Voice Based Automated Transport And Tourism Enquiry System

Mr.M.Thangamani[1], K.Abishegavalli[2], M.A.Daurani [3], S.Jesima nilofer[4], A.Nandhini[5]
[1, 2, 3, 4, 5]Department of Computer Science and Engineering
[1, 2, 3, 4, 5]Saranathan College of Engineering, Trichy-620 012, Tamil Nadu

Abstract-*The presence of location based services is growing exponentially .The need for live information relative to the context is important .The proposed model aims to achieve this with users posing queries to the system .The system responds by bringing out the results and also reading out the results .This application is primarily focused on tourism and transport .The two most important areas requiring technology to provide suitable solutions. It presents the implementation of enquiry system which operates on the voice from the user .There is no communication is more effective than voice which could be easily understood if there is well understanding what is the real request is?. This system too uses the voice commands and gives the required information in the form of voice .The presence of location based services is growing exponentially .The need for live information relative to the context is important. The proposed model aims to achieve this with users posing queries to the system. The system responds by bringing out the results and also reading out the results. This application is primarily focused on tourism and transport. The two most important areas requiring technology to provide suitable solutions.*

Keywords-Speech Recognition, Text to Speech, Speech Engine, Voice Commands, System, Request, Audio.

I. INTRODUCTION

Voice Based Automated Transport Enquiry System is developed for providing the information for the enquiry in transport terminals. This project is developed using .Net technology using c# Programming language. This uses sql server for storing the information to be provided to the user. This user Microsoft Speech recognition to detect the voice from the user and uses the speech control to deliver the voice output. This also displays the results on the screen for further verification.

A software design is a description of the structure of the software to be implemented, the data which is part of the system, the interfaces between the system components and sometimes the algorithms used. Designers do not arrive at a finished design immediately but develop the design iteratively through a number of different versions. The design process involves adding formality and detail as the design is developed with constant backtracking to correct earlier designs.

II. SPEECH RECOGNITION

You might have already used speech recognition in products, and maybe even incorporated it into your own application, but you still don't know how it works. This document will give you a technical overview of speech recognition so you can understand how it works, and better understand some of the capabilities and limitations of the technology. Speech recognition fundamentally functions as a pipeline that converts PCM (Pulse Code Modulation) digital audio from a sound card into recognized speech. The elements of the pipeline are:

1. Transform the PCM digital audio into a better acoustic representation.
2. Apply a "grammar" so the speech recognizer knows what phonemes to expect. A grammar could be anything from a context-free grammar to full-blown English.
3. Figure out which phonemes are spoken.
4. Convert the phonemes into words.

Transform the PCM digital audio the first element of the pipeline converts digital audio coming from the sound card into a format that's more representative of what a person hears. The digital audio is a stream of amplitudes, sampled at about 16,000 times per second. If you visualize the incoming data, it looks just like the output of an oscilloscope. It's a wavy line that periodically repeats while the user is speaking. While in this form, the data isn't useful to speech recognition because it's too difficult to identify any patterns that correlate to what was actually said. To make pattern recognition easier, the PCM digital audio is transformed into the "frequency domain." Transformations are done using a windowed fast-Fourier transform. The output is similar to what a spectrograph produces. In frequency domain, you can identify the frequency .The components of a sound. From the frequency components, it's possible to approximate how the human ear perceives the sound.

The fast Fourier transform analyzes every 1/100th of a second and converts the audio data into the frequency domain. Each 1/100th of a second results is a graph of the amplitudes of frequency components, describing the sound heard for that 1/100th of a second. The speech recognizer has a database of several thousand such graphs (called a codebook) that identify different types of sounds the human voice can make. The sound is "identified" by matching it to its closest entry in the codebook, producing a number that describes the sound. This number is called the "feature number." (Actually, there are several feature numbers generated for every 1/100 th of a second but the process is easier to explain assuming only one.)

