ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

TEXT-TO-SPEECH CONVERSION

[1]N.SWETHA [2]K..ANURADHA
[1]Department of Electronics and Communication Engineering
Sree Visvesvaraya institute of technology and science
nandhi.swetha@gmail.com
[2]Department of Electronics and Communication Engineering
Sree Visvesvaraya institute of technology and science
Mahabubnagar,India
kanuradha06@gmail.com

ABSTRACT

Text-to-speech (TTS) is the generation of synthesized speech from text. Our goal is to make synthesized speech as intelligible, natural and pleasant to listen, as human speech. Speech is the primary means of communication between people. During synthesis very small segments of recorded human speech are concatenated together to produce the synthesized speech.

The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood. A text-to-speech synthesizer allows people with visual impairments and reading disabilities to listen to written works on a home computer. Many computer operating systems have included speech synthesizers since the early 1990s.

Recent progress in speech synthesis has produced synthesizers with very high intelligibility but the sound quality and naturalness still remain a major problem. However, the quality of present products has reached an adequate level for several applications, such as multimedia and telecommunications.

The following thesis presents a brief overview of the main text-to-speech synthesis problems, and the initial work done in building a TTS in English.

Keywords: speech synthesis, syllable, phoneme ,concatenation, prosody.

1. INTRODUCTION

Language is the ability to express one's thoughts by means of a set of signs (text), gestures, and sounds. It is a distinctive feature of human beings, who are the only creatures to use such a system. Speech is the oldest means of communication between people and it is also the most widely used.

'Speech synthesis' also called 'Text to speech synthesis' is the artificial production of human speech. A computer system used for this purpose is called a **speech synthesizer** and can be implementd in software. A **text-to-speech (TTS)** system converts text to speech.

At first sight, this task does not look too hard to perform. After all we all have a deep knowledge of reading rules of our mother tongue. They were transmitted to us, in a simplified form, at primary school, and we improved them year after year. But in the context of TTS synthesis, it is impossible to record and store all the words of the language. Some other method has to be implemented for this purpose.

The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood. A text-to-speech synthesizer allows people with visual impairments and reading disabilities to listen to written works on a home computer. Many computer operating systems have included speech synthesizers since the early 1990s.

Astro- physician Stephen Hawkins, who is completely paralyzed, gives all his lectures using a TTS system.

This project gives an idea about developing a pc based text-to speech synthesizer using MATLAB.

2. SPEECH SYNTHESIS

2.1 WHAT IS SPEECH SYNTHESIS?

A Text-To-Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud, when it is directly introduced in the computer by an operator.

It is more suitable to define Text-To-Speech or speech synthesis as an automatic production of

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

speech, by 'grapheme to phoneme' transcription. A **grapheme** is the smallest distinguishing unit in a written language. It does not carry meaning by itself. Graphemes include alphabetic letters, numerical digits, punctuation marks, and the individual symbols of any of the world's writing systems. A **phoneme** is "the smallest segmental unit of sound employed to form meaningful utterances".

2.2 PHONETICS

In most languages the written text does not correspond to its pronunciation. so that in order to describe correct pronunciation some kind of symbolic presentation is needed. Every language has a different phonetic alphabet and a different set of possible phonemes and their combinations. The number of phonetic symbols is between 20 and 60 in each language. "A set of phonemes can be defined as the minimum number of symbols needed to describe every possible word in a language". In English there are about 44 phonemes. Due to complexity and different kind of definitions, the number of phonemes in English and most of the other languages cannot be defined exactly.

Phonemes are abstract units and their pronunciation depends on contextual effects, speaker's characteristics, and emotions. During continuous speech, the articulator movements depend on the preceding and the following phonemes. The articulators are in different position depending on the preceding one and they are preparing to the following phoneme in advance. This causes some variations on how the individual phoneme is pronounced. These variations are called allophones which are the subset of phonemes and the effect is known as coarticulation. For example, a word *lice* contains a light /l/ and *small* contains a dark /l/. These l's are the same phoneme but different allophones and have different vocal tract configurations. Another reason why the phonetic representation is not perfect, is that the speech signal is always continuous and phonetic notation is always discrete

The phonetic alphabet is usually divided in two main categories, vowels and consonants. Vowels are always voiced sounds and they are produced with the vocal cords in vibration, while consonants may be either voiced or unvoiced. Vowels have considerably higher amplitude than consonants and they are also more stable and easier to analyze and describe acoustically. Because consonants involve very rapid changes they are more difficult to synthesize properly.

