

A Multimedia Interactive System

Manjushree Mahajan, Sushma Kate, Sandhya Katkar, Parinita Chinchalikar, Jyoti Hajare
Computer Engineering Department
G.H.Raisoni college of Engg.(Pune University) ,Pune, India

Abstract-In today's world communication devices reduce distance between people . Everyone needs devices for easy interaction with world. There are lots of multimedia devices such as mobile phones, web cam, headphones etc. and are gradually reducing the problems in communication.

Today there is not any system which contain all utilities into an single system. Multimedia Interactive System(MIS) contains all these utilities into a single system that is Conversion of speech to text (STT) & text to speech (TTS) in languages Like ENGLISH, HINDI, MARATHI ,windows stimulation through voice command, System modify Microsoft's speech engine to accept & perform operation of STT and TTS on Indian languages. Developed the directory to match the word in ENGLISH and HINDI, ENGLISH and MARATHI. We developed a paintbrush tool to recognize the characters according to the pixels of each character .System scans the pixels of each individual character and then recognize it .This helps to teach the English language.

The paper mainly contains the proposed system which we are going to develop. The main part of this paper is the system overview and main working of a system. The main procedure is that how actually working of system takes place.

Keywords— Multimedia Interactive system, Application Programming Interface, Text To Speech, Speech To Text.

I. INTRODUCTION

Multimedia Interactive System has been subject to a lot of hype, among the common people because of the high market penetration of Multimedia devices. To build an application depending on Multimedia systems to develop interactive projects Multimedia devices such as mobile phones, Web cam, headphones etc. are gradually reducing the problems in communication and offering the possibility to capturing images, converting text to speech, converting speech to text, image reorganization and other services. Thus building applications, so that average user having multimedia device can perform all these task using a single application.

Microsoft has developed a tool which contains text to speech ,speech to text and some utilities but it contains many drawbacks like less accuracy, it can't work in noisy environment and it gives problem to conflicting of words . proposed system contains the important features like it has more accuracy, it also contains separate Digit Engine and Menu Engine, different voices can be handled by less if else statements reduces complexity. It avoid all interferences (background sound) and high recognition of words.

Innovativeness: Till now there is Text To Speech (TTS) software, image reorganization software but we don't have a single application having all these facilities having in it. This application is been designed to provide human sensing ability to PC with respect to Voice, and Graphics.

The developed utility would assist the end users to enhance multimedia input system so that they can interact with the software in short time and ease of handling computer system. The systems key focus on speech API's text to speech application, voice based Active X control.

Need and Origin:

In India Multimedia Based applications are widely used but the process of TTS and Speech Recognition (SR) is not that popular to use in every GUI applications. Hence the next need from the ever growing community of mobile users is shifting from typical ring tone and gaming to serious business use like checking credit card dues, checking Speech market details, checking Bank balance etc. Hence the need arise to develop a particular system which would assist the end users to track the speeches, they are interested in prior to or post their investment. Here the end user may or might-not is an entity dealing with the speech exchange either directly or indirectly.

II. BASIC ARCHITECTURE

The **Speech Application Programming Interface** or **SAPI** is an API developed by microsoft to allow the use of speech recognition and speech synthesis within windows applications. To date, a number of versions of the API have been released, which have shipped either as part of a Speech SDK, or as part of the Windows OS itself. Applications that use SAPI include Microsoft office, Microsoft Agent and Microsoft Speech Server.

In general all versions of the API have been designed such that a software developer can write an application to perform speech recognition and synthesis by using a standard set of interfaces, accessible from a variety of programming languages. In addition, it is possible for a 3rd-party company to produce their own Speech Recognition and Text To Speech engines or adapt existing engines to work with SAPI. In principle, as long as these engines conform to the defined interfaces they can be used instead of the Microsoft-supplied engines.

In general the Speech API is a freely-redistributable component which can be shipped with any Windows application that wishes to use speech technology. Many versions (although not all) of the speech recognition and synthesis engines are also freely redistributable.

There have been two main 'families' of the Microsoft Speech API. SAPI versions 1 through 4 are all similar to each other, with extra features in each newer version. SAPI 5 however was a completely new interface, released in 2000. Since then several sub-versions of this API have been released[1].

In system SAPI is mainly use to inbuilt the sound of the developer. User will inbuilt sound by give voice input. Computer will listen this voice and work out when user will provide command . It also do the speech to text conversion and text to speech conversion. Main important part is SAPI to work out with system.

Digital to Analog conversion(DAC)

A digital-to-analog converter is a device that converts a digital (usually binary) code to an analog signal (current, voltage, or electric charge).[3]

Analog to Digital conversion(ADC)

An analog-to-digital converter performs the reverse operation. Signals are easily stored and transmitted in digital form, but a DAC is needed for the signal to be recognized by human senses or other non-digital systems[2].

