International Journal of Computer Applications Technology and Research
Volume 4– Issue 2, 157 - 164, 2015, ISSN:- 2319–8656

Assistive Examination System for Visually Impaired

Manvi Breja
Manav Rachna College of
Engineering
Faridabad, Haryana, India

Abstract: This paper presents a design of voice enabled examination system which can be used by the visually challenged students. The system uses Text-to-Speech (TTS) and Speech-to-Text (STT) technology. The text-to-speech and speech-to-text web based academic testing software would provide an interaction for blind students to enhance their educational experiences by providing them with a tool to give the exams. This system will aid the differently-abled to appear for online tests and enable them to come at par with the other students. This system can also be used by students with learning disabilities or by people who wish to take the examination in a combined auditory and visual way.

Keywords: Speech recognition, speech synthesis, prosody analysis, phonemes, speech API.

1. INTRODUCTION

In today's era of rapidly evolving technological advances, a major change has occurred in the educational system prevalent in schools and colleges. Along with other changes in the teaching system, the examination system has evolved significantly Voice Enabled Examination System for the visually impaired people has been a very active research area for a long time and much more success is also achieved in this area.. According to the National Center for Educational Statistics , "*The number of students with disabilities attending higher education has increased. In a recent study, the number of postsecondary undergraduate students identified as having disabilities in the United States was found to be 428,280, representing 6% of the student body.*" With the growing number of blind people attending college, there has been a growing need for such system that can aid these visually impaired students .In today's era, the conventional pen and paper tests have been replaced by online examination systems. The word 'Online' here refers to not necessarily web-based or browser-dependent, but to a network that links all of the test takers to a common server. As of now, there are many companies in the industry that offer testing solutions, but none of them has on offer an examination system which is voice-enabled and assists the differently abled.

2. RELATED WORK

Technology has removed many barriers to education and employment for visually impaired individuals. Various technology exists for students with visual impairements. These include:

2.1 Screen Magnification

Screen magnification software is used by people with visual impairments to access information on computer screens. The software enlarges information on the screen by incremental factors (2 x magnification, 3x up to 20x magnification). Most screen magnification programs have the flexibility to magnify the full screen, parts of the screen, or a magnifying glass view of the area around the cursor or pointer. Commonly used screen magnification software are-

• MAGic, developed by Freedom Scientific Inc. Blind/Low Vision Group.
• ZoomText, developed by Ai Squared
• BigShot, developed by Ai Squared .

2.2 Screen Readers

A screen reader is a software application that attempts to identify and interpret what is being displayed on the screen. Screen reading software reads aloud everything on computer screens, including text, pull-down menus, icons, dialog boxes, and web pages. Screen readers run simultaneously with the computer's operating system and applications. There are mainly two types of screen readers- the CLI screen readers and the GUI screen readers.

2.3 Optical Character Recognition Systems

Optical character recognition (OCR) technology offers blind and visually impaired persons the capacity to scan printed text and then speak it back in synthetic speech or save it to a computer. There are three essential elements to OCR technology—scanning, recognition, and reading text. Initially, a printed document is scanned by a camera. OCR software then converts the images into recognized characters and words. Some of the most popular OCR systems are:

- Kurzweil 1000, developed by Kurzweil Educational Systems
- OpenBook, developed by Freedom Scientific Inc.
- Eye-Pal, developed by ABISee, Inc.

2.4 Electronic Portable Note-Takers

Electronic Braille note takers are small, portable devices with Braille keyboards for entering information. They use a speech synthesizer or Braille display for output. The user enters the information on the Braille keyboard and has the option of transferring it to a larger computer with more memory, reviewing it using the built in speech synthesizer or Braille display, or printing it on a Braille or ink print printer.
The commonly used note takers are:

- Braille 'n' Speak, developed by Freedom Scientific, Inc.
- Type 'n' Speak, developed by Freedom Scientific, Inc.
- PacMate Series, developed by Freedom Scientific, Inc.
- VoiceNote, developed by Pulse Data.

