Significance Of Paralinguistic Cues in Audio Rendering of Mathematical Content

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science by Research
in
COMPUTER SCIENCE

by

Venkatesh Potluri
201002175
venkatesh.potluri@research.iiit.ac.in

Speech and Vision Lab
International Institute of Information Technology
Hyderabad - 500 032, INDIA

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled Significance of Paralinguistic Cues in Audio Rendering of Mathematical Content by Venkatesh Potluri, has been carried out under my supervision and is not submitted elsewhere for a degree.

_____ _____

Date

Adviser: Dr. Kishore Prahallad
Dr. Priyanka Srivastava

To Family

Acknowledgments

I thank Dr. Kishore S Prahallad and Dr. Priyanka Srivastava for guiding me through to the completion of this project. They have helped me understand and properly progress through the different stages of the project. I thank Prof. Peri Bhaskararao for his contributions during the initial discussions related to developing ideas for this project. His suggestions equipped us with sufficient background knowledge to get started. I would also like to thank Dr. Radhika Mamidi for suggesting relevant literature. I thank my coauthors have helped me gain a sound understanding of the technical aspects involved in taking up this research. The participants have been extremely cooperative and patient while performing the listening tests. Vision Rehabilitation department, LV Prasad Eye Hospital has been very supportive in terms of providing their infrastructure and assisting me to identify the suitable set of visually impaired participants for the experiments. The speakers' experience was of great help in setting a baseline for my research. I must thank friends and family for their constant encouragement and support to take up research in the area of mathematical content and deal with content that is not accessible in this process. I thank the institute for providing me the opportunity, resources and a professional platform to get introduced to quality research. The educative environment at IIIT has instilled the desire to perform quality research and pursue further educative opportunities with a research component.

Abstract

Text To Speech (TTS) systems hold promise as information access tools for literate and illiterate including visually challenged. Current TTS systems can convert a typical text into a natural sounding speech. TTS is used in screen reader applications, IVR(Interactive Voice Response) systems, and other systems that output speech. TTS is therefore used to have text spoken. However, auditory rendering of mathematical content, specifically speaking equations and pie charts is not a trivial task. The nature of mathematical content requires them to be spoken in a way different from how traditional text is spoken. Mathematical equations have to be read so that appropriate bracketing such as parentheses, superscripts and subscripts are conveyed to the listener in an accurate way. Current implementations of TTS in screen readers cannot effectively render mathematical content in audio. Earlier works have attempted to use pauses as acoustic cues to indicate some of the semantics associated with the mathematical symbols. In this thesis, we first analyse the acoustic cues which human beings employ while speaking the mathematical content to (visually challenged) listeners. Speakers used linguistic and paralinguistic cues to speak the equations. The current thesis aims to analyse the significance of cues in synthesising mathematical content in audio. We propose four techniques which render the observed patterns in a text-to-speech system. The evaluation considered eight aspects such as listening effort, content familiarity, accentuation, intonation, etc. The effectiveness of the proposed techniques is gauged by performing an evaluation against people with and without vision impairment. Different methods for evaluation had to be followed with visually challenged participants. Our objective metrics show that a combination of the proposed techniques could render the mathematical equations using a TTS system as good as that of a human being. We employ the proposed technique to render statistical data in audio and measure the effectiveness of our implementation against people with vision impairment. Results show that the use of our technique increases the correctness by 30 %.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As humans, our ability to speak is innate to us. Technology has evolved to a great extent in the past 2 to 3 decades. A computer has evolved from a vacuum tube machine to a mobile, pocket sized one, and is now moving on to our body. There is a lot more to the change than just the size. Computers have evolved to a great extent in terms of their capabilities. One such interesting capability is the ability to synthesise speech. Speech synthesis is the artificial production of human speech. A computer used for speech synthesis is called a speech synthesiser. Speech synthesisers can be implemented both in hardware and software[1]. Human Computer Interaction through speech is growing rapidly. A good variety of interesting applications - screen readers, virtual assistants and IVR(Interactive Voice Response) systems have come into existence. To put it simple, the functionality of a computer software to convert a text into speech, synthesise it and speak it is called TTS. In addition to simplifying human computer interactions, TTS has opened up numerous possibilities for the print disabled. The key benefit TTS holds is that it gives the print disabled people a way to access information and engage in the digital world. Spoken alarms, airport or railway announcements are a few examples where one may have listened to synthesised speech. IVR systems for customer care is a very common example of a situation where a person interacts with a text to speech(TTS) system. TTS systems are used for a wide variety of purposes. In this context, the most relevant ones are the use of TTS in assistive technology (screen readers) and the use of TTS in virtual assistants. A screen reader is a software that gives spoken feedback of all system messages and text displayed on the computer. Screen readers have opened up a wide range of learning avenues for the print disabled. These software make use of TTS engines to generate speech for the text they encounter.

1.1 An Overview Of Text To Speech

Text to speech, or Text to speech systems automatically convert text to synthesised speech. Let us say we have a text document. If the document is given as an input to a text to speech system, the output will be a spoken version of the input document. Text to speech systems have a voice and the grammar for a language. They generate speech based on the input text, the voice and the language. There are 2 types of text to speech synthesis.

- Limited domain synthesis.
- Full synthesis.

1.1.1 Limited Domain Speech Synthesis

Limited domain speech synthesis [2] is a speech synthesis technique used in situations where the text to be spoken is of a limited domain. That is, the text to be spoken is fixed (fixed set of words). Unit selection [10] is one of the popular and effective techniques for limited domain synthesis. Unit selection is used on a database of words related to a domain for synthesis. One may have listened to limited domain synthesised speech in IVR systems, talking clocks and ATMs(Automatic Teller Machines), automated announcements, etc. Limited domain speech synthesis is said to produce a speech output of a high quality. This is due to the limited data that has to be processed for the synthesis. We will not be making use of TTS engines that synthesise speech using this technique in the context of this research. This is primarily because all of the terms used in mathematics can be synthesised by any TTS system that can synthesise english speech. A major setback in using limited domain synthesis is that the output may vary when our ideas are integrated into a mainstream screen reader.

1.1.2 Full Synthesis

Full synthesis is a more generic form of synthesising speech from text. Voices generated intended for this purpose use a significantly large amount of text recordings to build the voice. These voices are capable of synthesising any text of that language. That is, they can speak out any text of that language unlike the voices generated for the purpose of limited domain synthesis. Voices used in screen readers are generated using this approach. But TTS systems and screen readers have a good number of limitations. For instance, they are incapable of speaking graphical content and animated content. In the following chapters, we will be discussing about

a major limitation of TTS systems the incapability to effectively speaking mathematical content. Current day TTS systems and screen reader implementations are not capable of efficiently speaking mathematical content.

Mathematical equations comprise of different types of visual cues to convey their semantic meaning. Some of these visual cues are superscripts, subscripts, parentheses,etc. Despite advances in screen reading and text to speech technologies, the problem of speaking complex math remains majorly unresolved. Speaking the equation just as any other string of text, a line, or a sentence will not suffice to effectively render mathematics in speech. Consider the expression:

$$e^{x+1} - 1 \tag{1.1}$$

This expression denotes that the value "e" should be multiplied "x+1" times before subtracting 1 from it. However, when it is rendered in speech like a general string, it is difficult to identify the portion of the equation in the superscript and the remainder of it after the superscript. That is, the listener may assume the equation to be e^{x+1-1} or $e^x + 1 - 1$. To effectively resolve such ambiguities and identify such demarcations in mathematical content, information presented through visual cues such as spatialisation must be mapped to their auditory equivalent. Mathematics, in its visual form, gives the reader a very high level granularity in perceiving the equation. Mathematical equations, when presented in audio must be able to match the advantage in granularity provided in visual representation of mathematics. The typical issues in audio rendering of mathematical equations include quantification, superscripting and subscripting, and fractions.

1.2 Key Aspects Of Audio Rendering Of Equations

1.2.1 Quantification

Most of mathematical equations contain expressions in parentheses. For instance, consider the equation

$$(A + B) * (C + D) + E \tag{1.2}$$

It may seem that the equation can just be treated as a general string of text while speaking. However, this will create a confusion to the listener as there are two ways of expressing equation 1.2 The equation could be spoken as

- "left parenthesis A plus B right parenthesis times left parenthesis C plus D right parenthesis plus E "

- "A plus B times C plus D plus E".

In the former case, the listener will have to keep a track of all the parentheses when he or she listens to the equation. This becomes a hectic task for bigger equations and also results in deviating the listener's attention from concentrating on the actual contents of the equation. On the other hand, in the latter case, the listener gets an ambiguous representation of the equation. The listener could interpret the equation as

- $A + B * (C + D + E)$
- $A + B * C + D + E$

The spoken form of the equation should have additional information to the contents of the equation to solve this ambiguity.

