



Project

Text to Speech Conversion

INTRODUCTION

Speech is one of the oldest and most natural means of information exchange between human. Over the years, Many attempts have been made to develop vocally interactive computers in order to achieve voice/speech synthesis. Text-to-Speech Synthesis is a technology that converts written text from a descriptive form to a spoken language that the end user can understand. It is a computer based system that can be able to generate synthesized speech from text and read it aloud. A text-to-speech synthesizer allows people with visual impairments and reading disabilities to listen to written text on a home computer. It is also very helpful for kids and adults who struggle with reading. Great Astro- physician Stephen Hawkins, who was completely paralyzed, gives all his lectures using a TTS system.



Text to Speech

Objectives of the Project

The objective of our project is to implement an isolated whole word speech synthesizer that is capable of converting word and responding with speech. It consists of two main phases. The first is text analysis, in which the input text is converted into a phonetic or other linguistic representation, and the second is speech waveform generation, in which the output is generated from this phonetic information.



SPEECH SYNTHESIS

Speech synthesis is defined as the automatic creation of speech by 'grapheme to phoneme' translation. A grapheme is the smallest differentiating unit of written language. It has no meaning on its own. Alphabetic letters, numerical digits, punctuation marks, and individual symbols from any of the world's writing systems are all examples of graphemes. A phoneme is defined as "the smallest segmental unit of sound used to create meaningful speech. There are different ways to perform speech synthesis. The choice depends on the task they are used for, but the most widely used method is Concatenative Synthesis, because it generally produces the most natural-sounding synthesized speech. Concatenative synthesis done by concatenating pieces of recorded speech that are stored in a database of computer.

What is Phonemes?

A set of phonemes is the smallest amount of symbols required to describe every possible word in a language. The written text in most languages does not correspond to the pronunciation. As a result, some type of symbolic presentation is required to represent precise pronunciation. Every language has a unique phonetic alphabet as well as a unique set of phonemes and their combinations. In English there are about 44 phonemes. Out of these 44 phonemes - 22 sounds are of vowels, and 22 are of consonants. We created database for all these pronunciation sounds in form of .wav files. There is .wav files for each phoneme.

Phase one: Text to Phoneme conversion

Phonemic Chart Click on a symbol to hear its pronunciation

	monophthongs				diphthongs		
	i:	ɪ	ʊ	u:	ɪə	eɪ	
VOWELS	sheep	ship	good	shoot	here	wait	
	e	ə	ɜ:	ɔ:	ʊə	ɔɪ	əʊ
	bed	teacher	bird	door	cure	boy	show
	æ	ʌ	ɑ:	ɒ	eə	aɪ	aʊ
	cat	up	far	on	hair	my	cow
CONSONANTS	p	f	t	θ	tʃ	s	ʃ
	pea	fly	tea	think	cheese	see	shall
	k						
	b	v	d	ð	dʒ	z	ʒ
	boat	video	dog	this	June	zoo	television
	g						
	h	m	n	ŋ	r	l	w
	hat	man	now	singer	red	love	wet
	j						

Example of some phonemes -

Used In

- /f/



- Fan



- /s/



- Sip



- /v/



- Van



- /z/



- zip



Concatenation of segments of recorded speech

Phase Two: Phoneme to Speech conversion

For this step we use Concatenative synthesis method. We use Matlab software to read the word from input and generate audio signal wave for each phoneme contain in this word. Then we the concatenate each of these audio waves to produce final audio wave of word. Connecting prerecorded natural speech 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.

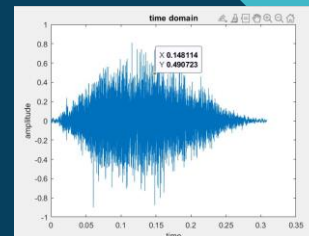
Implementation, Testing and Results

Audio wave of phonemes

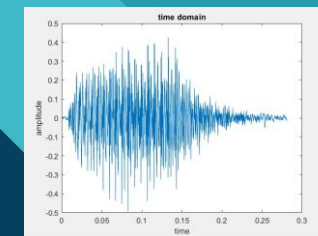
Here is a sample of our implementations –

➤ Input – **fish**

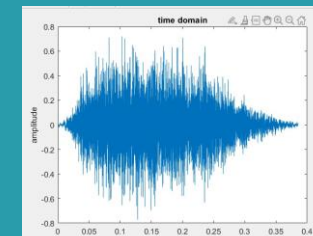
Word 'fish' have three phonemes - /f/, /i/, /sh/



