Development of Voice user interfaces (VUI)

Supriya Bachal,Aditya Joshi

Abstract— **Human machine communication is a major part of today's computer usability. When we talk about human machine interaction one of the main participants in this discussion would be user interfaces. User interfaces decide how the human interacts with the machine. As technology advances every application we build is made interactive for human feedback and preferences. This paper sheds light on the niche area of user interfaces for voice applications. Voice applications are in high demand as they are easy to use; however they have a more rigid set of rules for their user interface design. As speech synthesis and speech recognition technologies improve, the complexity and naturalness of applications also increases. A carefully crafted user interface helps the application become more palatable to the user.**

Index Terms—**: voice based interfaces, voice applications, mobile devices, voice recognition, voice user interfaces**.

INTRODUCTION

Technological improvements have shifted the focus of an UI from adapting the users input to fit the limitations of technology to facilitating interaction between the human and machine. Factors that is vital while designing voice interfaces are

1. The task requirements of the application
2. Capabilities and limitations of technology being used.
3. Characteristics of the user.

Voice interface is the most sought after user interface because it helps to reduce the cost, size & maintenance cost on the other hand it improves the durability. Speech requires more modest physical resources .speech based interactions can be scaled down to much smaller, readily available and cost effective form factors than visual or manual. Speech usually requires physical resources such as a microphone as input device and a headset/headphone or speakers as output device which are already a part of most computer systems, none the less they are small in size and inexpensive.

On a practical note, users with various disabilities who cannot use a mouse and a keyboard or who cannot see pictures on a screen depend largely on these on audio presentation of information.

As the day gets busier and life gets tougher, multitasking is no longer just another option is has become the only option to survival. In this situation many users encounter a situation where their eyes and hands are preoccupied with another task, in such circumstances the voice user interface (VUI) based applications are a boon.

Voice user interface and graphical user interface are the two major contenders for user interface development. VUI is preferable because it is natural to humans and puts minim strain and requires minimum effort from the user. Speech is descriptive while vision is referential and hence both VUI and GUI are complementary and can be used together for great effect. VUI's are invisible to the user and hence make the use less complex .VoiceXML applications are available which are meant to focus on VUI for mobile applications.

General structure of a VUI:

A voice can be used in three ways: to command a computer, to enter information, and to communicate with other people .components of building voice applications are:

1. End user
2. Front-end interfaces
3. Voice recognition system
4. Dictionary and text file database.

The figure below gives clear understanding of how the components of the voice user interface interact with one another.

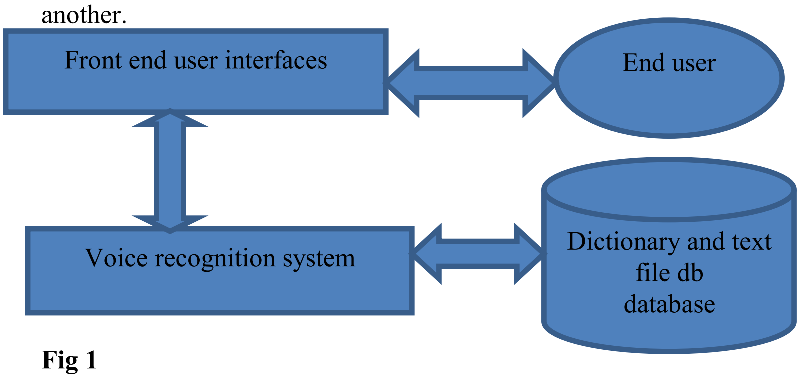

Front end user interfaces
End user
Voice recognition system
Dictionary and text file db database

Fig 1

Each component is explained as follows:

End users: generally end users is a term pertaining to the users of the device. They can use devices to establish communication and make voice feedback with the application. End users can be viewed as the users of a mobile phone, laptop, palmtop etc.

Front end interfaces

A front end interface means a user interface that can directly access and communicate with users by supporting speech I/O with graphic and icon based menus. It gets users speech and delivers them to a voice recognition system to recognize the voice inputs and generates a feedback to users after the commands have been executed through the several sub-functions of the system.

Voice recognition systems:

It is the heart of the voice application, which understands the user's voice, makes the application work and generate a voice feedback. This system is a vital part to allow a user to use voice as input. Thus to analyze the user voice and give feedback .The application should host a voice recognition system containing the process which maps an acoustic speech signal to text and form the abstract meaning of speech.