1.2.2 Superscript And Subscript

Today's screen readers and TTS engines do not effectively convey the equations with superscript and subscript content. They often do not speak out the parts of the equation contained in the superscript and subscript. They speak out such content continuously, with the rest of the equation. For instance, let us say the expression is E^X. With the currently available technologies, the expression may be rendered as "EX". This does not give the listener the information that X is in the superscript and the listener may understand the expression as $E * X$. In expressions where there are at least 2 variables that cause a phonetic sound when spoken together, the general TTS in screen readers may treat the expression as a complete word. Consider the expression A^B. The TTS may speak it as "ab". In case of numbers, say we have an expression 5^{2^5}, the TTS in screen readers reads it as "five hundred twenty five" or "five two five". We come across the same issues while trying to render subscript text. If a human speaks the expression, he may not make such mistakes. The challenge to the human speaker lies in effectively conveying the spatial orientation of the different parts of the equation. That is, the equation, presented in audio must give the listener a clear picture of what content is in the superscript and the subscript. The listener must also be able to observe the end of the superscript or subscript part of a mathematical expression. The listener should understand that any thing that he listens to after the end is in the baseline or the general part of the equation, unless specified. An example illustrating the conditions mentioned above(superscripting and subscripting, and the expression after the superscript or subscript) is given below.

$$(a + b_1)^2 + c = a^2 + 2 * a * b_1 + b_1^2 + c \tag{1.3}$$

To effectively render equation 1.3, The following factors must be taken care of.

- The listener must understand that the superscript 2 applies to the entire expression in parentheses and not just b_1.
- The superscript 2 applies only to the part of the expression enclosed in parentheses and does not apply to the $+c$ term in the equation.
- Care must be taken while rendering the b_1 term in audio. specially when rendering b_1^2.

To overcome the challenges explained in example equation 1.3, the spoken form of an equation should provide the listener with different cues for superscript and subscript content.

1.2.3 Fractions

Fractions, like the other mathematical concepts discussed above can not be treated like a general string of text. The key information that has to be conveyed to the listener in addition to the contents of the fraction is the beginning of the fraction, the content of the fraction in numerator and denominator and the end of the fraction. A good example to show these factors explained is given below.

$$a + \frac{b+1}{c+1} - 1 \tag{1.4}$$

The audio equivalent of equation 1.4 should be able to convey the following information.

- The beginning of the fraction after $a+$.
- The content in the numerator, $b + 1$.
- The content in the denominator, $c + 1$.
- The difference between the content in the numerator and the denominator.
- The end of the fraction.
- the part of the expression after the fraction, -1

The audio equivalent of the equation should effectively be able to convey nested fractions in addition to the regular fractions to the listener.
Effectively rendering mathematics is a major issue that is hindering people dependent on assistive technologies to efficiently understand and comprehend mathematical content. There have been attempts to overcome this hurdle of rendering mathematical content in alternative forms to regular print. A few of these attempts will be outlined in section 2.6.

Section 1.3 will give an outline of the thesis.

1.3 Organisation Of The Thesis

Chapter 2 outlines the current state of accessibility of different types of content. In chapter 3, we will discuss the significance of the task of rendering mathematical content in audio. Section 3.1 gives a brief account of the current methods followed by visually challenged individuals to access mathematical content. Section 3.1.1 gives an account of my experience in learning and in general, dealing with mathematical content in the course of my education. Section 3.2 sets the tone for the challenge at hand and shows the significance of rendering mathematical content in audio from an experimental standpoint. Section 3.4 explains the working and design of the prototype built for the purpose of this research. Sections 3.5 to section 3.8 explain the developed techniques. Section 3.9 gives the results of evaluation of the proposed techniques and gives an analysis on the test results. Section 3.10 concludes the discussion on the proposed ideas and sets the tone for further analysis of the developed concepts to render mathematical content in audio.

Chapter 4 provides details on the comprehension test conducted to evaluate one of our proposed technique to render equations and explains our approach to render pie charts in audio. We talk about the comprehension test performed on participants with and without vision impairment in section 4.1. In section 4.1.1, we talk about the participants for the comprehension test. Section 4.1.2 contains the details of the comprehension test. Section 4.1.5 shows the results of the comprehension test and gives details about the participant feedback for the experiment. More attention was given in taking feedback from people with vision impairment. This is due to the fact that a lot of factors (use of assistive technology, comfort level with using the computer and test environment, etc) play a critical role in the case of participants using assistive technology software. Section 4.4 explains the evaluation method used in the audio rendering of statistical data.

In chapter 5, I conclude the thesis. Section 5.1 summarises the idea documented in the thesis. Section 5.3 gives an overview of the various platforms where this research has been presented and published. Section 5.4 discusses the plan I have for this research and most likely my future attempts at research.

Chapter 2

Review of Types of Accessible Content

2.1 Overview

This chapter draws a big picture on the current scenario of online and digital information accessibility. It encompasses a non exhaustive list of the different kinds of information that is both accessible and inaccessible. I attempted to outline the reason behind the inaccessibility of certain types of information. This reasoning is based both on my personal experiences and existing literature.

People with print disabilities often perceive information that is trivially conveyed visually through alternative sensory inputs. Audio and Touch(haptic input) are very effective alternative inputs in the context of conveying visual information. In some situations, a combination of these inputs aid in providing a more detailed experience or idea of the visual information. In the past, conveying this information through these alternative sensory input channels was a task that required a lot of effort in terms of technical work and resources. With the development of technology, this task has been simplified to a great extent. However, there are a significant challenges in making information accessible to the print disabled.

2.2 Text Accessibility

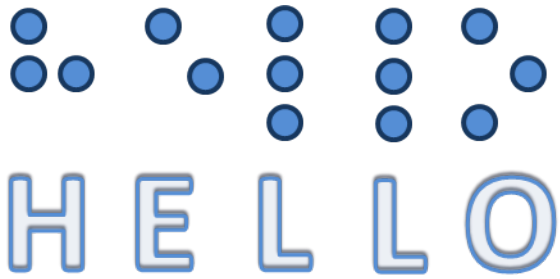

Figure 2.1 The english word “hello” in braille.

Text is one of the most prevalent and common forms of visual information. There are a number of ways in which text can be conveyed through a non visual input channel. Braille [17] is widely used to convey text to visually impaired users. It is a language in which each character is represented by raised dots. Different set of raised dots represent different characters (Numbers, letters, symbols, punctuations, etc). The figure 2.1 shows the representation of the word ”hello” in Braille. Each letter is in a cell and each cell can have 6 raised dots. The character can be identified by feeling the combination of raised dots in each cell. Braille is one of the de facto standard for text information access among the visually impaired. Production of Braille information requires additional training and resources. Perkins brailler is a good example of a hardware brailler. The user will have 6 keys to enter text in addition to a space key and a key to go to the next line. At a later stage, a variety of braille devices have been invented. Most of these devices interfaced between the computer and the braille. Braille printers and braille keyboards for the computers are good examples. One of the easier ways to access digital information in braille is through a refreshable braille display. A refreshable braille display [21] is a device that contains a set of cells which contain 6 movable pins. The pins corresponding to a character are pushed up by motors present in the device. The refreshable Braille display takes input from a computer, smartphone, tablet, etc. The refreshable Braille display then moves the pins corresponding to the dots in the letter in each cell. Development in text to speech technology has opened up a possibility for a new class of assistive technology software, screen readers [12]. A screen reader is a piece of software that interprets the information displayed through speech and sound icons. A screen reader does this with the help of a text to speech(TTS) soft-

ware. In addition to reading the content that is displayed, screen readers also have capabilities to simplify navigating through the information. They enable a user to do this by making use of keyboard shortcuts through organisation patterns [like headings and links] used for presenting the information. The text to speech engines are also capable of interpreting punctuations in text as appropriate variations in speech patterns. These patterns involve the variation in intonation. Despite advancements in screen reading and text to speech technology, a significant amount of text is inaccessible to the users. Reading printed text is still a major challenge for visually impaired users. Many OCR software and hardware products like Kurzweil reading machine [13], KNFB reader and openBook by freedom scientific enable access to printed material to some extent. These softwares provide a significant level of ability but are neither completely accurate nor easy to use. That is, they require additional equipment, technical and cited capabilities. Another kind of text that could be inaccessible to screen readers is found in software or apps that have heavily graphical user interfaces. In these apps, developers do not use the standard user interface elements provided by the operating systems and they do not pay attention to accessibility in the custom UI elements they develop. In some cases, they use an image containing text for aesthetic purposes. To the best of my knowledge, this is done to display static text. JAWS 15 and higher have the functionality to perform OCR on an application window. However, This is not a full proof solution. PDF documents containing text in the form of pictures are also inaccessible to a screen reader. They can be spoken only after OCR is performed on them. Again, the output is based on the OCR accuracy in recognising text from the PDF.