/f/

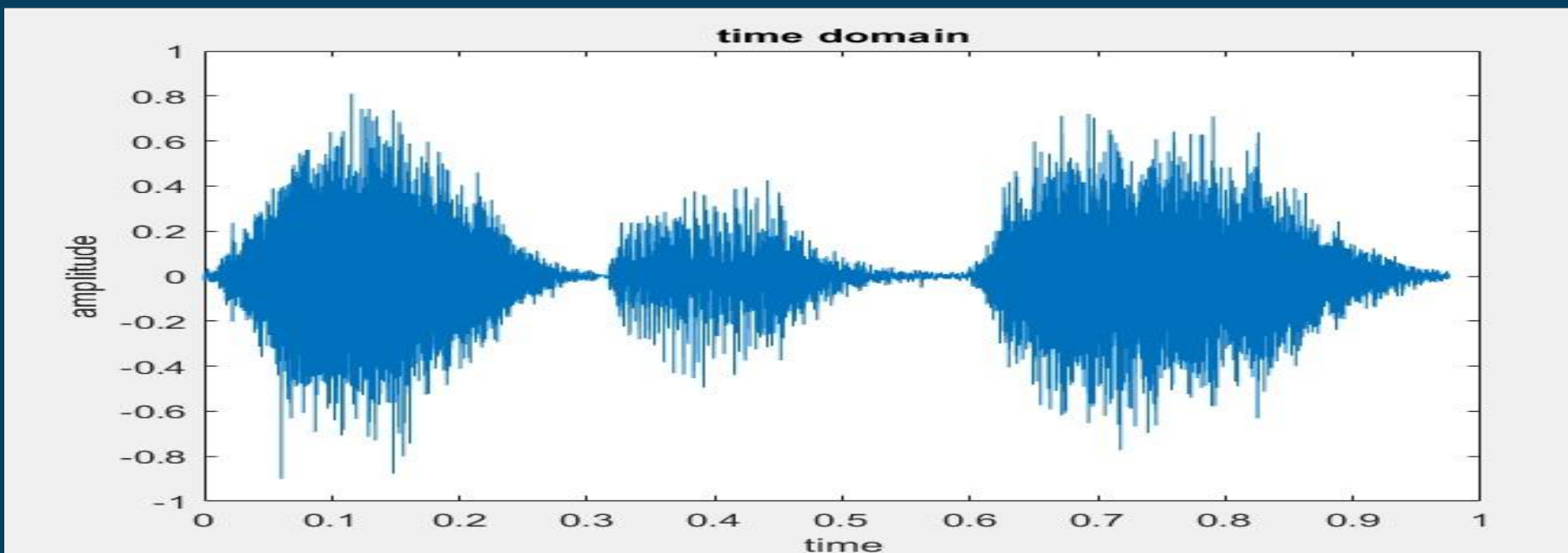


/i/

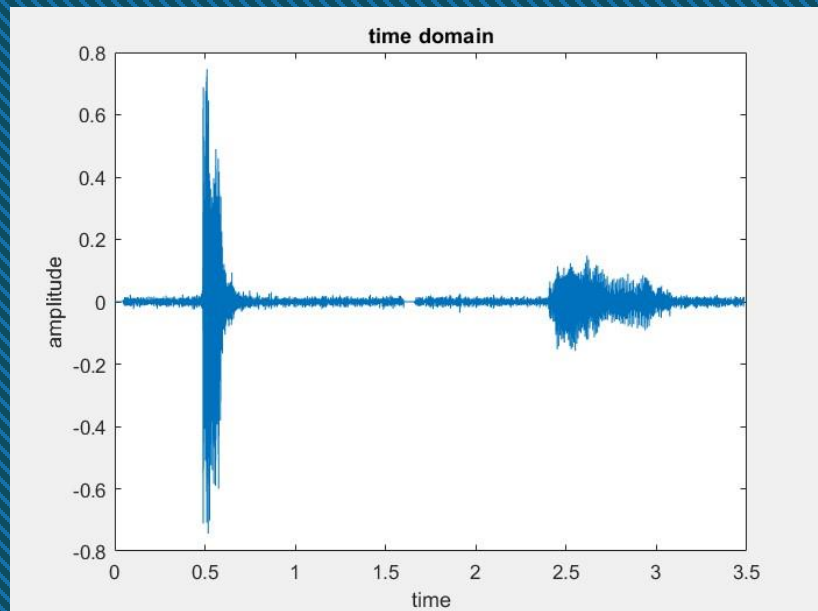


/sh/

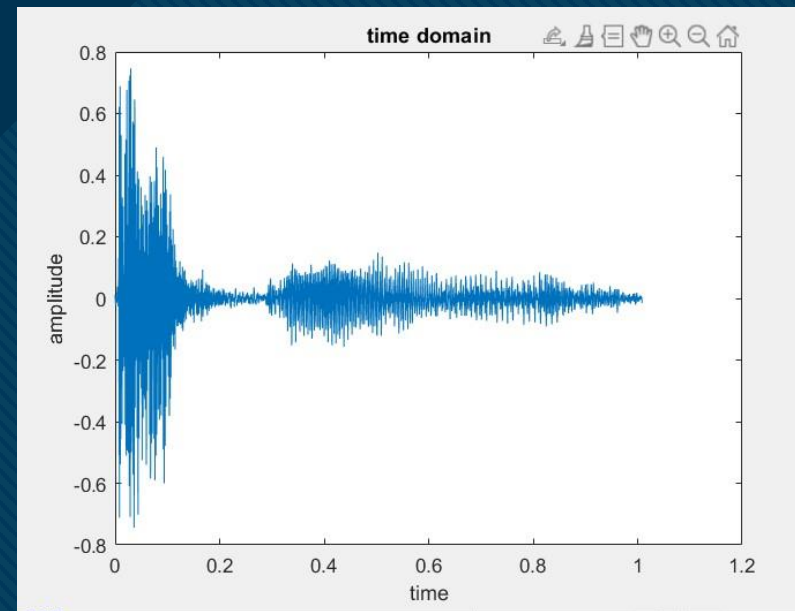
Concatenation result -



Before Modification



After Modification



ALGORITHM

How to run algorithm

- Run the main.m file in matlab.
- It will ask for user input.
- Enter an English letter having words >1.
- Press Enter key.
- Matlab program will run.
- Word what you entered will be played on speaker.
- A file named 'result.wav' will also be formed.
- 'result.wav' contains recording of the input word.

Software and source

- Software: Matlab is used for signal processing and signal synthesizing purposes.
- Algorithm is written in Matlab.
- Dataset: .wav files for 44 different phonemes of British English.



Drawbacks and Flaws in our implementation

- 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.



Thank You