Few examples of different phonetic notations are shown below.

IPA	EXAMPLE
i:	b*ee*t
I	b*i*t
ɛ	b*e*t
æ	*a*t
ə	*a*bout
ʌ	b*u*t

2.3 SYNTHESIZER TECHNOLOGY

The most important qualities of a speech synthesis system are *naturalness* and *intelligibility*. Naturalness describes how closely the output sounds like human speech, while intelligibility is the ease with which the output is understood. The ideal speech synthesizer is both natural and intelligible. Speech synthesis systems usually try to maximize both characteristics.

The primary technology for generating synthetic speech is concatenative synthesis.

2.3.1 Concatenative synthesis

Concatenative synthesis is based on the concatenation (or stringing together) of segments of recorded speech.

Connecting prerecorded natural utterances is probably the easiest way to produce intelligible and natural sounding synthetic speech. However, concatenative synthesizers are usually limited to one speaker and one voice and usually require more memory capacity.

One of the most important aspects in concatenative synthesis is to find correct unit length. The selection is usually a trade-off between longer and shorter units. With longer units' high naturalness, less concatenation points are achieved, but the amount of required units and memory is increased. With shorter units, less memory is needed, but the sample collecting and labeling procedures become more difficult and complex. In present systems units used are usually words, syllables, phonemes.

2.3.2 Domain-specific synthesis

Word is perhaps the most natural unit for written text and some messaging systems with very limited vocabulary. Concatenation of words is relative easy to perform**.**

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

Domain-specific synthesis concatenates prerecorded words and phrases to create complete utterances. It is used in applications where the system's output is limited to a particular domain, like transit schedule announcements or weather reports. The technology is very simple to implement, and has been in commercial use for a long time, in devices like talking clocks and calculators. The level of naturalness of these systems can be very high because the variety of sentence types is limited, and they closely match the prosody and intonation of the original recordings.

Because these systems are limited by the words and phrases in their databases, they are not general-purpose and can only synthesize the combinations of words and phrases with which they have been preprogrammed. However, there is a great difference with words spoken in isolation and in continuous sentence which makes the continuous speech to sound very unnatural. Because there are hundreds of thousands of different words and proper names in each language, it is quite clear that we cannot create a database of all words and common names in the world and so word is not a suitable unit for any kind of unrestricted TTS system.

Thus, for unrestricted speech synthesis (text-to-speech) we have to use shorter pieces of speech signal, such as syllables, phonemes or even shorter segments.

Phonemes are probably the most commonly used units in speech synthesis because they are the normal linguistic presentation of speech. The inventory of basic units is usually between 40 and 50, which is clearly the smallest compared to other units. Using phonemes gives maximum flexibility.

2.4 CREATION OF DATABASE

There are various critical factors to be considered while designing a TTS system that will produce intelligible speech. The first crucial step in the design of any TTS system is to select the most appropriate units or segments of speech that result in smooth utterance.

Building the unit inventory consists of three main phases. First, the natural speech must be recorded so that all used units (phonemes) within all possible contexts (allophones) are included. After this, the units must be labeled from spoken speech data, and finally, the most appropriate units must be chosen. Gathering the samples from natural speech is usually very time-consuming. The implementation of rules to select correct samples for concatenation must also be done very carefully.

The voice which is recorded manually contains some delay. This causes a greater time lapse between two consecutives utterances. This makes the speech a bit unpleasant and unnatural to listen. Hence there is a need to remove this delay.

Let us now get to know the algorithms used at different stages of implementation of text to speech conversion.

3. IMPLEMENTATION ALGORITHMS

3.1. CHARACTER –TO- VOICE:

Let us start text to speech synthesis with a simple character to voice conversion. The database required for character to voice conversion is recorded alphabets(a-z),digits(0-9) in the form of wave files(.wav).

