Voice Customizable Text-To-Speech Conversion System for Dumb

Gaurav Kumar, E. Yazhini, Pankaj Agarwal

Abstract -The text-to-speech conversion system is a useful interface for the speechless who wants to communicate effectively in public places and official areas. This system uses the MATLAB software, wherein the text entered is converted into its corresponding speech signal using a graphical user interface development environment. A text database is created and a set of corresponding signals are recorded, both of which are interfaced to the MATLAB code. The speech signal undergoes framing and de-framing in order to convert a non-stationary signal into a stationary one. This is necessary because any signal processing tool takes into assumption that the signal it receives would be stationary in nature. The MATLAB platform has been used to reduce the cost of this system.

Keywords: text-to-speech, MATLAB code, corresponding signals, recorded, signal processing.

I. INTRODUCTION

Dumbness is the condition of being unwilling or unable to speak. The problems faced by the dumb people in India can be broadly classified into categories like social interaction, language and communication, education, behavioral problems, mental health, and safety concerns. Their problems have been worsened by the absence of a proper sign language training centre in India. Due to their inability to express themselves, they lack good job opportunities, participation in public events and are often sidelined by common people. There have been instances where dumb people have been rejected in interviews despite their competency in their respective fields. Even if they are placed in a good job, they are unable to effectively participate in meetings and official discussions. Though sign languages and facial expressions do help them to some extent, but it is not that effective in official places. In order to bring them out from this plight, the text-to-speech conversion system was designed to help them communicate with normal people easily. In this system, a graphical user interface development environment was used wherein a set of texts or individual words was created as a database. Their corresponding voice signals were also stored. Each voice signal was processed and sampled to remove the non-stationary nature of speech signals using framing and de-framing concept.

For every sentence entered, the MATLAB code will search for each word in the sentence from the database and give their corresponding voice signals as output. The main concern with this system is that it must be cost effective. Moreover, it must produce a faithful output, ie speech signal with minimum noise.

II. SYNTHESIS

Various problems may arise when dealing with systems and software. These are mentioned as below.

1. In signal processing, noise is a general term for unwanted (and, in general, unknown) modifications that a signal may suffer during capture, storage, transmission, processing, or conversion. Noise in signal leads to system fidelity. Generally, a system must be able to reproduce state and behavior of real world condition to a great extent. But, systems in the past didn't have the ability to produce faithful output. Additional noise used to be a major limiting factor giving inappropriate output. The manufacturer-specific algorithms are developed to acoustically analyze the incoming signal and alter the gain/output characteristics according to their predetermined rules.

2. System developers may make incorrect assumptions about the requirements of certain modules. It may also be made on the part of the entity requiring the software system. A software developer should know all the prospective of a system and he/she must also know the direction in which the system is going and the problems that are going to get solved in the coming future. Moreover, he/she must take into account that the system is adaptable to the future developments. A poor user interface may cause significant problems for users of the system and thus greatly increase the likelihood of those users introducing to the system data for example that causes system failure.

3. A software developer must keep in mind that the system is implemented on a good user-interface so that users may not face problem while using it. If people can't figure out how the application works, they'll get confused. Over-clarification must be avoided to make it concise. It should work fast and should appear as if people have encountered it before. It must be consistent so that it allows users to develop good usage patterns. It must be attractive and appealing to the user. It should also be able to perform functions faster and with less effort. The interface must also be forgiving to save users from costly mistakes.

4. System developers end up designing a system which is very expensive. A good design is one which is cost-effective and at the same time it must produce the desired level of output. Various parameters must be kept in mind before constructing a system that is cost-effective and producing considerably good and faithful output. He/She must be able to decide upon a good alternative to cut-down the cost without affecting the system fidelity.

III. METHODOLOGY

A. Introduction about MATLAB software.

This system for the speechless was built on MATLAB platform. MATLAB is a multi-paradigm numerical computing environment and fourth-generation programming language. It is a proprietary programming language developed by Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages. It provides a platform to bulid a user-interface system using Graphical User Interface Development Environment. MATLAB was interfaced with the text database and their corresponding voice signals were stored. Both of these were interfaced using a user-interface system called the guide using MATLAB coding.

B. Text database

The system requires the creation of a database of words, in other words, a dictionary so that for every sentence entered by a dumb from the keyboard, the words in the sentence were picked up from the database. The database was created in Microsoft Excel where the individual words were contained in each cell. When this sheet was interfaced to the MATLAB, each word in a cell was stored as variables in the form of individual files. The reason behind choosing Excel was that it was easier to interface with MATLAB as compared to other database systems such as SQL.

C. Voice signals

Voice signals for various words were stored in the MATLAB software itself. Each word is stored in individual file. It was also required that the speech signals were sampled with appropriate sampling period to avoid system fidelity.

D. GUIDE (Graphical User Interface Development Environment)

The GUIDE facility in the MATLAB was used as a user-interface for the dumb to enter set of words to be converted to its corresponding speech signal. Moreover, it also consisted of facilities to extend database of texts and its corresponding voice signal and also the number of recordings up-to-date.

E. Speech signal processing

We first define the term what are stationary and non-stationary signals. A stationary signal is one in which the frequency remains constant from one time interval to another. On the other hand, a non-stationary signal is one in which the frequency changes in one or more time intervals.