Dictionary and test files database:

Depending on the type of the device used and the requirements of the user the application is meant to support some distinguished type of input or provide special voice feedback. The application can provide users with additional information or help functions based on these database files. An additional text file database is installed in order to update the application.

I. TYPES OF COMMUNICATION METHODS

VUI usually uses voice as principle input. In mobile systems, voice interaction is associated with recognizing a user voice and generating desired phone sound depending upon the system environments. Many different methods are used to achieve this some of the most popular methods are listed below.

Speaker dependent method:

This method works on the principle that voices are prerecorded into the device. At an instance where voice command is entered, it is compared to the record message and the action is executed. This approach is used to a certain extent for feedback such as explanations or responding messages based on user requests. When voice authentication process is used for security of the phone, this method is adopted.

Speaker independent method:

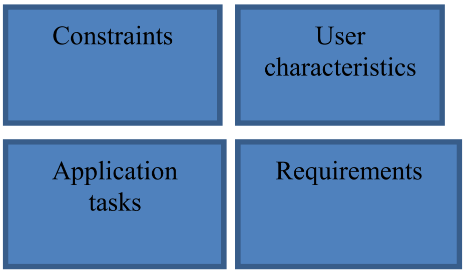
This method was developed with the objective to recognize spoken input without the need to train the device for individual voices. A voice recognition system creates a separate electronic template for each word that is spoken by each user. It is able to recognize anyone's voice for some specified vocabularies. This method is specially effective for telephone network conversations' where callers from various dialects, genders, tones, pitches are exposed to the system.

Text to speech (TTS):

Text to speech is an approach which uses a certain engine to derive or synthesize speech or audio output from text input. The above two methods use voice input however this method gives voice output. TTS is mostly entirely implemented in software and only standard audio capability is required. TTS has various applications such as reading messages for users, reading out electronic mails, or reading out a book to the user.IT can also be used to generate spoken prompts in voice applications on mobile devices.

Designing strategies:

Designing strategy is dependent upon the following interrelated factors:



1. Dialog design
2. Accommodation for various environments and dialects
3. Prompts
4. Confirmation and error detection and error handling
5. Learning user speech.

Dialog design:

Dialog design is the very basis of the VUI. The dialogue design models the interaction and the flow of the conversation and responses to likely instances.it is the determinant for the applications functionality. Dialog design is derived from logical and linguistic criteria, and callers' mental models. Dialog design helps to set some restrictions on the user and helps to guide him through the process of maintaining these set of rules. Some of the pointers for a good dialogue design are listed as follows:

1. Dialogs have to be efficient and short

2. Dialogs have to be clear and structured

3. Novices have to be guided. They need to know what kind of information to provide. Experienced users know what to say and need fast outputs.

Accommodation for various environments and dialects:

For practical use of VUI applications, the environment cannot be predicted; hence the designing of the VUI should be done such that it can accommodate at least most environments if not all. There might be difference in the quality of sound recorded, noise interference may vary, level of unwanted background sound is another variable factor. Different regions of the world have different dialects for the same language this has to be taken into consideration.

Prompts:

Prompt serves the dual purpose of giving instruction to the user as well as informing the user of the status of the application. It gives the user statistical and content information. Prompts should be clear, to the point and not too wordy. They should be conventional and colloquial the goal is to make the dialogue as intuitive for the caller as possible.

Confirmation, error detection and error handling:

Error detection and confirmation are essential for the application to run smoothly. Confirmation is a situation where the input given by the user is repeated by the system and confirmed or verified. Error detection is a situation where the system cannot recognize any or some part of the input. If there is no utterance from the user side the system prompts the user for input when expected. If the system cannot identify the speech or input is unacceptable then phrases such as “please repeat”, “please say again” etc. are used this is called error handling.

Learning user speech:
A voice application should have a learning technique it increases the accuracy with which it identifies user voices; thus the application remembers a user’s speech of each word as people speak. This should allow a voice application, even though everyone speaks with varying accents and inflection. Apart from learning how a user pronounces words, a voice application uses grammatical context and frequency of use to predict the next word. These statistical tools allow the software to function efficiently.