The next step in converting text to speech is to create a text file(.txt). once the file is created, it is opened and read in matlab.

In matlab all the data is stored in the form of a matrix. For every element read, corresponding wave file is played so as to output the sound of that character. we can read as big file as possible but only character wise.

ALGORITHM

STEP1: Create a database of various wave files

STEP2: Create a text file (.txt)

STEP3: Open the .txt file in matlab.

STEP4: Read the file opened.

STEP5: For every character read, play the corresponding wave(.wav) file.

But as said earlier, there exist some delay by default while recording a sound. This delay has to be removed to get a continuous utterance of speech.

Fig 3.1 shows the plot of sound recorded manually and the default delay. This delay is removed and sound file is re-written. Fig 3.2 shows the plot of the sound after the delay is removed.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

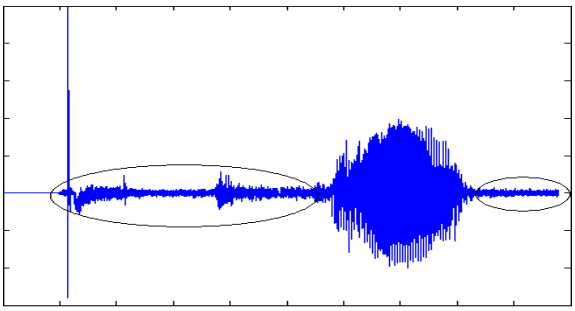

Fig 3. 1: sound with default delay

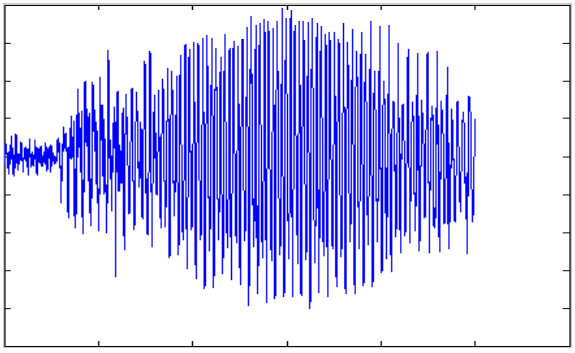

Fig 3.2: sound after the delay is removed.

3.2 WORD PRONUNCIATION

Character to voice is not a really big task. This is because there are only 26 characters in English and each character has a unique pronunciation. However when we have to read lengthy texts, character to voice is not recommended at the user level ,as it is difficult to make out a word from the characters read.

As we have played the wave file corresponding to every character read, in character to voice conversion, we can also play the wave file for every word read. But practically it is impossible to record all the words of a dictionary. Hence there is a need to think of some other alternative. In the first attempt to play a word as a whole we can think of playing syllables of a word.

SYLLABLE:

Syllable is a part of a word that contains a single vowel sound and that is pronounced as a unit. For example “book” has one syllable, and “reading” has two syllables. The word “syllable” itself has three syllables ‘syl-la-ble’.

There are 44 pronunciation sounds (Phonemes) of which 22 sounds are of vowels, and 22 are of consonants. The database of all the pronunciation sounds is created. Once the text is read, for every syllable the corresponding wave files are concatenated and played.

Example:

Word: mahabubnagar

Syllables: ma-ha-b-u-b-na-ga-r

When the above word is played it looks like stammering and it becomes difficult to make out what is played. This is because every syllable is played separately and it may not be in continuation with the previous sound played. Hence choosing sound unit with proper length is important, so that the word is natural and understandable when synthesized.

ALGORITHM

STEP1: create a database of pronunciation sounds (phonemes).

STEP2: create a text file.

STEP3: open the file.

STEP4: read the file.

STEP5: concatenate the .wav files accordingly and play them.

4. PROBLEMS IN SPEECH SYNTHESIS

The problem area in speech synthesis is very wide. There are several problems in text pre-processing, such as numerals, abbreviations, and acronyms. This chapter describes the major problems in text-to-speech research.