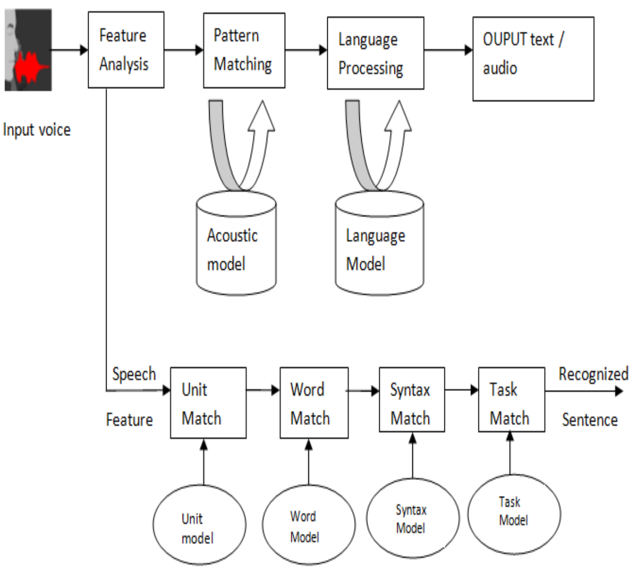


Figure 1.1 : Speech recognition

III.TEXT-TO-SPEECH

We might have already used text-to-speech in products, and maybe even incorporated it into your own application, but you still don't know how it works. This document will give you a technical overview of text-to-speech so you can understand how it works, and better understand some of the capabilities and limitations of the technology. Text-to-speech fundamentally functions as a pipeline that converts text into PCM digital audio. The elements of the pipeline are:

1. Text normalization
2. Homograph disambiguation
3. Word pronunciation
4. Prosody
5. Concatenate wave segments I'll cover each of these steps individually.

The "text normalization" component of text-to-speech converts any input text into a series of spoken words. Trivially, text normalization converts a string like "John rode home." to a series of words, "john", "rode", "home", along with a marker indicating that a period occurred. However, this gets more complicated when strings like "John rode home at 23.5 mph", where "23.5 mph" is converted to "twenty three point five miles per hour".

IV. PROPOSED SYSTEM

The proposed model uses a follow me cloud model. The user inputs his location. The system tracks all the users requirements from the server. When the user provides a voice command. The results related to the particular location are displayed. Apart from that the information is also read out using text to speech Very useful when touring or in outstations for locating cars, places of stay, routes, ATM'S, food, directions, transport, etc. Independent and autonomous solutions are provided. Location specific and easy to use. Even the not so tech savvy can use with simple voice commands. Provides display and multiple results with follow me options. Google maps are provided for easy to understand routes. Common utilities like ATM'S, hospitals, luggage rooms etc are included.

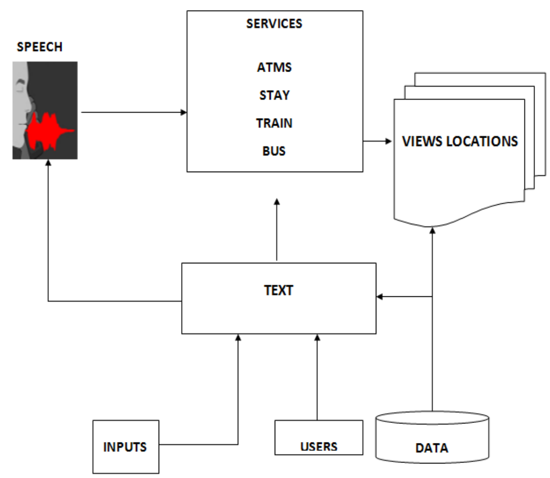


Fig 3.1 System Architecture

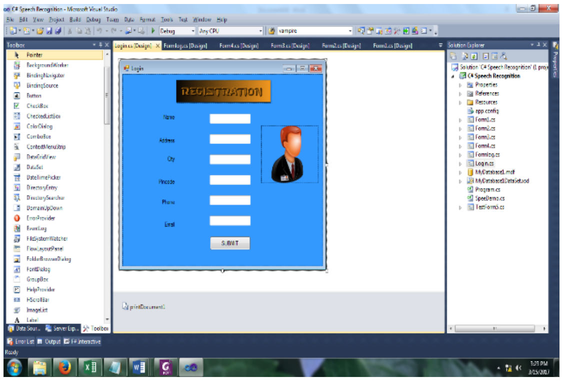
The following are 5 major modules in this proposed system