III WORKING

The system mainly contains pipes and filter architecture which is given below:

Pipes-And-Filters

A very simple, yet powerful architecture, that is also very robust. It consists of any number of components (filters) that transform or filter data, before passing it on via connectors (pipes) to other components. The filters are all working at the same time. The architecture is often used as a simple sequence, but it may also be used for very complex structures.

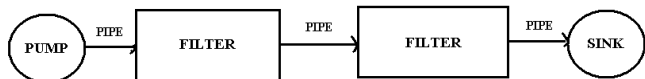


fig.1:pipes and filter architecture

How does it work

The application links together all inputs and outputs of the filters by pipes, then spawns separate threads for each filter to run.

Here's an idea of the relationships that can be created between the different filter processes, through pipes.

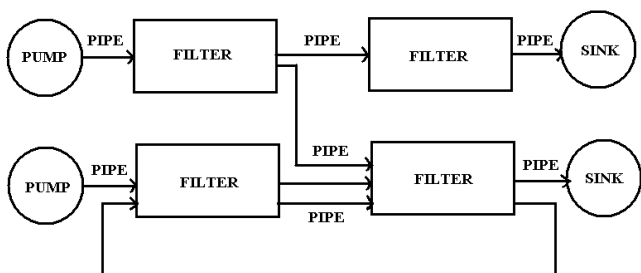


fig.2: working of pipes and filter

All filters are processes that run (virtually) at the same time. That means, they can run as different threads, co routines, or be located on different machines entirely. Every pipe connected to a filter has its own role in the function of the filter. So if you connect a pipe, you also need to specify the role it plays in the filter process. The filters should be made so robust that pipes can be added and removed at runtime. Every time the filter performs a step, it reads from its input pipes,

performs its function on this data, and places the result on all output pipes. If there is insufficient data in the input pipes, the filter simply waits.

In system pipes and filter architecture the data stream is entered through the input. For example working of menu engine, digit engine data input given in analog form i.e. in voice it will first of all convert analog to digital conversion which done by sound card inbuilt into computer then it further provide to SAPI. SAPI convert it into text .data gets filter and provide to require process like menu engine ,digit engine ,command control etc.

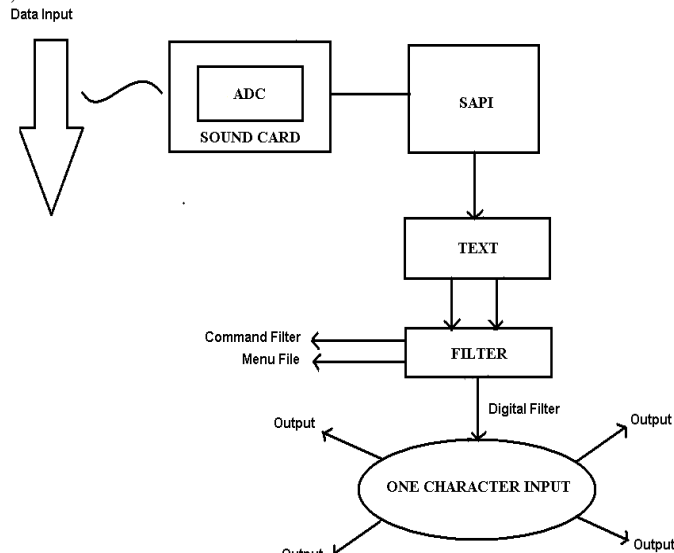


fig.3 :working of pipes and filter in system

Word relevance

A thing might be relevant, a document or a piece of information may be relevant. The basic understanding of relevance does not depend on whether user speak of "things" or "information".

In system when inbuilt the sound into computer the relevance of some word taking place .From this words we select the most common words and inbuilt that words into program.

Command and control mechanism

When working with voice command the commands are provided by voice and control with that command which are inbuilt is taking place. Control is like display the menu bar when user speak as a 'MENU'.

Devices :

A fundamental demand for the Multimedia device is that it must be able to connect to a network or PC Interface through Sound Card to initiate Speech tracking. The network could be either Wireless or Internet and the data providing instance could be user based Speech or .wav file or any active Windows application .The main qualification to be considered as a Multimedia device is obviously that the end-user is able to use it independently of any cables i.e. the end-user shall be able to carry the device with him/her and be able to use it independently without any strings attach (wires for network

connections or electric current cables) in case of mobile based Speech Recognition .The devices used are:

- Speaker / Headphone with Microphone
- Microphone Wired/wireless Compatible to Soundcard.
- Mobile phones.
- Laptops.

focus on the mobile phones (a GSM handset) as the device carrying out m-speech dictation and synthesis.