4.1 Text-to-Phonetic Conversion

The first task faced by any TTS system is the conversion of input text into linguistic representation, usually called text-to-phonetic or grapheme-to-phoneme conversion. The difficulty of conversion is highly language depended and includes many problems. In some languages, such as Hindi or Telugu, the conversion is quite simple because written text almost corresponds to its pronunciation. For English and most of the other languages the conversion is much more complicated. A very large set of different rules and their exceptions is needed to produce correct pronunciation and prosody for synthesized speech.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

4.2 Text preprocessing

Text preprocessing is usually a very complex task and includes several language dependent problems. Digits and numerals must be expanded into full words. For example in English, numeral 243 would be expanded as *two hundred and forty-three* and 1750 as *seventeen-fifty* (if year) or *one-thousand seven-hundred and fifty* (if measure). Fractions and dates are also problematic. 5/16 can be expanded as *five-sixteenths* (if fraction) or *May sixteenth* (if date). Same kind of contextual problems are faced with roman numerals. Chapter III should be expanded as *Chapter three* and Henry III as *Henry the third* and *I* may be either a pronoun or number.

For example kg can be either *kilogram* or *kilograms* depending on preceding number, St. can be *saint* or *street.* In some cases, the adjacent information may be enough to find out the correct conversion, but to avoid misconversions the best solution in some cases may be the use of letter-to-letter conversion.

Special characters and symbols, such as '$', '%', '&', '/', '-', '+', cause also special kind of problems. In some situations the word order must be changed. For example, *$71.50* must be expanded as *seventy-one dollars and fifty cents* and *$100 million* as *one hundred million dollars*, not as *one hundred dollars million.*. Some languages also include special non ASCII characters, such as accent markers or special symbols.

4.3 Pronunciation

The second task is to find correct pronunciation for different contexts in the text. Some words, called *homographs*, cause the most difficult problems in TTS systems. Homographs are spelled the same way but they differ in meaning and usually in pronunciation (e.g. fair, lives). The word *lives* is for example pronounced differently in sentences "Three *lives* were lost" and "One *lives* to eat". Some words, e.g. *present*, has different pronunciations depending on the context. (I was *present* there when he received the *present*).

The characters 'th' in ‘mother’ and ‘think’ is pronounced differently. Some sounds may also be either voiced or unvoiced in different context. For example, phoneme /s/ in word *dogs* is voiced, but unvoiced in word *cats* .

Finding correct pronunciation for proper names, especially when they are borrowed from other languages, is usually one of the most difficult tasks for any TTS system. Unfortunately, it is clear that there is no way to build a database of all proper names in the world.

4.4 Prosody

Finding correct intonation, stress, and duration from written text is probably the most challenging problem for years to come. These features together are called prosodic features and may be considered as the melody, rhythm, and emphasis of the speech at the perceptual level. The intonation means how the pitch pattern or fundamental frequency changes during speech. The prosody of continuous speech depends on many separate aspects, such as the meaning of the sentence and the speaker characteristics and emotions .Unfortunately, written text usually contains very little information of these features and some of them change dynamically during speech. However, with some specific control characters this information may be given to a speech synthesizer.

If there is no breath pauses in speech or if they are in wrong places, the speech may sound very unnatural or even the meaning of the sentence may be misunderstood. For example, the input string "John says Peter is a liar" can be spoken as two different ways giving two different meanings as "John says: Peter is a liar" or "John, says Peter, is a liar". In the first sentence Peter is a liar, and in the second one the liar is John.

5. APPLICATIONS OF SYNTHETIC SPEECH

Synthetic speech may be used in several applications. some applications, such as reading machines for the blind or electronic-mail readers, require unlimited vocabulary and a TTS system is needed.

The application field of synthetic speech is expanding fast whilst the quality of TTS systems is also increasing steadily. Speech synthesis systems are also becoming more affordable for common customers, which makes these systems more suitable for everyday use. For example, better availability of TTS systems may increase employing possibilities for people with communication difficulties.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

Applications for the Blind

Probably the most important and useful application field in speech synthesis is the reading and communication aids for the blind. Before synthesized speech, specific audio books were used where the content of the book was read into audio tape. It is clear that making such spoken copy of any large book takes several months and is very expensive. It is also easier to get information from computer with speech instead of using special bliss symbol keyboard, which is an interface for reading the Braille characters.