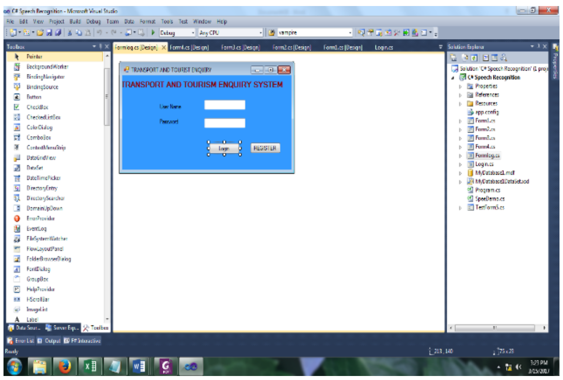
- User registration
- Service requestor
- Voice command process
- Result
- Text to speech

1) USER REGISTRATION

In this module, The user should register with the system for utilizing the services.The user is provided a login and password.The user may select services in advance – Follow Me Services.The follow Me services mean the system

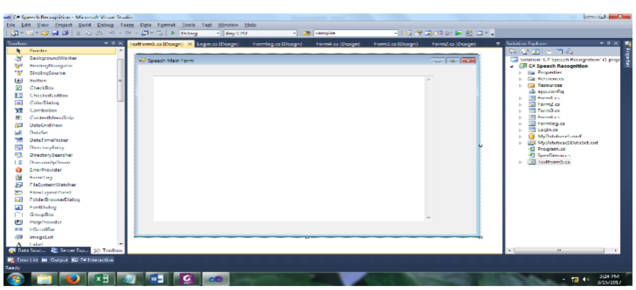
follows the user's location and provides info about certain common services mentioned by the
user in advance.





2) SERVICE REQUSTOR

The user input the location and requests the service .The system queries the web database. The user has the privilege to request any sort of information from the system such as ATM, Stay, Bus, Train etc. these request are given as the input to the system as voice and this voice is recognized. The user's current location is get and the information about the request made is around the user's location the ATM is displayed. Which is nearest and how many meters far away from the users location is also fetched from the database and displayed to the user and also in the form of audio to the user.



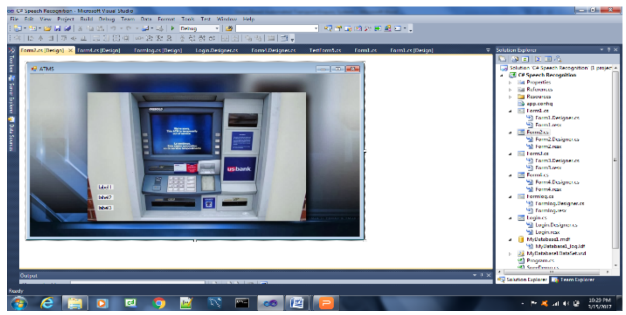
3) VOICE COMMAND PROCESS

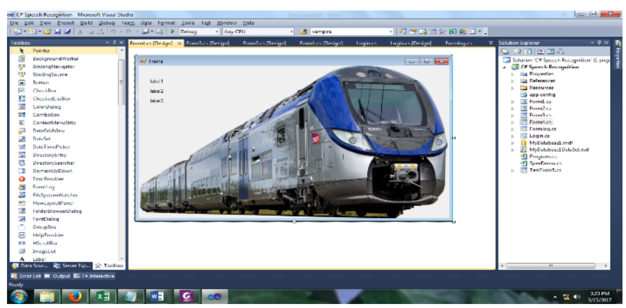
In this module, the user's voice commands are processed by the system