2.5 Portable Reading Devices

One of newer blind technologies is the portable reading device, which downloads books and then reads them out loud in a synthesized voice. They are specially designed with the blind and visually impaired community in mind. The Victor Reader Stream and the BookSense audio book are popular models.

3. PROPOSED SYSTEM

The prime interest behind the development of this system is to implement speech technology in an application in such a way that it enables the visually-challenged candidates to appear for a computer-adaptive online examination. The system is a stand-alone application which uses Speech-To-Text (STT) and Text-To-Speech (TTS) technology to provide the users almost all of the capabilities of a conventional online examination.

The online examination system is adaptable to different types of questions pertaining to different subjects, different time limits and different marking schemes, and can be customized according to the needs of any organization. All the data pertaining to the test is stored in a database which is linked to the application.

The Voice Enabled Examination System is able to read aloud the questions and the different options available to the test-taker. The candidate has to answer the question by speaking out the option number. The system registers the answer given by the candidate and moves on to the next question. At the end of the test, a report is generated by the system.

This system is equipped with the following functionalities:-

- Authentication of candidates via a mechanism of unique registration id and FolderLocking.
- Reading out of the questions by the application.
- Registering the answer of the candidate which has been spoken by him/her.
- Announcement of score to the candidate at the end of the exam.
- Folder Locking which ensures encryption of candidates' data in a folder and allows only the administrator to unlock that folder containing the candidates' details of the exam.
- Sending the resultant score sheet of the Examination to the registered mail id of the Candidate as well as to the Administrator.
- Generating the Certificate of scored marks and printing it at the end of the exam.
- Voice notifications to the candidates about the status of time left for each question.
- will inculcate the feature of Photograph Matching of the candidate while he appears for an exam from the administrator side for the security purposes.

The remainder of this paper is organized as follows. Section IV introduces some introduction on speech synthesis, Section V presents the concepts of speech recognition, Section VI discusses the application frameworks used, Section VII presents the snapshots of the results, Section VIII presents the future scope. Finally Section IX concludes the paper.

4. SPEECH SYNTHESIS

Speech synthesis is the artificial production of human speech. A synthesizer can be implemented in software or hardware.. A Text-To-Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems that simply concatenate isolated words or parts of sentences, denoted as Voice Response Systems, are only applicable when a limited vocabulary is required (typically a few one hundreds of words), and when the sentences to be pronounced respect a very restricted structure. It is thus more suitable to define Text-To-Speech as the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter [4].

The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood. An intelligible text-to-speech program allows people with visual impairments or reading disabilities to listen to written works on a home computer. Many computer operating systems have included speech synthesizers since the early 1980s. The text-to-speech (TTS) synthesis procedure consists of two main phases. The first one is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation, and the second one is the generation of speech waveforms, where the acoustic output is produced from this phonetic and prosodic information. These two phases are usually called as high- and low-level synthesis. A simplified version of the procedure is presented in Figure.

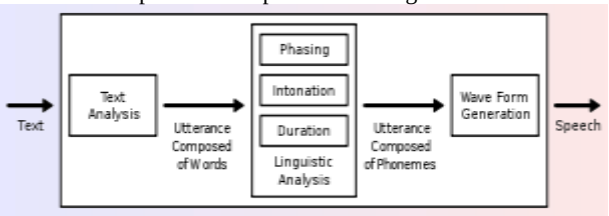


Figure 1: Process of Text-to-Speech Synthesizer

The input text might be for example data from a word processor, standard ASCII from e-mail, a mobile text-message, orscanned text from a newspaper. The character string is then preprocessed and analyzed into phonetic representation which is usually a string of phonemes with some additional information for correct intonation, duration, and stress. Speech sound is finally generated with the low-level synthesizer by the information from high-level one [8].

4.1. Process of Speech Synthesis

4.1.1. Structure analysis

Processes the input text to determine where paragraphs, sentences, and other structures start and end. For most languages, punctuation and formatting data are used in this stage.

4.1.2. Text pre-processing:

Analyzes the input text for special constructs of the language. In English, special treatment is required for abbreviations, acronyms, dates, times, numbers, currency amounts, e-mail addresses, and many other forms. Other languages need special processing for these forms, and most languages have other specialized requirements.