A blind person can not also see the length of an input text when starting to listen it with a speech synthesizer, so an important feature is to give in advance some information of the text to be read. For example, the synthesizer may check the document and calculate the estimated duration of reading and speak it to the listener. Also the information of bold or underlined text may be given by for example with slight change of intonation or loudness.

Applications for the Deafened and Vocally Handicapped

People who are born-deaf can not learn to speak properly and people with hearing difficulties have usually speaking difficulties. Synthesized speech gives the deafened and vocally handicapped an opportunity to communicate with people who do not understand the sign language.

Educational Applications

Synthesized speech can be used also in many educational situations. A computer with speech synthesizer can teach 24 hours a day and 365 days a year. It can be programmed for special tasks like spelling and pronunciation teaching for different languages. It can also be used with interactive educational applications.

Applications for Telecommunications and Multimedia

The newest applications in speech synthesis are in the area of multimedia. Electronic mail has become very usual in last few years. However, it is sometimes impossible to read those E-mail messages when being for example abroad. There may be no proper computer available or some security problems exist. With synthetic speech e-mail messages may be listened to via normal telephone line. Synthesized speech may also be used to speak out short text messages (sms) in mobile phones.

Other Applications

In principle, speech synthesis may be used in all kind of human-machine interactions. For example, in warning and alarm systems synthesized speech may be used to give more accurate information of the current situation. Using speech instead of warning lights or buzzers gives an opportunity to reach the warning signal for example from a different room

6. IMPLEMENTATION, TESTING AND RESULTS

Character-voice

The first step of our project is to create the required database. Initially, we have started with recording alphabets and numerical digits in a mobile and then converted them to wave format (.wav) in a pc. All the .wav files are labeled appropriately.

Testing and problems encountered:

A text file with few characters is created. The file is then opened and read. For every character read, corresponding wave file is played. Initially these sound files are played separately.

The plot of the sound files of characters when played individually is shown in Fig 6.1, Fig 6.2 below.

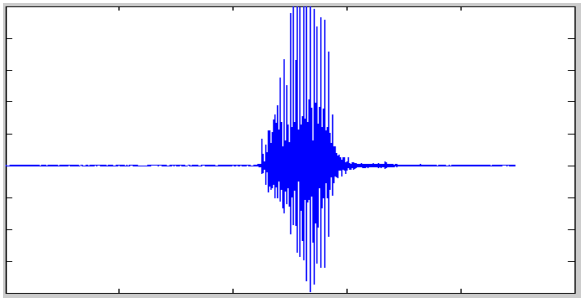

Fig 6.1: Plot of the sound ‘a’

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

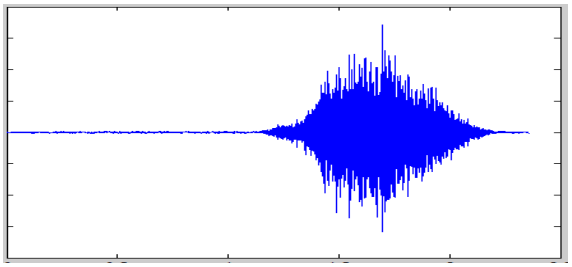

Fig 6.2: Plot of the sound 's'

The Fig 6.3 shows the plot of the sound, when 'a','s' are concatenated and played together.

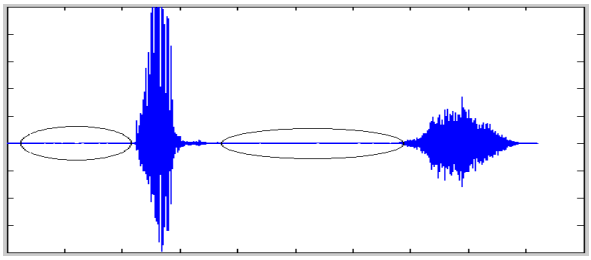

Fig 6.3: Plot of 'as'

But from the above figure, it is quite clear that there exists a large time lapse between two consecutive sounds.it sounds like 'a s' and not 'as'. The reason for this time lapse is the default delay that is present during recording the sound file.