2.3 Web Content Accessibility

A significant amount of web content is accessible to screen reader users. HTML, the language used to design web pages is quite structured and enables assistive technology softwares to present information in a suitable manner. A few areas that are inaccessible include:

- The use of custom CSS style sheets for presenting information, links, buttons, etc.
- The use of images containing text instead of the text.
- Heavy use of flash content.
- Absence of labels for buttons, etc.
- Absence of alt text for images. Screen readers speak the alt text in place of the image. This can be used to describe the image.

With HTML 5 now ready for mainstream adoption, a significant improvement can be expected in web accessibility. In the recent times, web applications are becoming very dynamic and interaction rich. This could prove to be a hinderence for assistive technology software. The use of web standards like Aria can enable these highly interactive web applications to be accessible. The world wide web consortium has proposed a set of recommendations known as Web Content Accessibility Guidelines(WCAG). As of now, the version that is in adoption is 2.0. [4]

2.4 Mathematical Content Accessibility

Mathematical content is one of the most inaccessible forms of content to screen readers. Mathematical equations have a lot of symbols and the information is visually organised. Website developers face difficulties to display mathematical content on the web. Web developers used GIFs to display math on the web. In some cases, Websites had photographs of hand written equations. This made the mathematical content inaccessible to assistive technology softwares. Screen readers could not extract the content out of such images. Images and GIFs also made it difficult for screen magnification software to magnify the mathematical expression appropriately. In addition to accessibility implications, use of GIFs and photographed equations made it difficult to save the content of these webpages. The use of these formats to have mathematical content on the web also makes it inaccessible to search engines. The advent of languages like LaTeX[14], MathML, SVG[19], etc enable developers to make mathematical content more presentable on the web. Content written in these languages is more structured and there is a possibility to make it more accessible. Mathematical content presented in these formats will make it more accessible to search engines. Section 2.6 contains more details.

My research efforts are confined to making mathematics more accessible through audio. Making mathematics in different forms (Images, GIFs, etc) accessible to a screen reader is not in the purview of this thesis. MathML, a W3C standard for displaying mathematical content on the web is the input format considered.

2.5 Attempts To Render Mathematical Content In Alternate Forms

There have been several attempts to present mathematical content through alternative modes to vision. 2.5.1 gives an account of a few existing devices that are used to accomplish this task.

2.5.1 Existing Devices To Convey Math To The Visually Challenged

The abacus is a device to perform calculations with numbers. It consists of sets of beads that can be moved across a bamboo stick. This tool can be useful to solve basic math containing numbers. It cannot be used to convey a mathematical equation to a person.

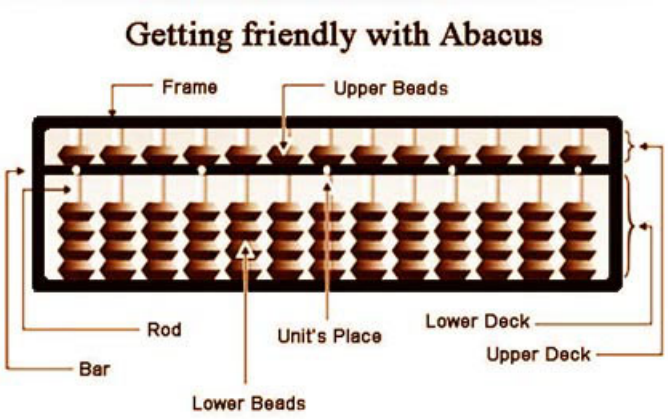

Figure 2.2 Illustration of an Abacus.

The taylor frame is a device to present mathematical content in a form readable by visually challenged people. Numbers and symbols are presented using pegs placed on an alluminium board. The reader gets a different tactile feedback for each number and symbol. A major problem with this device is that the trainer must also be trained to use the device to effectively convey mathematics to a visually challenged student. Using this device also does not facilitate storing equations for later reference. The user would have to clear the board in order to use it for different equations.

These are 2 heavily prevalent devices among educators and students. However, none of these devices are capable of storing mathematical content for later reference. That is, a visually challenged individual should depend on a person trained in using these devices to get access to mathematical content. With the increase in digital content and availability in assistive technology, there are possibilities to more effectively render mathematical content in audio. A few attempts will be discussed in 2.6.

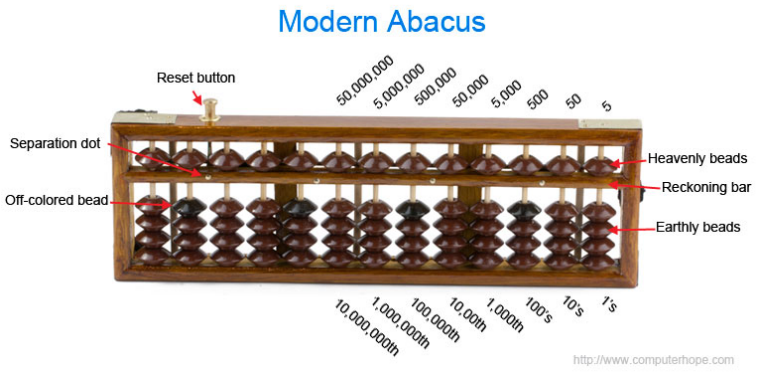

Figure 2.3 Calculation using Abacus.

2.6 Attempts To Render Math In Audio And Braille

Efforts have been made to formulate standards for presenting math through Braille and speech. Nemeth Code[16] is a special type of Braille used for math and science notations. With Nemeth Code, one can render all mathematical and technical documents into six-dot Braille. This code could also be used to speak mathematical content. Dr Nemeth's idea of speaking mathematics using the Nemeth code is illustrated in a paper on mathspeak [15]. MathSpeak is Dr. Nemeth's method of reading mathematics aloud so that it can be written down while the reader is speaking in either Nemeth Code or print. In fact, it conforms exactly to the structure of the Nemeth Code. As the key area of consideration is rendering mathematical equations in audio, an example of speaking an equation with the help of Nemeth Code is given below.

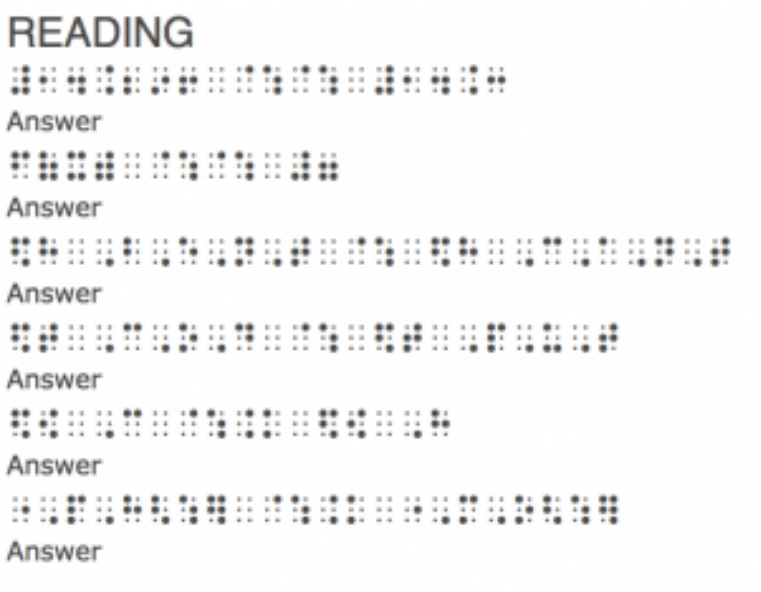

Figure 2.4 Nemeth code example.

An article [24] describes an attempt to simplify speaking Nemeth code (1972 revision). The article gives an understanding of the Nemeth Code from a design standpoint. Dr T.V Raman has developed an audio system for technical readings (ASTER)[20]. ASTER is a computing system for producing audio renderings of electronic documents. The present implementation works with documents written in the TEX family of markup languages: TEX, LaTeX and AMS-TEX. A more recent attempt has been made by a company called Design Science. They developed an internet explorer plugin called MathPlayer [23] that displays and speaks out mathematical content marked up in MathML [11]. This plugin works with existing assistive technology software like JAWS(Job Access With Speech) [22]. Mathematical expressions often have text with different font sizes. MathPlayer enables a user to dynamically change the size of the expression. The plugin also changes the font size of the text of the expression if the font size of the browser window is changed. According to this article, the plugin does not make use of the features provided by a speech synthesiser to introduce pauses and other prosodic elements. It makes use of commas and periods to introduce pauses. There have been attempts to form a set of guidelines to effectively speak mathematics in audio. The handbook for spoken mathematics [5] gives an account of such an attempt. This handbook proposes certain conventions to speak mathematical equations. The author's intension is to make the handbook useful for people interested in synthesising equations in speech. An article on how to speak math also describes the challenges in speaking mathematics to and by a computer [6]. The author talks about using

speech in combination with other forms of input(hand writing, etc). The author address 2 key aspects; spoken mathematics to and by a computer. The latter aspect is of relevance in the context of this thesis. In this article, the author proposes that spoken math can be used in 3 different ways.