The remaining steps convert the spoken text to speech:

4.1.3. Text-to-phoneme conversion:

Converts each word to phonemes. A phoneme is a basic unit of sound in a language.

4.1.4. Prosody analysis:
Processes the sentence structure, words, and phonemes to determine the appropriate prosody for the sentence.

4.1.5. Waveform production:
Uses the phonemes and prosody information to produce the audio waveform for each sentence [3].

4.2. Synthesizer Technologies

The most important qualities of a speech synthesis system are naturalness and intelligibility. The ideal speech synthesizer is both natural and intelligible. Speech synthesis systems usually try to maximize both characteristics.
The two primary technologies for generating synthetic speech waveforms are concatenative synthesis and formant synthesis. Each technology has strengths and weaknesses, and the intended uses of a synthesis system will typically determine which approach is used [7].

5. SPEECH RECOGNITION SYSTEM

The speech is primary mode of communication among human being and also the most natural and efficient form of exchanging information among human in speech. Speech Recognition can be defined as the process of converting speech signal to a sequence of words by means of an algorithm implemented as a computer program. Speech processing is one of the exciting areas of signal processing[1]. Since the 1960s computer scientists have been researching ways and means to make computers able to record interpret and understand human speech. Throughout the decades this has been a daunting task. Even the most rudimentary in the early years. It took until the 1980s before the first systems problem such as digitalizing (sampling) voice was a huge challenge arrived which could actually decipher speech. Off course these early systems were very limited in scope and power. Communication among the human being is dominated by spoken language, therefore it is natural for people to expect speech interfaces with computer .computer which can speak and recognize speech in native language[2]. Machine recognisation of speech involves generating a sequence of words best matches the given speech signal. Some of known applications include virtual reality, Multimedia searches, auto-attendants, travel Information and reservation, translators, natural language understanding and many more Applications.

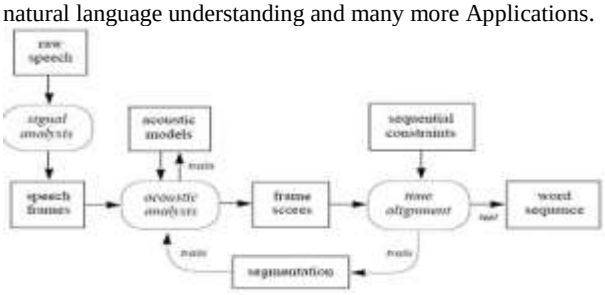


Figure 2: Structure of a standard speech recognition system.

A standard speech recognition system consists of the following components [5] **:-**

Raw speech. Speech is typically sampled at a high frequency, e.g., 16 KHz over a microphone or 8 KHz over a telephone. This yields a sequence of amplitude values over time.

Signal analysis. Raw speech should be initially transformed and compressed, in order to simplify subsequent processing. Many signal analysis techniques are available which can extract useful features and compress the data by a factor of ten without losing any important information.

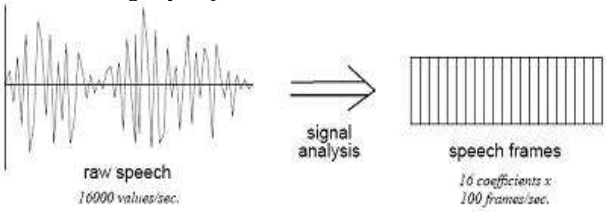


Figure 3: Conversion of raw speech to speech frames through signal anaysis

Speech frames. The result of signal analysis is a sequence of speech frames, typically at 10 msec intervals, with about 16 coefficients per frame. These frames may be augmented by their own first and/or second derivatives, providing explicit information about speech dynamics; this typically leads to improved performance. The speech frames are used for acoustic analysis.

Acoustic models. In order to analyze the speech frames for their acoustic content, we need a set of acoustic models. There are many kinds of acoustic models, varying in their representation, granularity, context dependence, and other properties.