Amendments and results

This problem is eliminated by removing the delay in .wav files (using matlab) and rewriting the sound file (without delay).when the text file is now played, the wave files corresponding to the characters read were played continuously without delay.

The following figures show the plot without delay.

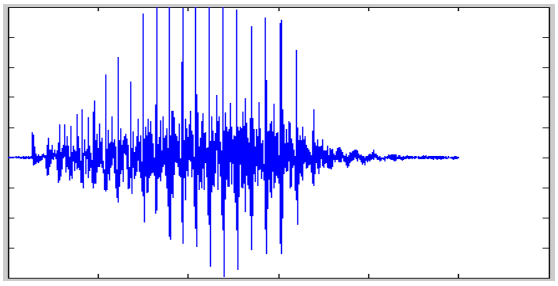

Fig 6.4: Plot of 'a' without delay

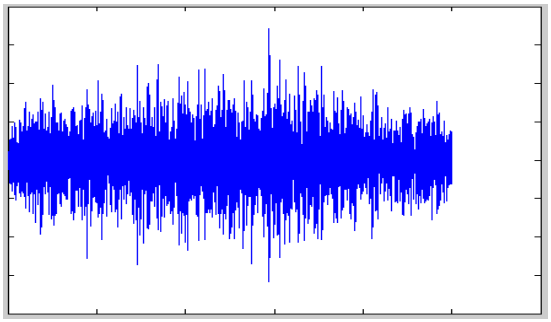

Fig 6.5: Plot of 's' without delay

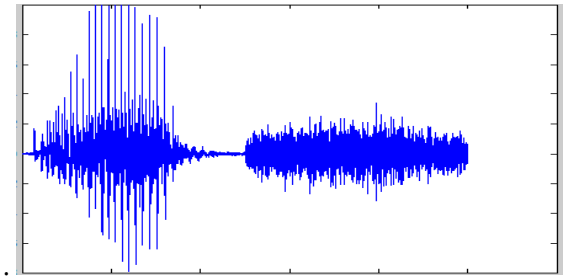

Fig 6.6: Plot of 'as' without delays

When the sounds without delays are concatenated and played, it now sounds like 'as' and not 'a s' . During character to voice conversion, it is observed that, the better the delays are removed ,the natural the output sounds.

WORD-VOICE

The next idea was to play the word as a whole. For this, we started thinking about pronunciation sounds. All the 44 sounds were recorded and converted into wave format. As the first step to play the word, we started playing individual sounds of the word separately and then concatenated them to form a word.

Testing and problems encountered

Here is an example to play the sounds of the word 'sky' separately .

Fig 6.7 shows the plot of the phoneme 's'. When the corresponding wave file is played, it sounds like 'sa' and not 's'.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

Fig 6.8 shows the plot of the phoneme 'k'. When the corresponding wave file is played, it sounds like 'ka' and not 'k'.

Fig 6.9 shows the plot of the phoneme 'aI'.

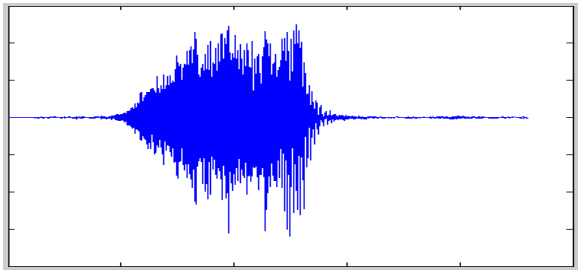

Fig 6.7: Plot of the syllable 's'

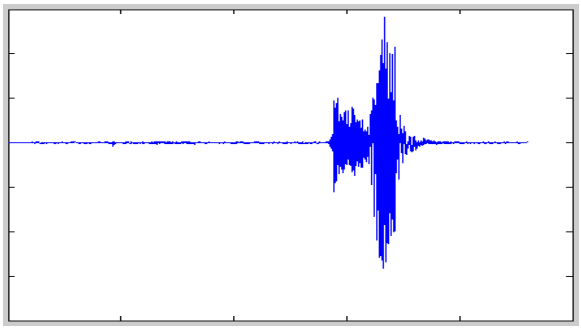

Fig 6.8: Plot of the syllable 'k'