- Speech can be used as a primary method for conveying mathematics.
- It can serve as an auxiliary method of conveying mathematics in a multimodal input or output environment.
- It could be used as an error correction command language.

Earlier works discussed so far, have not effectively used paralinguistic cues and variations in the equation. However, humans use a lot of cues when reading out a mathematical equation which helps in understanding the semantics of it. Usage of the cues similar to the humans would result in more effective rendering of the equations. Most of the research efforts outlined so far did not focus much on measuring the effectiveness of using different methods of rendering mathematics in speech and audio. [7] Describes an attempt to improve accessibility of mathematical content and emphasises to a good extent on evaluating the effectiveness of the improvements on the target audience (people with vision impairment).

The objective of this work is to analyse the way these visual cues are presented in an auditory format by human speakers who are well acquainted with speaking the mathematical content, especially to visually challenged individuals. A subjective and objective analysis is performed on the equations recorded by the speakers. Based on this analysis, we make an attempt to form specific rules. The rules map the visual cues to their auditory equivalents, which facilitates to programatically and unambiguously render the mathematical content in audio using a text-to-speech system.

Chapter 3

Significance of Paralinguistic Cues in the Synthesis of Mathematical Equations

In this chapter, I will discuss the importance of rendering mathematical equations in an auditory format. I will discuss the reasons that make rendering mathematical content in audio a challenging task. As mentioned in chapter 1, currently available screen readers and TTS systems can not unambiguously convey mathematical equations. With the advent of speech driven and voice based user interfaces, the problem of conveying math will persist due to the limitation in existing TTS technology.

The improvements to audio rendering of mathematics will not only benefit the general public but also to those with print disabilities. The people with print disabilities who could benefit include those suffering from vision impairment, blindness, dyslexia, cognitive disorders, etc. Scientific content has not been accessible to this segment of people at large. This is mainly due to the fact that scientific content has significant amount of mathematical equations. This has proven to be a major hindrance for people with print disability to pursue education and careers in scientific disciplines like engineering, mathematics and chemistry. In addition to limitations in learning, people attempting to pursue scientific disciplines of education have to deal with giving tests with scientific content. However, in my attempt, only a part of the latter problem of examination is handled. With methods to effectively render mathematical equations in audio in place, the student with print disability will be able to understand a question at the bare minimum. Methods to effectively take responses containing heavy mathematical equations is out of the purview of this thesis.

In section 3.1, I give a brief account of my experience and challenges in understanding mathematical content. I will also give a brief overview of the laws in place for examinations for people with print disabilities, primarily visually challenged test takers.

3.1 Current Methods Followed By Visually Challenged Individuals

Despite the many difficulties present in accessing mathematical content, people with print disabilities have made attempts to understand mathematics with the help of existing technology in addition to making use of a few methods they have developed. A good number of individuals dealing with mathematical content make use of existing devices like the ones mentioned in 2.5.1. However, there has always been a need for human intervention. That is, there was always the need for a human to convey mathematics orally. Section 2.6, talks about efforts to convey mathematics through speech. However, from my experience and my interaction with people in this regard, these developments are not in use. One significant reason for this observation could be that most of these efforts are still in the form of research projects. On observing and using different assistive technology, I understand that different people with the same form of disability have different ways of using the same assistive technology. My opinion is that, these products and ideas neither provide a fool proof solution to the challenge of rendering mathematical content in audio nor do they cater to the different needs of different users.

3.1.1 My Experience Learning Mathematical Content

At the time of my basic mathematics education, assistive technology was quite in its infancy and a computer speaking mathematical content was far from reality. I made use of the abacus and the tailerframe for basic mathematical equations(arithmetic, algebra and trigonometry). In addition to these, I used a tactile geometric kit for basic geometry. A lot more than what the assistive devices offered was instrumental in helping me understand scientific content. For instance, clay models and general toys similar to Lego were a good tool to get a perception of 3-dimenssional geometry. Despite making use of all these assistive aids, I think I have benefited the most with the help of a sited person speaking math to me. Speech feedback has served as a supplement to the information provided by these assistive aids in some cases. While having mathematical content spoken to me and making use of screen readers to gain access to non mathematical content, I realised the value added by human speech while speaking mathematical content. This “value add” will be explained and it will be experimentally validated in the next section.

3.2 Cues In Spoken Equations

Our study is based on the preposition that treating a mathematical expression as a regular English sentence while speaking is not an effective way to present mathematical content in an auditory form. In order to test this observation, we asked a set of 15 people to rate mathematical equations spoken by a traditional TTS system. Then we conducted the same experiment on spoken equations (i.e., equations spoken by a human-being). The details of the listening tests are as follows.

A set of 15 participants were made to listen to the recorded equations. Each participant was made to listen to the equations using headphones and the responses were recorded. The listening test was self paced and also the users were informed that they were free to listen to the equation any number of times till they felt comfortable that they could recall the equation. Similarly, the same participants were also made to listen to speech of mathematical equations generated by a TTS system. The participant will have to reproduce the equation he/she listens to. In addition to reproducing the equation, the participant will have to evaluate the spoken equation based on eight other parameters, i.e., perform objective analysis. We arrived at these parameters partly by following the listening test procedures followed in the Blizzard challenges [9] and our own analysis.

3.2.1 Selection Of The Equations

Selection of suitable equations is a critical component to analyse the auditory presentation of mathematical content. We hand picked a few equations which had variations in number of variables, number of sub expressions and length of the equation. The equations can be found in appendix A.1. Each of the equations is semantically unrelated, that is, the equations have mathematical content but the listener may not have come across the exact same equation prior to listening to them from our recordings . The reason behind choosing the equations in such a way is to ensure that the listener's prior knowledge does not influence the ability to recall the equation. If the listener is able to recall the equation even before he or she listens to it completely, the listener is benefitting from memory, not the spoken equation.

Parameter	Spoken	Synthesized(Technique 1)
Listening Effort	2.5	4.4
Content Familiarity	2.7	2.7
Effectiveness of additional cues	3.2	1.2
Accentuation	4.3	2.5
Intonation	4.26	1.6
Pauses	3.1	2.15
Number of repetitions (Mode)	2	4
Mean Opinion Score	4.42	1.89

Table 3.1 Evaluation of Spoken Math vs TTS

3.2.2 Selection Of Speakers to Record the Equations

We have chosen 2 sets of speaker participants to record the equations. The first set comprises of people who have certification to teach mathematical content to visually challenged students. The other set comprises of those who have trained themselves maybe with a little bit of assistance to teach math to visually challenged students. These speakers had to train to teach mathematical content to visually challenged students due to the need to help friends or family who needed to be taught math in this way. The speakers who have trained themselves have been able to teach mathematics and record equations as effective as those who have been formally trained and certified. Table 3.2 gives relevant details about the speakers who volunteered to record equations. In total, we had 4 speakers record the equations. However, we could only make use of equations recorded by 3 of the 4 speakers. One of the speaker had heavy native language accent.

Name	Age	Profession	Certification	No of students	Experience (years)	Level
Speaker1	43	Consultant of Vision Rehabilitation	Yes (Masters Degree in Vision Rehabilitation)	nearly 400	19	up to 10th grade
Speaker2	25	Working in hospitality finance	No (Teaching a family member)	1	7	up to 12th grade
Speaker3	45	Special Educator	Yes (Diploma in Community Based Rehabilitation)	50	8	Primary school
Speaker4	48	Teacher	No (Teaching a family member)	2	20	up to 10th grade

Table 3.2 Details of speakers

3.2.3 Parameters For Objective Analysis

On a scale of 1 to 5, the participants were asked to evaluate the spoken equations on the following parameters.

- Listening effort (1 = low, 5 = high)
- Intonation (1 = ineffective and 5 = very effective)
- Acceptance (1 = poor, 5 = good).
- Speech pauses (1= not noticeable and 5 = very prominent)
- Accentuation (1 = poor and 5 = very prominent).
- Content familiarity (1 = totally new concept and 5 = very familiar). Here 1 indicates that the user is not acquainted to the terminology used in the equation. In this case, the participants' response for that particular equation can not be considered completely as he may have entered a wrong response due to the lack of domain knowledge, not due to the lack of understanding of the audio.
- Effectiveness of additional cues such as sounds, pitch and rate variations, change in direction, etc. (1 = hardly noticeable and 5 = very helpful).
- Number of repetitions of each equation.

3.3 Inferences From The Listening Tests

The results of this experiment, shown in the Table 3.2.1 indicates that the equations are not intelligible enough if it is spoken as a plain text using a text-to-speech system. The mean

opinion scores of spoken equations indicate a human-being use several acoustic cues to manifest the semantics of the mathematical symbols in audio mode. It was noticed that the trained speakers brought certain variations in their speech while speaking specific aspects of the mathematical expression. The variations are noticed in pauses and pitch variations (intonation). A careful analysis revealed that the acoustic variations were introduced by the speakers to unambiguously speak 1) quantification, 2) superscripting and subscripting and 3) fractions in mathematical equations.