The major steps of a typical speech recognizer are as follows:

Grammar design:
Defines the words that may be spoken by a user and the patterns in which they may be spoken.

Signal processing:
Analyzes the spectrum (i.e., the frequency) characteristics of the incoming audio.

Phoneme recognition:
Compares the spectrum patterns to the patterns of the phonemes of the language being recognized.

Word recognition:
Compares the sequence of likely phonemes against the words and patterns of words specified by the active grammars.

Result generation:
Provides the application with information about the words the recognizer has detected in the incoming audio.

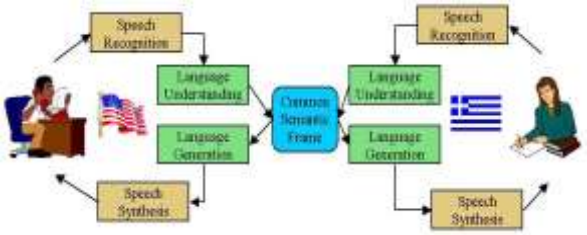


Figure 4: Process of Speech Recognition and Speech synthesis

5.1. Speech Recognition Techniques

The goal of speech recognition is for a machine to be able to "hear," understand," and "act upon" spoken information. The

earliest speech recognition systems were first attempted in the early 1950s at Bell Laboratories, Davis, Biddulph and Balashek developed an isolated digit Recognition system for a single speaker. The goal of automatic speaker recognisation is to analyze, extract characterize and recognize information about the speaker identity. The speaker recognisation system may be viewed as working in a four stages

1. Analysis
2. Feature extraction
3. Modeling
4. Testing

5.1.1. Speech analysis technique
Speech data contain different type of information that shows a speaker identity. This includes speaker specific information due to vocal tract, excitation source and behavior feature. The information about the behavior feature also embedded in signal and that can be used for speaker recognition. The speech analysis stage deals with stage with suitable frame size for segmenting speech signal for further analysis and extracting [6] .

5.1.2. Feature Extraction Technique
The speech feature extraction in a categorization problem is about reducing the dimensionality of the input vector while maintaining the discriminating power of the signal. As we know from fundamental formation of speaker identification and verification system, that the number of training and test vector needed for the classification problem grows with the dimension of the given input so we need feature extraction of speech signal

5.1.3 Modeling Technique
The objective of modeling technique is to generate speaker models using speaker specific feature vector. The speaker modeling technique divided into two classification speaker recognition and speaker identification. The speaker identification technique automatically identify who is speaking on basis of individual information integrated in speech signal The speaker recognisation is also divided into two parts that means speaker dependant and speaker independent. In the speaker independent mode of the speech recognisation the computer should ignore the speaker specific characteristics of the speech signal and extract the intended message .on the other hand in case of speaker recognisation machine should extract speaker characteristics in the acoustic signal. The main aim of speaker identification is comparing a speech signal from an unknown speaker to a database of known speaker. Speaker recognition can also be divide into two methods, text- dependent and text independent methods. In text dependent method the speaker say key words or sentences having the same text for both training and recognition trials. Whereas text independent does not rely on a specific texts being spoken

5.1.4 Matching Techniques
Speech-recognition engines match a detected word to a known word using one of the following techniques:
.
5.1.4.1. Whole-word matching
The engine compares the incoming digital-audio signal against a prerecorded template of the word. This technique takes much less processing than sub-word matching, but it requires that the user (or someone) prerecord every word that will be recognized - sometimes several hundred thousand words. Whole-word templates also require large amounts of storage (between 50 and 512 bytes per word) and are practical only if the recognition vocabulary is known when the application is developed .

5.1.4.2. Sub-word matching
The engine looks for sub-words – usually phonemes and then performs further pattern recognition on those. This technique takes more processing than whole-word matching, but it requires much less storage (between 5 and 20 bytes per word). In addition, the pronunciation of the word can be guessed from English text without requiring the user to speak the word beforehand to discuss that research in the area of automatic speech recognition had been pursued for the last three decades.