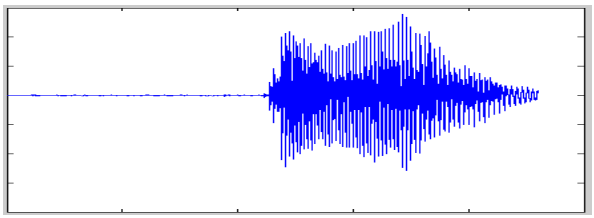

Fig 6.9: Plot of the sound ' aI'

These sounds are concatenated to get the word 'sky'. The following figure 6.10 shows the plot the word 'sky'.

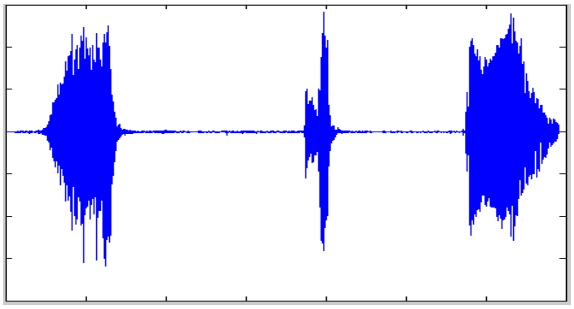

Fig 6.10: Plot of the word 'sky'

But from the above figure it is clear that, it sounds as "sa *ka ai*" and not '*sky*'. This is because the length of the sounds is by default too large than required.Hence, the length of the corresponding wave files is edited to get the desired sound. All the corrections made and results obtained are shown below**.**

Amendments and results:

Fig 6.11 shows the plot of the shortened sound of phoneme 's'. When the .wav file corresponding to the phoneme 's' played, it now sounds as 's' and not 'sa'.

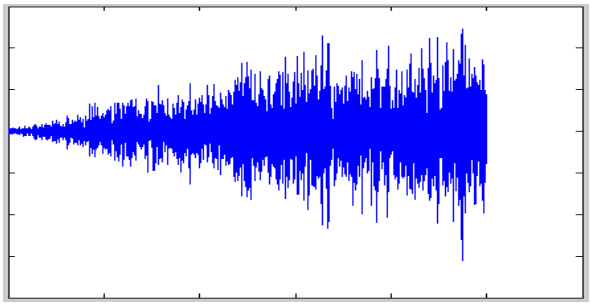

Fig 6.11: Plot of the syllable 's' whose length is shortened.

Fig 6.12 shows the plot of the shortened sound of the phoneme 'k'. When this .wav file corresponding to the syllable 'k' is now played, it sounds as 'k' and not ka'.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

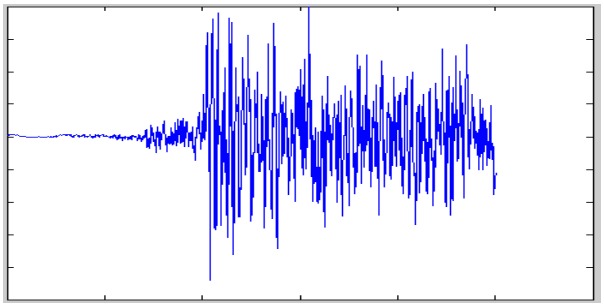

Fig 6.12: Plot of the syllable 'k' whose length is shortened.

Fig 6.13 shows the plot of the shortened sound of the syllable 'aI'.