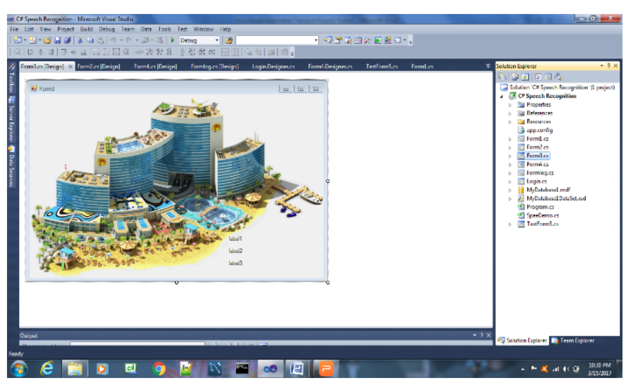
The answers for the queries are fetched from the server in real time.

IV. RESULT

The results fetched are displayed to the user in real time. The follow me services are fetched in advance.







5) TEXT TO SPEECH

The system also provides assistance in the form of text to speech for the displayed results. The user's current location is get and the information about the request made is around the user's location the ATM is displayed. Which is nearest and how many meters far away from the users location is also fetched from the database and displayed to the user and also in the form of audio to the user.

V. CONCLUSION

Voice Based Automated Transport Enquiry System is developed for providing the information for the enquiry in transport terminals. This project is developed using .Net technology using c# Programming language. This uses sql server for storing the information to be provided to the user. This user Microsoft Speech recognition to detect the voice from the user and uses the speech control to deliver the voice output. This also displays the results on the screen for further verification In the future, this model will be universally adopted for persons with the use of information technology, and this will result in less turnaround time and more efficiency in post-disaster settings.

The centralized process adds both security and at the same time accurate utilization of resources and essential items. The arrangement of the items are also convenient without any clutter and can be used by most of the users in the project without any problems .In future the application may be processed for using web services app and hence installed at any place. The model will be implemented as web services in the future so that all can access the services

REFERENCES

[1] S. Takamichi, T. Toda, A.W. Black, and S. Nakamura, “Modulation spectrum-constrained trajectory training algorithm for gmm-based voice conversion,” in ICASSP 2015, 2015.

[2] R. Aihara, T. Nakashika, T. Takiguchi, and Y. Ariki, “Voice conversion based on non-negative matrix factorization using phoneme-categorized dictionary,” in ICASSP 2014, 2014.

[3] S. Desai, E.V. Raghavendra, B. Yegnanarayana, A.W. Black, and K. Prahallad, “Voice conversion using artificial neural networks,” in ICASSP 2009, 2009.

[4] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigne, “Re-´ structuring speech representations using a pitch-adaptive timefrequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds,” Speech Communication, vol. 27, pp. 187–207, 1999.

[5] Y. Stylianou, “Voice transformation: A survey,” in ICASSP, April 2009, pp. 3585–3588.

[6] A. Kain and M.W. Macon, “Spectral voice conversion for text-to-speech synthesis,” in Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, 1998, vol. 1, pp. 285–288 vol.1.

[7] K. Y. Park and H. S. Kim, “Narrowband to wideband conversion of speech using gmm based transformation,” in ICASSP, 2000, vol. 3, pp. 1843–1846 vol.3.

[8] P. Taylor, Text-to-Speech Synthesis, Cambridge University Press, 2009.

[9] K. Fujii, J. Okawa, and K. Suigetsu, “High-individuality voice conversion based on concatenative speech synthesis,” International Journal of Electrical, Computer, Energetic, lectronic and Communication Engineering, vol. 1, no. 11, pp. 1617 –1622, 2007.

[10] D. Sundermann, H. H¨ oge, A. Bonafonte, H. Ney, A. Black, and¨ S. Narayanan, “Text-independent voice conversion based on unit selection,” in Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. IEEE, 2006, vol. 1.