Research and Applications (IJERA)*, *3*(1), 253-258.

Robert, E. Z., & Venson , S. (2013). "User Profile Based Speech to Text Conversion For Visual Voice Mail", United State Patent No US 8,358,752.

Sharma, K., & Haksar, P. (2012). Speech denoising using different types of filters. *International Journal of Engineering Research and Applications*, *2*(1), 809-811.

Sharma, M., & Sarma, K. K. (2016). Soft-Computational Techniques and Spectro-Temporal Features for Telephonic Speech Recognition: an overview and review of current state of the art. In *Handbook of Research on Advanced Hybrid Intelligent Techniques and Applications* (pp. 161-189). IGI Global.

Swamy, S., & Ramakrishnan, K. V. (2013). An efficient speech recognition system. Computer Science & Engineering, 3(4), 21.

Weber, D. (2003): "Interactive user interface using speech recognition and natural language processing." U.S. Patent 6,499,013, issued December 24, 2002.