Based on the feedback received from participants, we can infer that the use of these additional cues can effectively and unambiguously present mathematical content in audio. The question is how to introduce such cues to synthesise a mathematical equation using a text-to-speech system.

3.4 System Design

With the advent of languages like MathML, it is possible to programatically identify different attributes and visual cues of a mathematical expression. This possibility can in turn be leveraged to make some modifications while generating speech for mathematical content. We propose four techniques that could enhance the way mathematical content is rendered in audio.

The Equation was first converted into the Math Markup Language format. We chose "Presentation" Markup style to represent the equations. It is then text processed to identify and segregate the different terms occurring in the equation. The following terms have been segregated.

- Subscripts and superscripts
- Fractions
- Square root terms
- Overscripts and underscript

The MathML representation is processed to convert it into natural language and the acoustic cues such as pauses, intonation are incorporated to generate a file in the SABLE [1] markup language. The SABLE file is input to the speech synthesis system which generates the audio

[1]SABLE is mark up language due to collaboration between Sun, AT & T, Bell Labs, Edinburgh and CMU to devise a standard cross synthesizer standard mark up language. The language is XML-based and allows users to add addition controlling commands in text to affect the output. An implementation exists in Festival speech synthesis system.

form of the equation with specified pauses and intonation. We have generated the audio files using the Festival Speech Synthesis System[3]. Sections 3.5 through 3.8 discuss each of the four proposed techniques.

3.5 Technique 1 : Rendering equations With Pauses And Special Sounds

In visual communication, icons and symbols are used as indications for some types of information. In the context of mathematical expressions, the user can perceive the type of elements (superscripts, subscripts, etc) by glancing at the equation. A person has the advantage of perceiving a lot of information of the equation even before looking at the actual contents of the equation. This technique attempts to present the equation in a manner that a person gets a similar advantage when he listens to it.

In this concept, we made use of special sounds or ear cons while presenting the equations. However, replacing speech with sounds alone is not the most effective way to tackle the problem of presenting mathematic equations in audio. We made use of paralinguistic cues including, but not limited to sounds.

The cues presented in this method include:

- **Pauses** to convey certain parts of an equation. These pauses are mainly used to separate the parts of mathematical expressions. Consider $(A + B)^2$ and $(A + B^2) + 1$. It would sound more natural and intuitive if the expressions are spoken as "the quantity A + B pause superscript 2" and "the quantity A + B superscript 2 pause + 1."

- **Sounds** to indicate certain symbols and mathematical operations. Sounds are used to indicate superscripts, subscripts, roots, under scripts, over scripts and under script-over script combination.

We chose the sounds(such as the sound "ding") such that would be pleasant to the ear and that are passively noticed by a listener so as not to distract too much, at the same time, are loud enough not to go unnoticed. The sounds show a transition from high to low and low to high when there is a subscript and superscript respectively. Any other type of sounds and their variations could also be applied in this technique.

3.6 Technique 2 : Rendering Equations With Pitch And Rate Variations

Table 3.3 Pitch and rate variations

Term	Pitch variation	Rate variation
Superscript	50	20
Subscript	-50	-20
Fraction	25	-25
Underscript	-60	-25
Overscript	60	25

Screen Reader users are familiar to pitch changes. Generally, a high pitch is used to denote capitals and a low pitch is used to denote tool tip messages. On observing the human recorded equations explained in Section 3.2, we observed that speakers tend to modulate the pitch as they read aloud certain parts of a mathematical expression. It has been observed that certain parts of a mathematical expression are spoken at a faster rate to indicate that it is a sub expression and to isolate it from the rest of the expression.

In this technique, we use pitch and rate changes to denote the presence of certain mathematical attributes. The pitch and rate increase while speaking out the superscript text and decrease while speaking the subscript text. A similar method can be employed to properly render fractions. The numerator is spoken in a higher pitch and the denominator is spoken in a lower pitch. Similarly, quantities in a root are spoken at a faster rate. Table 3.3 shows the pitch and rate variation(in percentage) that are applied to the Mathematical equation. The variation is with respect to the base pitch and rate of the TTS.

3.7 Technique 3: Rendering Equations With Audio Spatialisation

Table 3.4 Sets of HRTF angles for audio spatialisation

Term	Elevation Angle	Azimuth Angle
Superscript	90	30
Subscript	-90	30
Fraction	270	45
Underscript	-90	45
Overscript	90	30

In this technique, we made an attempt to draw a closer analogy to the spatial positioning of various variables and numbers of a mathematical equation. The listener can be given the illusion that the superscript part of the math expression is spoken from above his head and the rest at the usual level using the Head Related Transfer Function (HRTF) [8]. Table 3.4 shows the sets of angles chosen for the different parts of the equation such as superscript, etc. We identify the portions of a mathematical expression that require modification in spatial orientation of sound. Based on the attribute, we apply the HRTF function with the required angles. section 3.7.1 will contain details related to the process of synthesising equations having 3-D audio using HRTF.

3.7.1 Determining The HRTF Parameters

The Head Related Transfer Function(HRTF) is a method developed to render sound in 3-D. As detailed in [8], the HRTF model also takes into account the diffraction and reflections due to the position of a person's head and torso. The HRTF model we used is developed to suit any person in general. It is developed keeping in mind that it is not feasible to compute HRTF models for each individual. We first generated the mathematical equations in audio with variations in pitch, rate and pauses(As explained in section 3.5 and 3.6). These generated equations were marked with regions where HRTF had to be applied. The regions to be marked were identified with the help of the variations in speech parameters introduced in the previous step. We then synthesised an equation containing

3.8 Technique 4 : Rendering Equations With Pitch Variations And Special Tones

In this technique, we render the equations in audio by varying the pitch, adding pauses, emphasising the speech and adding sounds at required parts of a mathematical expression. As explained in section 3.6, we can make pitch and rate manipulation while rendering superscripts, subscripts, fractions, under scripts and over scripts. In addition to the variations in speech, we have also added sounds to indicate the listener before hand that he must expect one of the above mentioned variations (superscripts, subscripts, etc). The sounds used here are the same as the ones mentioned in section 3.5. The pitch and rate variations that are introduced are the same as the percentage values given in table 3.3.

$$(X + Y)^{4-2} \tag{3.1}$$

3.9 Analysis Of The Listening Test

Parameter	Technique#1	Technique#2	Technique#3	Technique#4
Intonation Variation	2.3	**4.7**	4.32	4.68
Pitch Variation	1.4	4.43	**4.82**	4.36
Pauses	**4.15**	3.7	3.7	3.87
listening Effort	3.5	**2.3**	2.64	2.47
Content Familiarity	2.7	2.7	2.7	2.7
Effectiveness of additional cues	1.82	4.32	**4.37**	4.23
Accentuation	**3.47**	2.3	3.2	3.6
Number of repititions(Mode)	3	2	2	2
Mean Opinion Score	2.27	4.37	**4.62**	4.35

Table 3.5 Evaluation of the proposed techniques

- Technique 1: Pauses
- Technique 2: Special Sounds
- Technique 3: Spatial Audio
- Technique 4: Use of sounds, pitch and rate variations.

Table 3.6 Acceptance of the parameters

Parameter	Technique#1	Technique#2	Technique#3	Technique#4
Intonation Variation	3	**12**	10	11
Pitch Variation	4	11	10	12
Pauses	5	7	10	10
Listening Effort	4	10	14	14
Content Familiarity	8	8	5	6
Effectiveness of additional cues	1	12	11	14

A system was built to render mathematical expressions implementing each of the proposed ideas. An experiment procedure similar to the initial listening test is followed. 30 participants were made to participate in the experiment. The table 3.9 contains the normalised scores(1 to 5) calculated over the responses for the equations. The number of repetitions of the equation has the mode value(most occurring value).

The table 3.6 shows the participants' acceptance of a particular characteristic in the spoken equation. In our experiment, we have asked participants to rate equations on different parameters. The ratings have been taken on a scale of 5. To show acceptance, we consider ratings >=3 to be positive. The table shows the number of participants that accepted the effectiveness of the particular characteristic. On analysing the experiment as described in Section 3.3 it is observed that the participants are able to understand the human spoken equations. Moreover, it can be clearly understood that generating spoken forms of mathematical equations without making any enhancements is not capable of rendering math effectively. It can also be inferred that making use of just a few paralinguistic cues, sounds and pauses as explained in section 3.5 will not suffice either. The pitch and rate changes while rendering certain parts of the mathematical expressions have proven to be helpful to the participants in understanding the expression. In the method described in section 3.7, the listener has been able to draw an analogy to the print form of mathematics. It has been observed that the method explained in section 3.5 did not prove to be helpful to the listeners. However, from the table 3.9 and the values corresponding to the technique explained in section 3.8, it is evident that use of cues (pauses and rate variations) in addition to special sounds can be significantly effective in helping a listener.

Here are the box plots corresponding to the listening tests.