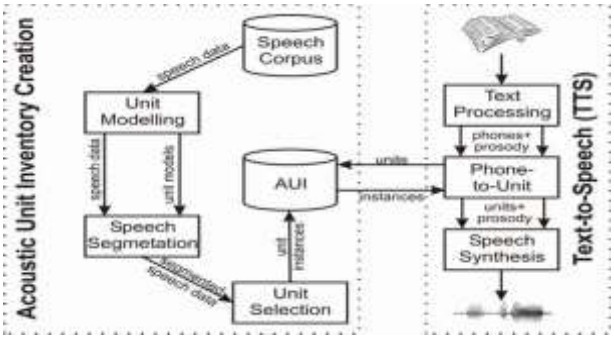


Figure 5: The general scheme of a concatenative text-to-speech system.

6. APPLICATION FRAMEWORKS

Several methods and interfaces for making the implementation of synthesized speech in desired applications easier have been developed during this decade. It is quite clear that it is impossible to create a standard for speech synthesis methods because most systems
act as stand alone device which means they are incompatible with each other and do not share common parts. However, it is possible to standardize the interface of data flow between the application and the synthesizer. Usually, the interface contains a set of control characters or variables for controlling the synthesizer output and features. The output is usually controlled by normal play, stop, pause, and resume type commands and the controllable features are usually pitch baseline and range, speech rate, volume, and in some cases even different voices, ages, and genders are available. Most of the present synthesis systems support so called Speech Application Programming Interface (SAPI) which makes easier the implementation of speech in any kind of application. For Internet purposes several kind of speech synthesis markup languages have been developed to make it possible to listen to synthesized speech without having to transfer the actual speech signal through network.

6.1. Speech Application Programming Interface

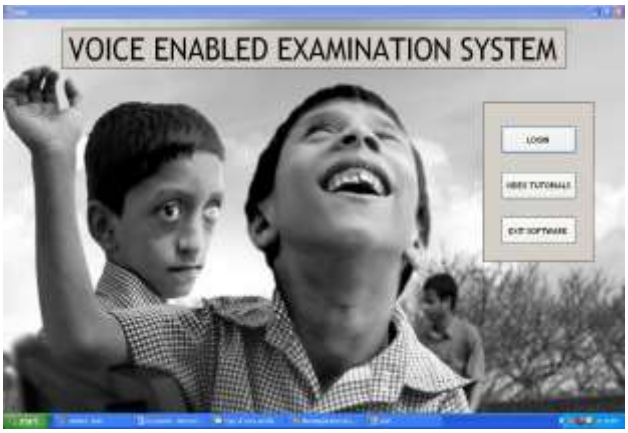
SAPI is an interface between applications and speech technology engines, both text-tospeech and speech recognition (Amundsen 1996). The interface allows multiple applications to share the available speech resources on a computer without having to program the speech engine itself. Speech synthesis and recognition applications usually require plenty of computational resources and with SAPI approach

lots of these resources may be saved. The user of an application can also choose the synthesizer used as long as it supports SAPI. Currently SAPIs are available for several environments, such as MS-SAPI for Microsoft Windows operating systems and Sun Microsystems Java SAPI (JSAPI) for JAVA based applications. SAPI text-to-speech part consists of three interfaces. The *voice text* interface which provides methods to start, pause, resume, fast *attribute interface* allows access to control the basic behavior of the forward, rewind, and stop the TTS engine during speech. The TTS engine, such as the audio device to be used, the playback speed (in words per minute), and turning the speech on and off. With some TTS systems the attribute interface may also be used to select the speaking mode from predefined list of voices, such as female, male, child, or alien. Finally, the *dialog interface* can be used to set and retrieve information regarding the TTS engine to for example identify the TTS engine and alter the pronunciation lexicon.

7. RESULTS

The proposed system opens up with the interface which has the two functionalities: login for the administrator and the student and watching the video tutorials on the related test topic.

Figure 6: Cover Page that is opened when the project is run



After clicking the login button, login page for candidate opens and from that administrator can also login. Login requires the authenticated username and password.