In order to make it appear stationary, the signal is divided inot frames of say around 250ms.In case of a speech signal, it is divided into 20ms.Any signal processing tool or system takes into assumption that the signal it receives is stationary in nature. So it is required that a non-stationary undergo framing. Let us take the following illustration.
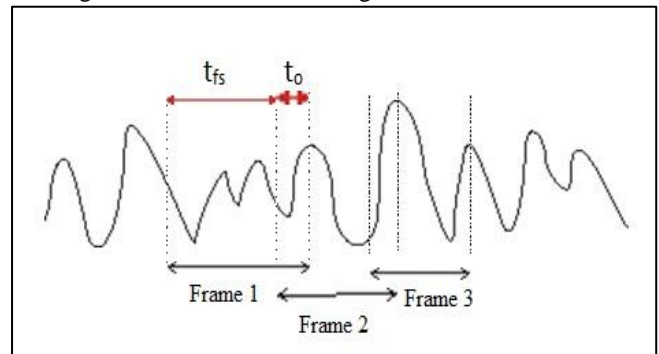
The above diagram shows a speech signal that is divided into frames. t_{fs} is the time between the start time of each frame. The overlap to is defined as the time from where a new frame starts until the current stop. $t_{fl} = t_{fs} + t_o$. The window has to be of length t_{fl} which corresponds to the $t_{fl}.f_s$ samples. After performing computations on each individual frame it is desirable to combine the frames into a coherent speech signal. This task contains two processes: taking account of the formerly used window and to combine the samples shared by different frames. A possible way of removing window is to multiply the frame by the reciprocal window. But before doing this it is essential to be cautious that the window is different from zero because it would prevent the number of possible windows. The common samples of two frames can be combined in such away that the closer the samples get to the edge of the frame the less weight they are given. The fig1.1 shows two functions averaging the beginning and end of frame. There is an overlap from t1 to t2 and from t3 to t4. The solid line is a half triangular window in each end and the dashed line is a half hanning window. The triangular window provides a constant amplitude but gives a point of discontinuity. On the other hand a hanning window doesn't provide a constant amplitude but it is continuous. Fig1.2 shows the function shown in fig1.1 divided by hamming window to compensate for the windowing performed by framing signals. Multiplication by one of these functions will de-window the signal and average the overlapping sections. This product can be converted to speech signal by joining the frames by addition in the overlapping zones.
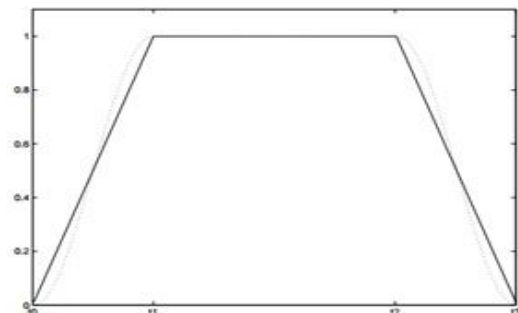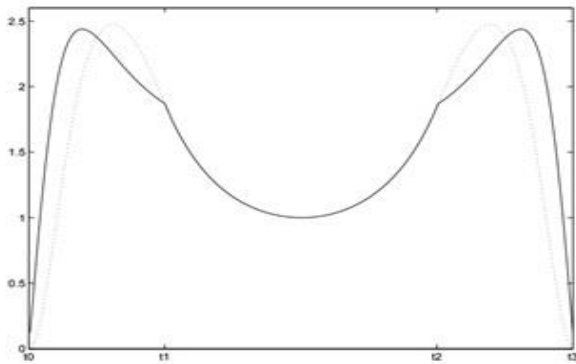

Fig 1.1

Fig 1.2

By the framing and de-framing procedure the noise was reduced to a great extent. In brief, the GUIDE was used as a platform in MATLAB in which, a person entered a sentence. The words in the sentence were picked up from the database and its corresponding voice signal was given as an output. So the system replicated what a dumb enters on the keyboard as speech.

IV.   RESULT AND DISCUSSIONS
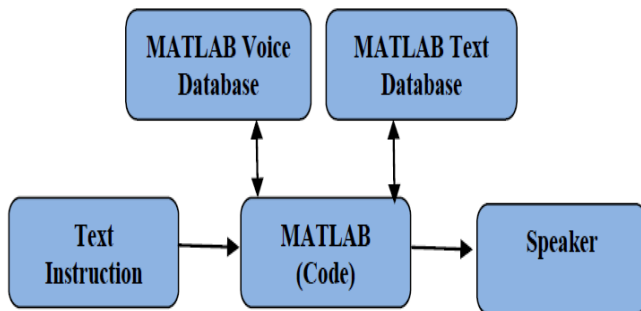
The voice customizable text-to-speech conversion system, produced a faithful speech signal as an output. This system didn't require any external interfacing components and expensive IC's. The MATLAB software along with Excel sheet as database was enough to produce the desired result, thereby making this system cost-effective. The GUIDE tool was used to provide a good user-interface for the dumb people. So the problem of poor interface with the user was solved. The speech signal was divided into frames to exhibit stationary nature of signals. All the frames were then de-framed to converge them into speech signals. The use of windows thus removed the noise, thereby producing faithful output.

REFERENCES

1. Yan-You Chen, Yu-Wei Bai1, Chun-Yu Tsai, Jhing-Fa Wang, Bo-Wei Che, IEEE Digital Library
2. Fredrik Berglund, Hans Jahanesson, Goran Gustafsson, Chalmers University of Technology, Goteborg   Sweden, Multidisciplinary project based product development.
3. Helenca Duxans I Barrobes, Dr. Antonio Bonafonte Cavez, Universitat Politechnica, Voice Conversion applied to text-to-speech systems.
4. K. Sreenivasa Rao, B. Yegnanarayana, IIT Guwahati, Voice Conversion and voice tract modification.