Dialogue:
Its aim is to bring the pros of web based development and content delivery to interactive voice applications. Since the technology uses XML syntax it is easy to use with other XML based technologies.
VUI is usually in a form where the user has to memorize and reproduce information; hence they cannot be content rich as it is not possible to memorize pages of verbal information. The design should be simple, with short and clear dialogues.
Voice interface design and development tools such as heyanita, tellme studio and many more can be used to develop VUI’s.

Method of implementation:
Voice user interfaces can be implemented using various ways one of the most common ways to implement voice user interfaces is through using voiceXML. VoiceXML is aW3C standard markup language for scripting voice interactions between a computer and a certain person. Its designs renders it the ability to create audio dialogues that feature synthesized speech, digitized audio, identification of DTMF (dual tone multi-frequency) key input, recording spoken input and telephony. VoiceXML uses natural dialogue. Its aim is to bring the pros of web based development and content delivery to interactive voice applications. Since the technology uses XML syntax it is easy to use with other XML based technologies.

VUI is usually in a form where the user has to memorize and reproduce information; hence they cannot be content rich as it is not possible to memorize pages of verbal information. The design should be simple, with short and clear dialogues.
Voice interface design and development tools such as heyanita, tellme studio and many more can be used to develop VUI’s.

Conclusion:
The paper stated the development process and the need for Voice User Interface. We mainly gave insights into mobile based VUI for applications. A discussion for the various components of voice user interface was provided and the desirous qualities of a voice user interface were determined. In totality the paper is an overview on designing a voice user interface for an application.
.

REFERENCES

1. Clark, Herbert H. Arenas of Language Use. University of Chicago Press, Chicago, IL, 1992.
2. Grice, H. P. "Logic and Conversation," Syntax and Semantics: Speech Acts, Cole & Morgan, editors,Volume 3, Academic Press, 1975.
3. Grosz, Barbara, and Candy Sidner. "Attention, Intentions, and the Structure of Discourse," Computational Linguistics, Volume 12, No. 3, 1986.
4. Kamm, Candace. "User Interfaces for Voice Applications," Voice Communication Between Humans and Machines, National Academy Press, Washington, DC, 1994.
5. Kitai, Mikia, A. Imamura, and Y. Suzuki. "Voice Activated Interaction System Based on HMM-based Speaker-Independent Word Spotting," Proceedings of the Voice I/O Systems Applications Conference, Atlanta, GA, September 1991.
6. Ly, Eric, and Chris Schmandt. "Chatter: A Conversational Learning Speech Interface," AAAI Spring Symposium on Intelligent Multi-Media Multi-Modal Systems, Stanford, CA, March 1994.
7. Martin, Paul and Andrew Kehler. "SpeechActs: A Testbed for Continuous Speech Applications" (7 PostScript pages), AAAI- 94 Workshop on the Integration of Natural Language and Speech Processing, 12th National Conference on AI, Seattle, WA, July 31-August 1, 1994.
8. Nielsen, Jakob. "The Usability Engineering Life Cycle," IEEE Computer, March 1992.
9. Roe, David, and Jay Wilpon, editors. Voice Communication Between Humans and Machines, National Academy Press, Washington, DC, 1994.
10. Schmandt, Chris. Voice Communication with Computers: Conversational Systems, Van Nostrand Reinhold, New York, 1994.
11. Stifelman, Lisa, Barry Arons, Chris Schmandt, and Eric Hulteen, "VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker, ACM INTERCHI ‘93 Conference Proceedings, Amsterdam, The Netherlands, April 24-29, 1993.
12. Yankelovich, Nicole. "Talking vs. Taking: Speech Access to Remote Computers" (2 PostScript pages). ACM CHI ‘94 Conference Companion, Boston, MA, April 24-28, 1994.
13. Yankelovich, Nicole and Eric Baatz. "SpeechActs: A Framework for Building Speech Applications" (9 PostScript pages). AVIOS ‘94 Conference Proceedings, San Jose, CA, September 20-23, 1994.
14.Design a Natural User Interface for GestureRecognition Application-Zhaochen Liu

15.Voice Communication Between Humans and Machines By for the National Academy of Sciences

Author Profiles:

Supriya Bachal
Student of Bachelors of Engineering (Computer Science)
Savitribai Phule Pune University

Aditya Joshi
Student of Bachelors of Engineering (Computer Science)
Savitribai Phule Pune University