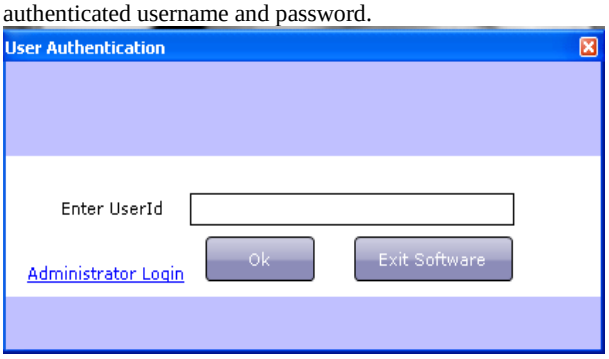


Figure 7: Login form for candidate

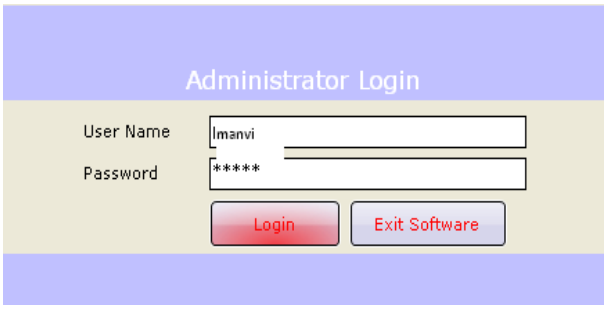


Figure 8: Login form for administrator

The administrator has the functionalities of registering the user for the system, adding the details of the users, ability to view all the user's details, folder lock facility so that the folder is authenticated, can be viewed by only the adminstrator.

Figure 9: Admin panel that opens up whenever administrator logins

The users who have come to appear for the test need to give their details to administrator for maintaining the record.

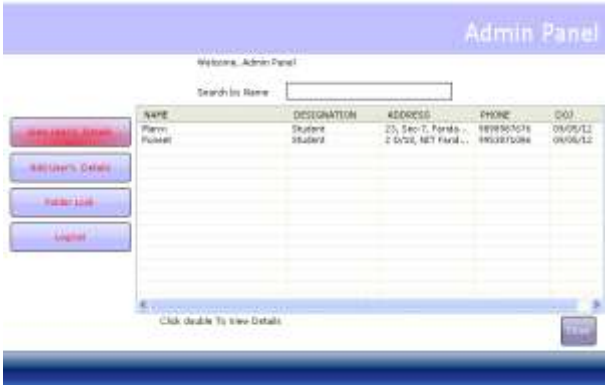


Figure 10: Display of user's details in grid layout format

The administrator has the ability to search the record of any candidate, adding the new candidate, deleting, updating the details, adding the photograph of the candidate.

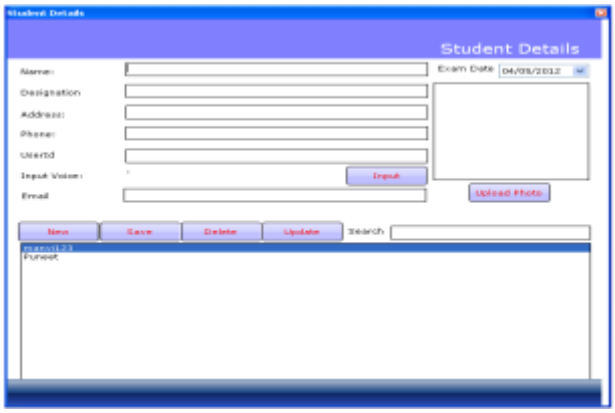


Figure 11: Add user detail form to be filled for each candidate

The administrator uses the facility of the folder lock i.e. folder in which candidate's records are kept can be accessible by only the administrator.

Figure 12: When the administrator clicks on folder lock

Figure 13: choosing the password to lock the desired folder

Figure 14: Displaying the locked status of folder

Authentication for the candidate who have come to appear for the test is also done. Since the candidate are blind, more security needs to be encorporated. Voice clip is taken when the students comes for registration. When the student will come to appear for the test, his voice clip is matched. Only when the authentication is done, the user is allowed to appear for the test.