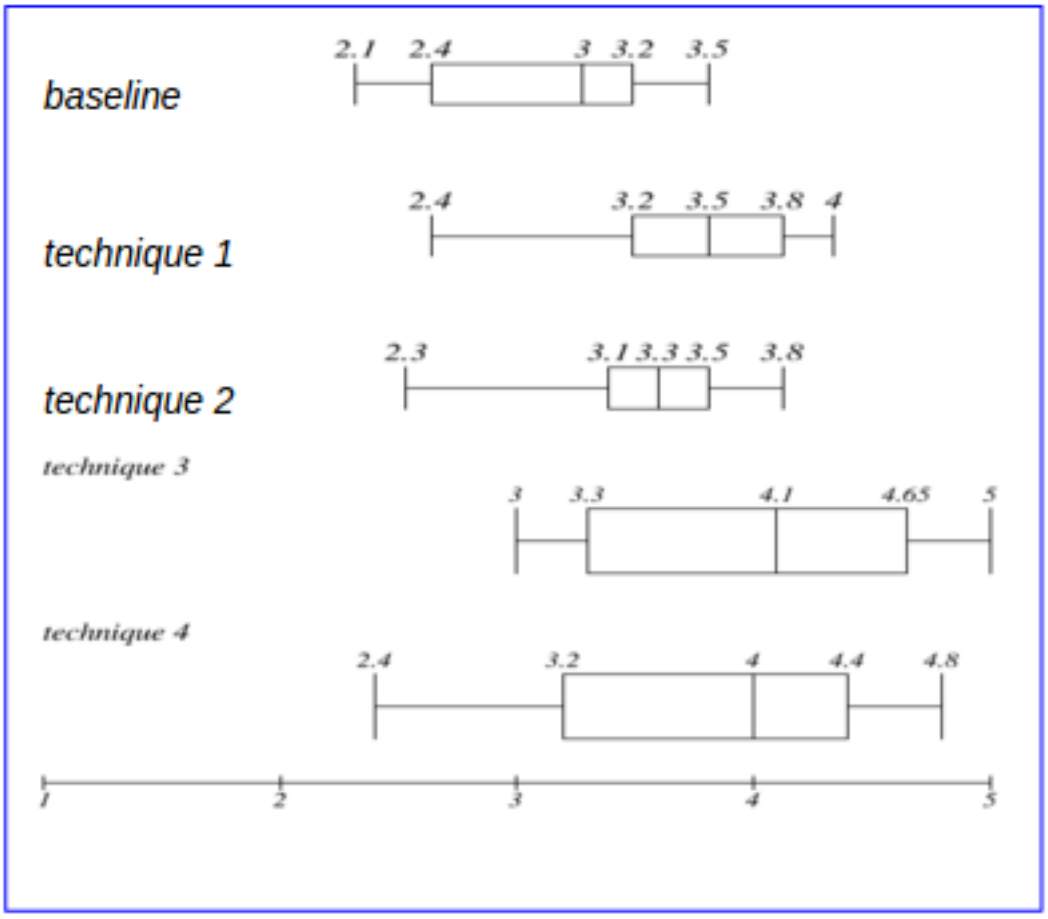

Figure 3.1 Box plots corresponding to the subjective evaluation

3.10 Conclusions From The Listening Test

From the analysis and the proposed ideas, we can say that there is a possibility to unambiguously render mathematics in audio. With the increase in voice driven interfaces and information access through audio, rendering mathematical content in audio could also help more effectively present such content in these interfaces. Personal assistance or any other voice driven UIs can more effectively render mathematical content to the listener. In addition to this, effectively rendering mathematical content in audio can be of a great advantage for people with print disabilities including, but not limited to vision impairment, dyslexia and cognitive impairment. With currently available assistive technology, understanding mathematical content is very dif-

ficult and almost impossible. The ideas explained in sections 3.5 to 3.7 improve the scenario of understanding mathematical content through a non visual input mode. as explained in section 3.8, there is also a chance that a combination of the proposed ideas are more effective than each of the ideas alone.

To validate the claims, we performed a comprehension test on the technique explained in 3.7 on participants with and without vision impairment.

Chapter 4

Comprehensive Evaluation and Audio Rendering of Statistical Data

4.1 Comprehension Test For Equations

In our previous experiments, we showed that the traditional TTS technology can not convey mathematics unambiguously and evaluated our proposed ideas based on a fixed set of parameters. From this evaluation, we could observe that people were able to get a clearer understanding of the equations with spatially oriented audio(section 3.7). To objectively validate this observation, we performed a comprehension test. We perform this experiment on participants with and without vision impairment. The idea behind performing this experiment on people with vision impairment is that they are more acquainted with receiving information through audio. They often make use of assistive technology software for information access (screen readers).

4.1.1 Participants For The Experiment

We chose participants with and without print disability (Vision impairment). All participants belonged to the 18 to 24 year age range and were mostly comprised of students. All participants had sufficient educational background to understand the mathematical concepts involved in the equations. All the participants with vision impairment were capable of using the computer and a web browser with the help of assistive technology. The test computers had JAWS For Windows(V12 and 13) installed and Firefox and chrome were used to participate in the experiment. This is critical to the experiment as the participants had to be on equal grounds in terms of taking the test. That is, participants with vision impairment must be able

by all means to access the experiment as efficiently as the sited participants. Moreover, A quiet testing environment was chosen for the participants to be able to concentrate well.

4.1.2 The Experiment

To test for the ability to comprehend equations rendered in the selected technique, we setup a web-based experiment. The test page has a simple layout. It has:

- the equation number.
- playback controls.
- A text field to enter the user response.
- A submit button to submit the answer and proceed to the next equation.

On visiting the web page, the participant will be given a brief overview of the efforts to render mathematical equations in audio. He will have options to start the comprehension test or learn more about the project. On clicking the "Start comprehension test" link, The participant will be taken to a page containing the instructions for the test. On clicking the "continue" button, the participant will be asked to enter his name, age and email address. Once he clicks the submit button, the participant will be taken to the first expression.

On completion of the experiment, we engaged the participants in a dialog about their experiences during the experiment and their thoughts on the existence of these ideas in a real assistive technology software.

4.1.3 Instructions To Be Followed For The Comprehension Test

Here are the instructions (from the test portal) for the comprehension test.

Welcome to the comprehension test for audio rendering of mathematical equations. In this test, you would be required to do some basic mental mathematics. You are allowed to use a calculator to solve this test; however, this may not be necessary. You will be made to listen to 20 questions involving basic mathematics (addition, subtraction, multiplication, exponents, etc). For each question, you are required to fill a blank with the answer. Ex: (2+3)/(4+5) or (3 + (2/4)). If you are unable to understand the question or you do not know the answer, Please enter 000 as your response. You are required to use headphones for this test. The estimated time for this test is about 20 minutes.

4.1.4 Selection Of Equations For Comprehension Test

This experiment contains two sets of equations.

- Equations rendered with no modifications in speech. This is to replicate the general speech output from a traditional text-to-speach system.
- Equations rendered with appropriate modifications in pitch and rate. These equations are also manipulated to produce a 3-dimensional audio.

The equations contain numbers and basic mathematical operations. The equations are designed such that the participants will be able to solve them mentally. We have chosen equations with simple numbers so as to not deviate from the goal of evaluating the system. Finding the answers to the equations will result in simple numbers.Choosing complex equations may result in the participants stressing more towards solving the problem and this may hinder them from completely understanding the equation first. This holds true especially for people with vision impairment. There are no effective means to write or store the intermediate steps in solving the equation. The participant will be made to listen to equations of both types in random order. For each equation, the participant will have to enter his or her response. The equations used for this test are given in appendix A.3.

4.1.5 Results Of The Comprehension Test

From the table 4.1.5, we can observe that participants could solve the equations better when the equations are rendered using the proposed technique. On comparing the results for participants with and without vision impairment, the participants with vision impairment have almost the same correctness rate when compared to those with no vision problems. This observation holds true for equations rendered using our technique(3.7). It must be noted that participants who used assistive technology took around 30 to 35 minutes to complete the test. Use of assistive technology may result in people accessing the computer at a slower rate. Moreover, the assistive technology (JAWS) and the questions in our test are in the form of speech and audio. So, visually challenged participants took time to get acquainted with the experiment interface. the screen reader's speech had to be interrupted when the question was spoken. On talking to the participants, we understood that they were excited about the fact that there is a possibility for a computer speaking out mathematics similar to a person (there usual cited reader, family member, etc). The participants expressed concern over the quality of the voice used in the synthesis of equations. However, they clearly stated that the voice used was not unintelligible. Moreover, the participants were allowed to repeat the equations and the test was self paced.

The dialog set a tone that they expect a better-sounding voice for long time use. The participants say that the additional verbosity, the pauses and the variations in pitch and rate made understanding the equation an easier task. The participants felt that similar methods to render mathematics in audio could prove beneficial to them in terms of information access, primarily educational content.