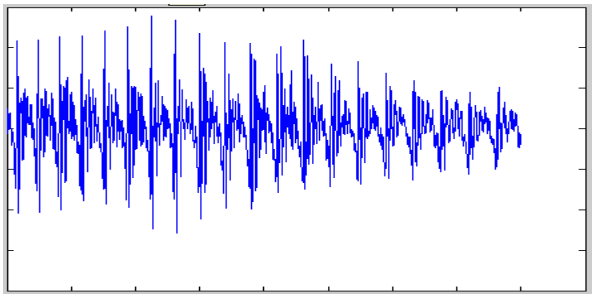

Fig 6.13: Plot of the sound 'aI' whose length is shortened

Now all the above shortened sounds without delay are concatenated and played. It now sounds as 'SKY' ,which is desired. Fig 6.14 shows the plot of the sound 'sky'.

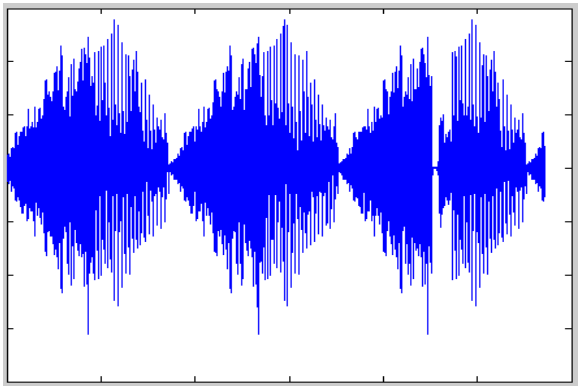

Fig 6.14: Plot of the sound 'sky'

READING STREAM OF WORDS WITH DELIMITERS:

Now, we tried of playing continuous stream of some random words with only one syllable each such as *"ash sky my school mass."*. String of characters until a delimiter (space, comma,fullstop) is encountered is recognized as a word. Fig6.15 shows the plot of the above mentioned stream of words.

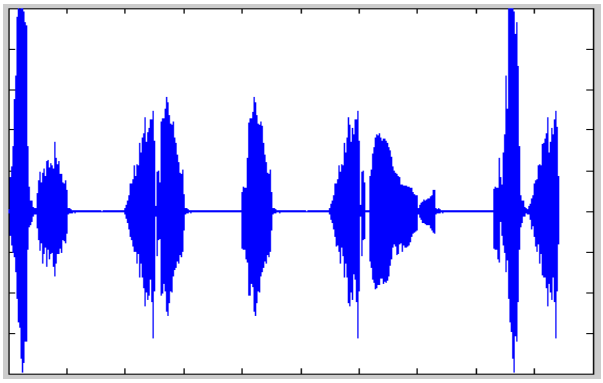

Fig 6.15: plot of the sound when 'ash sky my school mass.' are played.

When we are successful in playing a word correctly, then stream of few words can be easily played. And the results obtained were quite satisfactory.

7. CONCLUSIONS

7.1. Advantages

- Speech synthesis is advantageous for people who are visually handicapped. It helps them listen to the written works.
- People who are vocally handicapped can communicate with people who do not understand the sign language.
- Children may find this interesting to learn the pronunciation of words.
- TTS system can be used in domain specific applications such as train announcements.

7.2. Drawbacks

- Sound quality and naturalness is lacking.
- Concatenating the sounds depending on the word.
- Creating the logic for pronunciation of all the words of a dictionary accurately is difficult.
- High accuracy is difficult to achieve.

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.6, Pages : 269-278 (2013)
Special Issue of ICETEM 2013 - Held on 29-30 November, 2013 in Sree Visvesvaraya Institute of Technology and Science, Mahabubnagar – 204, AP, India

7.3. FUTURE SCOPE

We have seen that delay in sound wave causes the speech to look very unnatural. Removing the delay manually every time may be a tedious job. To overcome this problem, we can think of a method which recognizes the delay and automatically remove it. And the best method would be 'integration'. If we can integrate the synthesized speech, we can not only avoid delay but also get a continuous flow of speech. In this project we have worked only on text documents. Further we can think of reading word files, scanned data, PDF files etc.

REFERENCES

1.www.ims.uni-stuttgart.de/~moehler/synth**speech**/

2.http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/contents.html

3.http://books.google.co.in/books/about/An_Introduction_to_Text_To_Speech_Synthe.htm**L**

4.http://www.abelard.org/briefings/phonetic_chart_british_english.php