Data flow

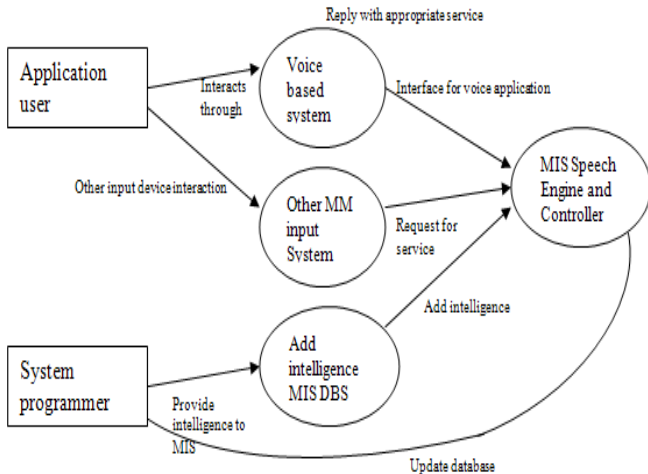


fig.4:DFD level 1

The system mainly contains two type of users system programmer and application user .As the system takes multimedia type of input the application user provide input in any type like speech, text, mobile input etc .As like in diagram voice input is given for further processing .It will further processing related with voice based system like Text to speech conversion or speech to text conversion. This provide to MIS controller system which is main part of system all the processing done in this part.

Any other type of multimedia input like mobile command is given it will request for service to the system. The system will identify the device and related processing is done on that. It again give to main part that is MIS controller system.

System programmer is main developer of the system can also do many task in the system. As shown in diagram he is going to provide intelligence to system by filtering the command provided to system. By filtering the command it add intelligence to the data base. It will further add intelligence to the computer system and querying the filtered data. It again goes to the main part that is MIS controller and filtered out the data .the data updating can be also done by programmer and it can directly done at MIS controller system.

All the working mainly done at the MIS controller.

SNAPSHOT

Description: This application reads the input provide by voice command and menu operation perform work by listening voice command.

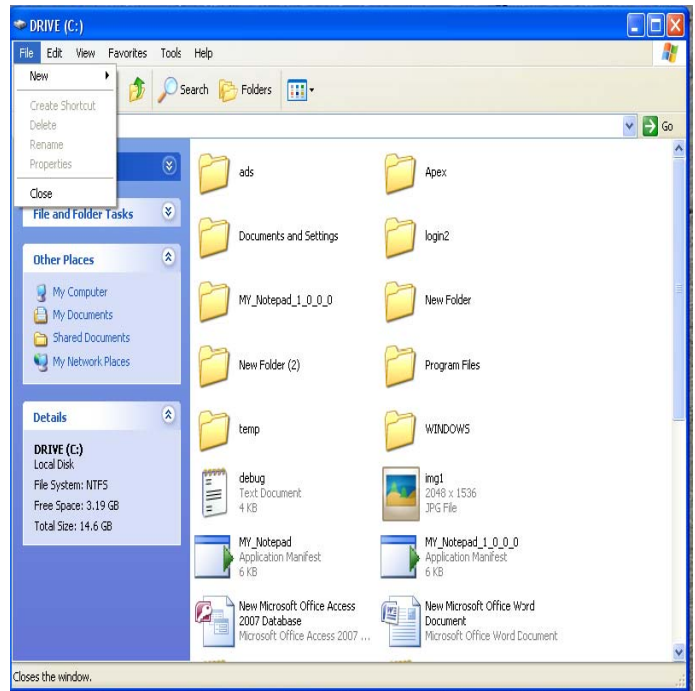


fig.5: Windows Menu Operation

IV. CONCLUSION

The Multimedia interactive system provide intelligence to the system through multimedia devices by using SAPI.

REFERENCES

- [1]Microsoft Systems Journal Whitepaper by Mike Rozak on the first version of SAPI.
- [2]www.wikipedia.org analog to digital conversion
- [3] en.wikipedia.org/wiki/digital-to-analog conversion
- [4] Course on Software Architectures - Pipe-And-Filter The Evolution of the Unix Time-sharing System
- [5]. Hjørland, B. & Sejer Christensen, F. (2002). Work tasks and socio-cognitive relevance: a specific example. Journal of the American Society for Information Science and Technology, 53(11), 960-965.
- [6] S. Kim, J. Lee, and G. Lee, "Incorporating second-order information into two-step major phrase break prediction for Korean," in Proceedings of International Conference on Spoken Language Processing, 2006, pp. 1487-1490