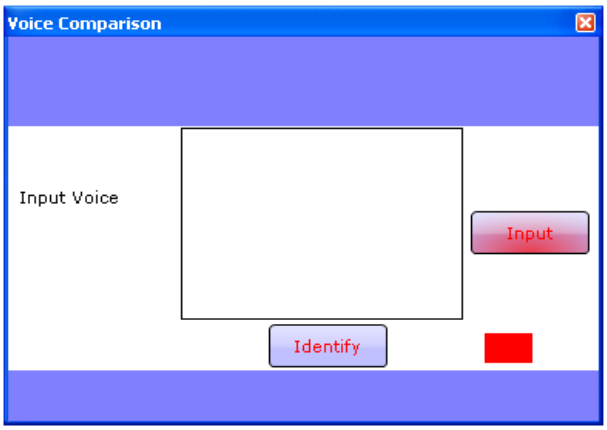


Figure 15: If authentication succeeds, matching video clip is to be browsed

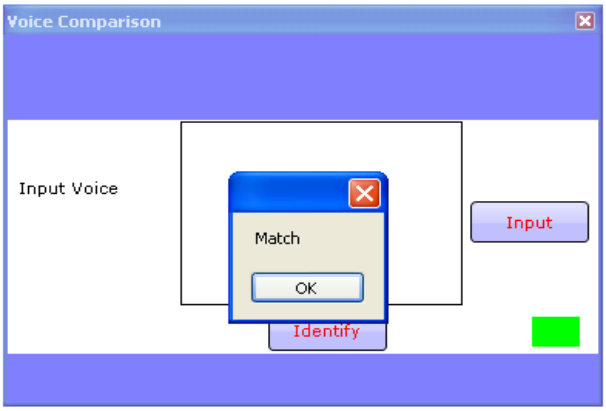


Figure 16: Displaying the status that the voice clip is matched

Now when the authenticated user login for the test, the system read aloud the initial instructions for the test.

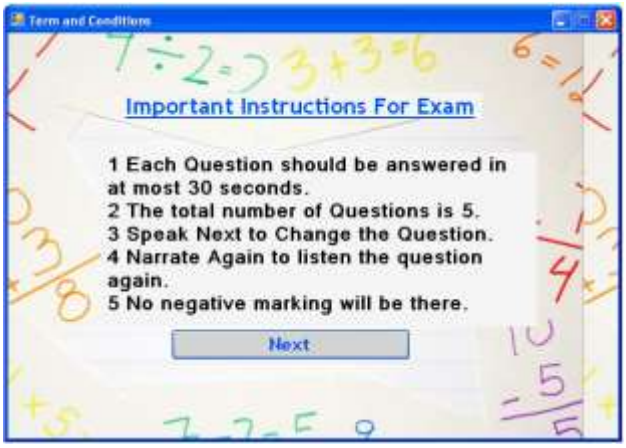


Figure 17: Instructions and guidelines for the exam

When the student listens the instruction and speaks Next, the test starts. The system reads aloud the questions and options, the student in turns speaks the option for answer. If the candidate wants to skip a certain question, he can simply go to the next one by speaking 'Next Question' or to the previous one by saying 'Previous Question'. The sytem is having the facility to tell the status of the left out time also.

Figure 18: Displaying the voice enabled exam screen

After the completion of the test, the system will announce the candidates score to him. Also, the system will generate a print of his score card and will mail a copy of the same to his e-mail id also. When the test has been completed, the system will encrypt the candidate's score and store it in the database.

8. FUTURE SCOPE

The applications of speaker recognition technology are quite varied and continually growing. Below is an outline of some broad areas where speaker recognition technology has been or is currently used.

8.1. Access Control:

Originally for physical facilities, more recent applications are for controlling access to computer networks (add biometric factor to usual password and/or token) or websites (thwart password sharing for access to subscription sites). Also used for automated password reset services.

8.2. Transaction Authentication:

For telephone banking, in addition to account access control, higher levels of verification can be used for more sensitive transactions. More recent applications are in user verification for remote electronic and mobile purchases (e- and m-commerce).