Table 4.1 Accuracy Percentages in Comprehension Evaluation of proposed technique vs TTS

Person	Proposed Technique	Standard TTS (Traditional TTS)
Normal	96	73
Visually Challenged	95.7	29.7

4.2 Audio Rendering Of Statistical Data

In the previous chapters, we have emphasized on rendering mathematical equations in audio. We have proposed techniques to render equations in audio in 3.5 to 3.8. We could experimentally validate that the technique (3.8) produces understandable equations in table 4.1.5. Mathematical content however is not confined to equations. A lot of numeric information is presented in statistical data. Most of this information is in the form of bar graphs and pie charts. This chapter demonstrates the applicability of our approach to statistical data (pie charts).

Bar graphs and pie charts are a graphical means of displaying statistical data. Though they look like figures, the underlying information is highly numerical and statistical. Let us look at an example where these are used. The pie chart in the figure 4.1 pictorially shows the percentage values of different vegetables and fruits sold by a vender. Currently available text to speech systems are not capable of speaking this form of data. If the numbers and different categories are spoken, the listener would have to memorise the data to later use it. For example, a listener would have to remember all the data in a pie chart to answer a question related to it in an examination. If this data is rendered with the use of some cues, it would be easier for the listener to remember it and easily recall the relevant information.

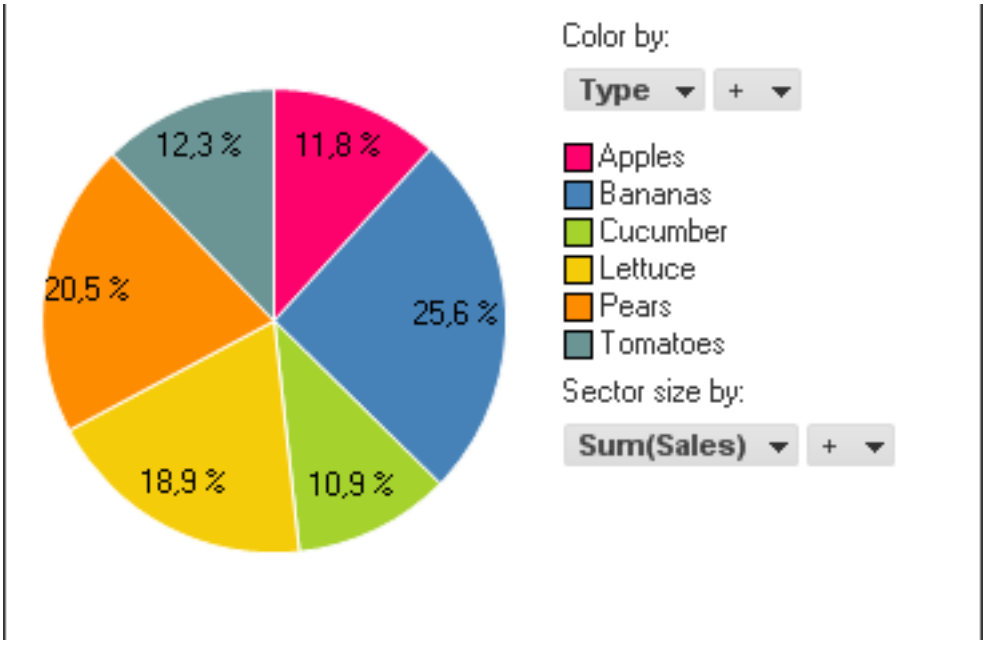

Figure 4.1 An example pie chart showing the amount of different vegetables and fruits sold by a vender.

4.3 Techniques To Render Pie Charts

Similar to the methods followed to render equations, we use a format called SVG(Scalable Vector Graphics) as an input format for our system. We then extract the relevant statistical information that is required to render the given chart or graph in audio. The system identifies areas where relevant cues are to be inserted. The system then generates a sable file with the data and the cues. This sable file is given as the input for a text to speech engine, festival in our case and is spoken. figure 4.2 depicts this workflow.

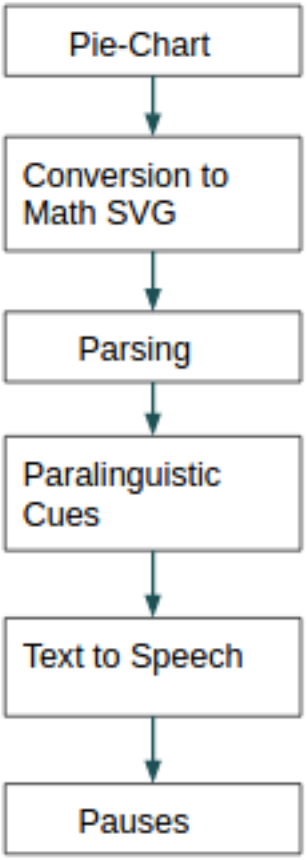

Figure 4.2 A figure representing the workflow to render pie charts in audio.

4.3.1 Rendering Charts With No Cues

In this technique, we get the required numerical data, the data in the figure. This data is spoken in the order it is depicted in the figure. In this technique, No cues or variation in speech patterns are introduced. It is considered the trivial way of rendering these figures. For the purpose of our study, We will be referring to this technique as the baseline technique. There is no implementation that would render charts and graphs that could be used as a base case for our study.

4.3.2 Rendering Statistical Data With Pitch Variation

In this technique, we make use of variations in pitch to render statistical data. The relevant data is extracted from the SVG Markup. Areas that require modification in pitch are identified. The Pitch must be increased accordingly. For instance, if the category K holds P percentage in the pie chart, the pitch must be increased by P

4.4 Evaluation Of The Proposed Techniques

We perform a comprehensive evaluation on the proposed techniques. Each participant was presented with 2 to 3 questions from each technique. Each question will contain a graph or a chart followed by a question. The experiment had 2 types of questions.

4.4.1 Types of Questions

There are 2 types of questions.

- Direct answers. In this type of questions, the participant is made to listen to a pie chart and is asked how much percentage a particular category or a person has or what is the remaining percentage for a particular category for which the percentage value is not spoken.

- comparison questions. In this type, the participant could be asked questions like: Who has the maximum, minimum, second highest or second lowest percentage? is percentage of A $<$ or $>$ percentage of B?

4.5 Results

We presented participants with 10 questions. five questions were rendered with no cues and the other 5 were rendered with pitch variations. The correctness percentage and the number of repetitions were recorded.

Table 4.2 Evaluation of audio rendered pie charts.

Measure	Proposed Technique	Standard TTS (Traditional TTS)
Accuracy	94	64
Number of Repetitions	2	5

Chapter 5

Conclusions

5.1 Summary

Let us classify data based on the structure and organisation. We have text, which is very structured. That is, we are able to process the information about a particular text and get various outcomes. Natural Language Processing is an entire field that is a result of this possibility. The information about a particular text is used to make text to speech sound more natural. The next category of data is mathematical equations. Equations are less structured when compared to text. Currently available text to speech systems are not capable of effectively rendering mathematical content in speech. Our attempts to do this have been detailed in chapter 3. The third category of information we attempted to render in 4.2 is graphical data(bar and pie charts).

In this research effort, we embark on the task of rendering mathematical content in audio. We feel that technological advancements can aid in making assistive technology more informative and improve the accessibility of different types of the content to everyone. We embark on the task of making mathematical content accessible to the print disabled. We made individuals trained to teach mathematics to visually challenged students to record a set of equations. We then generated the same set of equations without adding any additional cues using synthesised speech. These sets of equations were evaluated by individuals and it was evident that equations recorded by human speakers were better understood by participants. We analysed these equations and proposed 4 techniques. Each of the techniques make use of different linguistic and paralinguistic cues. These techniques have been explained in sections 3.5, 3.6, 3.7 and 3.8. We then evaluated these techniques. The evaluation procedure and parameters were similar to the ones used to evaluate human spoken equations. The results of this evaluation are given in table 3.9. We pick the 2 best performing techniques (explained in sections 3.6 and 3.7). We render

statistical data (pie charts) using a modification of the technique explained in 3.6. We perform a comprehensive evaluation of the technique explained in section 3.7. We could show significant improvement in correctness when compared to the current methods used. The comprehensive test resulted in a 95 % correctness when answering equations and 93 % correctness when answering pie chart questions rendered in 4.3.2 for visually challenged participants. Moreover, our comprehension test showed an improvement in correctness with participants without vision impairment. The results of the comprehension test for equations and charts are given in tables 4.1.5 and 4.5 respectively. The sample equations rendered using different techniques and the experiments can be found at http://goo.gl/6YujaS and http://projects.venkateshpotluri.me/math

5.2 Major Challenges

Defining the problem statement and coming up with the techniques for this research effort had its fair share of hurdles. firstly, we had to decide on the scope of the problem. We felt that it might not be possible to encompass the entire jargon of scientific content. However, we wanted our approach and our solution to be as broad as possible with minimal fragmentation. That is, we wanted to develop a single method that would be applicable to render mathematical content of different types. To do this, we had to identify demarkations in equations that are common across different types of equations. We chose bracketing, fractions, superscripting and subscripting and underscript-overscript markup as the demarkations that require modification of cues. In addition to these, we observed from the human recorded equations that some mathematical symbols such as roots require modifications in the way they are spoken.