8.3. Law Enforcement:

Some applications are home-parole monitoring (call parolees at random times to verify they are at home) and prison call monitoring (validate inmate prior to outbound call). There has also been discussion of using automatic systems to corroborate aural/spectral inspections of voice samples for forensic analysis.

8.4. Speech Data Management:

In voice mail browsing or intelligent answering machines, use speaker recognition to label incoming voice mail with speaker name for browsing and/or action (personal reply). For speech skimming or audio mining applications, annotate recorded meetings or video with speaker labels for quick indexing and filing.

8.5. Personalization:

In voice-web or device customization, store and retrieve personal setting/preferences based on user verification for multi-user site or device (car climate and radio settings). There is also interest in using recognition techniques for directed advertisement or services, where, for example, repeat users could be recognized or advertisements focused based on recognition of broad speaker characteristics (e.g. gender or age).

8.6. Aids for the disabled:

One of the longest-established applications of TTS synthesis is in reading machines for the blind. The first such machine, combining an optical character reader with a TTS synthesizer, was produced by Kurzweil Computer Products in the 1970s. Even now, this speech synthesis task is very difficult as the machine must cope with any arbitrary text, and the quality of the speech that is generated would be regarded as insufficient by many people. However, these systems provide the visually impaired with the facility to read text that would not otherwise be available to them.

8.7. Remote e-mail readers:

A specialized but very useful application of TTS synthesis is to provide remote access to e-mail from any fixed or mobile telephone. For an e-mail reader, a full TTS conversion facility is required because the messages may contain any text characters.

E-mail messages are often especially challenging, due to the tendency to errors of spelling and grammar as well as the special nature of the language, abbreviations and so on that are often used. There are also many formatting features that are specific to e-mail.

9. CONCLUSION

The system has led to the development of a voice enabled examination system, as a tool for giving voice enabled exam. The system has been designed keeping in view the requirements of visually impaired students to aid them to keep pace with ordinary people in the field of education.

The testing application was based on Text-to-Speech(TTS) and Speech-to-Text(STT) technology which was implemented using Microsoft Speech API(SAPI). This web based academic testing software would provide an interaction medium for blind or partially sighted students to enhance their educational experiences.

While designing the software, the developers have used the natural voice that read aloud the questions in the test which the blind students have to answer and taken care of the accuracy of the synthetic pronunciation so as to provide the appropriate answer to the question.

The developers have made the use of Speech Application Programming Interface (SAPI) which act as an interface between applications and speech technology engines, both text-to-speech and speech recognition.

Many assistive technologies was present but no such system was offering an easy, accessible and intelligible interaction of the visually challenged with the computer. This system will prove to be an indispensable tool.

Incorporating and implementing the accessibility technology during the development of this testing application provided numerous advantages to the sighted users, thereby giving the test more efficiently. Thus voice enabled examination system, is a vital technology that can be beneficial for all types of users.

10. REFERENCES

[1] R. Klevansand, R. Rodman, “Voice Recognition, Artech House, Boston, London,1997.

[2] Samudravijaya K. Speech and Speaker recognition tutorial TIFR

[3] http://www.w3.org/TR/speech-synthesis/

[4] Nurulisma Ismail and Halimah Badioze Zaman, Search Engine Module in Voice Recognition Browser to Facilitate the Visually Impaired in Virtual Learning (MGSYS VISI-VL), World Academy of Science, Engineering and Technology, Volume 71, 2010.

[5] Michael Koerner, 1996, Speech Recognition: The Future Now, Prentice Hall Professional Technical Reference, 306.

[6] GIN-DER WU AND YING LEI “ A Register Array based Low power FFT Processor for speech recognition” Department of Electrical engineering national Chi Nan university Puli ,545 Taiwan

[7] .Dutoit, T. (1999). A short introduction to text-to-speech synthesis [Web page]. Mons: TCTS Lab, Faculté Polytechnique de Mons. Retrieved from http://tcts.fpms.ac.be/synthesis/introtts_old.html.

[8] http://en.wikipedia.org/wiki/Speech_synthesis- speech