Another major challenge was to choose the right speakers. We initially recorded data from four speakers. However, one of the speakers had spoken mathematical content with heavy accent. The speaker's recordings had heavy influence of their native language. Though the equations were spoken correctly and were comprehendible, we were in the opinion that understanding equations with that particular accent may be a difficult task to many. In order to avoid the possibility of error, the decision to not use the recordings for experimentation was taken.

We used 3-D audio to render equations in 3.7. To achieve this, we had to choose the perfect set of elevation and azimuth angles to get the best 3D effect. There was no automated or exact way to do this. We rendered equations with all possible combinations, and listened to those with a significant intervals. We then chose a set of elevation and azimuth angles to render mathematical equations. Another major challenge was to understand the relation between a pie chart's pictorial characteristics and SVG markup. SVG markup had a lot of text and numbers. Reading the SVG data using a screen reader was a very tedious task. Moreover, SVG tags and

the pictorial characteristics did not have an easily understandable mapping. I will explain this with an example. The MathML tag for root is $< mroot >$. This is easy to understand to some extent. I had to depend on my colleagues to understand the relation between Pie chart SVG data and it's pictorial characteristics. There are no comprehensive tutorials on the web that explain this. Extra care had to be taken while performing experiments with visually impaired participants. These participants use screen readers to access the computer. We had to make the test accessible. Moreover, both the test and the screen reader convey information through audio. It is very important to ensure that the speech output of one does not interfere with the other. There was no programatic way to do this. The participants initially found this difficult. On explaining them the workaround (to mute the screen reader's speech as soon as the question page loads), the participants were able to perform the listening tests.

5.3 Publications And Presentations

Parts of our research have been accepted for publication at the International Conference on Natural Language Processing 2014 held in Goa, India in December 2014. The paper is titled Significance of Paralinguistic Cues in the synthesis of Mathematical Content [18]. The audience appreciated our approach to render mathematical equations in the reviews as well as at the presentation.

This research has been presented in the 30th Annual International Technology and Persons with Disabilities Conference, held in March 2015 in San Diego, USA. The presentation was titled Synthesis of Mathematical Content with Audio Cues. The audience acknowledged our research effort. The audience found relevance in the problem we chose to take on and felt that our approach was effective.

5.3.1 Where Do We Stand

I have recently attended a few presentations on rendering mathematics in audio at the 30th Annual International Technology and Persons with Disabilities Conference. Most of the implementations and improvements presented focused on navigating within the equations. However, there has not been much focus on presenting a holistic view of the equation. Efforts have been made by IBM and Freedom Scientific in its JAWS software to render mathematical content in audio. The updated implementations do not make use of cues (speech and non speech) to render mathematical content in audio. Specifically, I did not come across implementations that used sounds, pitch and rate variations and 3D audio. the implementation in JAWS16 relies on

mathML equations. To the best of my knowledge, they are introducing a math mode. In this mode, the user will be able to navigate the expression at different granularity. Though it is very essential to be able to navigate an expression at different levels, I strongly believe that it is important to be able to present a listener a holistic view of mathematical content. From IBM's presentation, I could conclude that their implementation will be integrated into analytics tools developed by IBM. It was clear that their implementation will not be available to the consumers directly in the form of a software package or a plugin of any form. IBM's implementation is aimed at making navigation across mathematical content easy. There has been emphasis on bar graph and other forms of statistical data.

Our approach is focused on providing an intended listener a holistic view of the equation. That is, the listener should be able to get most of the information by listening to the equation a minimum number of times.

5.4 Future Direction

Through this research, we have attempted to show a new approach to math accessibility by making effective use of speech and nonspeech cues. The idea was to demonstrate a very effective approach to the way math is read and understood by the print disabled. More importantly, the aim was to design and demonstrate an effective method for mathematical content to be spoken by any speech interface or system. It could be a screen reader, a virtual assistant or some form of a speech UI in the future.

Mathematical content is very different from regular english sentences. A regular TTS engine and the methods and rules to synthesise english sentences will not suffice to effectively render an equation or a piechart in a way that it could be understood by a person without looking at it. Hence, we performed the analysis explained in 3.2. We wish to make this available to people and we are making the code available on github. This research effort is just the beginning to my ambition to improve the state of the art in assistive technology. There are a lot more audience to whom mathematics are inaccessible. and there are a lot more people to whom various kinds of information is inaccessible. For instance, teaching something as simple as adding two two digit numbers is an almost out of reach and impossible task to children with multiple disabilities. By children with multiple disabilities, I am talking about those with finger dexterity limitations, Tactile impairments, etc. I conclude this document by expressing my hope that there should be a day when limitation to assistive technology should not be the reason for any individual with a disability to pursue education and learn something.

Appendix A

A.1 Equations recorded by Human voice

$$X + Y = z \tag{A.1}$$

$$\frac{X + Y}{K} = \alpha \tag{A.2}$$

$$(X + Y)^{P+Q} = X^{P*Q} + Y^{P} * Q - P + \frac{Q}{Y} - \frac{P}{Q - X} \tag{A.3}$$

$$\frac{(P + X) * (Q - Y)}{(X + Y)^{K}} = \frac{P}{X + K} - Q * (\frac{K^{x}}{Y - P}) \tag{A.4}$$

$$(X + Y)^{K} = 3 * X^{K} + 4 * X^{y} - 5Y^{K+X} \tag{A.5}$$

$$(X + Y)^{P+Q} = X^{P*Q} + Y^{P} * Q - P + \frac{Q}{Y} - \frac{P}{Q - X} \tag{A.6}$$

$$\frac{(P+X)*(Q-Y)}{(X+Y)^K} = \frac{P}{X+K} - Q*(\frac{K^x}{Y-P}) \tag{A.7}$$

$$\frac{X+Y}{K} = \alpha \tag{A.8}$$

$$(X+Y)^K = 3*X^K + 4*X^y - 5Y^{K+X} \tag{A.9}$$

A.2 Equations for testing the Systems

$$1+2+3-5+4+2+3 = (3+2)*(1+1) \tag{A.10}$$

$$\lim_{x\to+\infty} \frac{3x^2+7x^3}{x^2+5x^4} = 3. \tag{A.11}$$

$$\frac{\partial}{\partial x} x^2 y = 2xy \tag{A.12}$$

$$\frac{\partial u}{\partial t} = h^2 - E^{n+1} - 1 \tag{A.13}$$

$$\int_0^R \frac{2x\,dx}{1+x^2} = \log(1+R^2) \tag{A.14}$$

$$\int_0^{+\infty} x^n e^{-x}\,dx = n!. \tag{A.15}$$

$$(P+Q)^K + R = P^K * Q + Q^K * P + R^{P*Q} * K + \frac{P^Q * K + 1}{R} \tag{A.16}$$

$$(P+Q) * (R+K) = (P+R)^Q - (K+R^Q) + \frac{R+Q^K}{(R+Q)^K + 1} \tag{A.17}$$

$$(P+Q) * (R+K) = (P+R)^Q - (K+R^Q) + \frac{R+Q^K}{(R+Q)^K + 1} \tag{A.18}$$

$$\frac{X_1^K + X_2^K}{P_3^X * 5_4^x} + E^X = e^{\frac{X_{K+1} + X_{K+2}}{(X+Y)}} \tag{A.19}$$

$$\sqrt[P+Q]{A + K^P + A^{K+P}} = \frac{(K+P)(K-P)}{K * (P+K)} \tag{A.20}$$

$$\sum_{i=1}^{\infty} \frac{1}{i^2} + 5i + \sqrt[3]{i+1} = \frac{\pi^2 + 4\pi^3 + \sqrt[\pi+i]{9 * \pi}}{6} \tag{A.21}$$

$$(\frac{X+Y}{K} + 1)^3 = \sqrt[3]{X} + \sqrt[3]{Y} + (X * Y)/3 + \frac{X+Y}{3+K} + 3 \tag{A.22}$$

A.3 Equations used for Comprehension Test

$$(2+3) * (5+4) \tag{A.23}$$

$$\frac{(10+8)}{(7+2)} \tag{A.24}$$

$$\frac{3+6}{2+1} \tag{A.25}$$

$$(3+1)*2+1 \tag{A.26}$$

$$1+\frac{3+7}{2}-5 \tag{A.27}$$

$$(13-5)^2 \tag{A.28}$$

$$19+9^2 \tag{A.29}$$

$$(8-3)^2*4 \tag{A.30}$$

$$\frac{(2+6)^2}{(1+3)^2} \tag{A.31}$$

$$\frac{6+9}{8-2^2} \tag{A.32}$$

$$(9+7)*2^2 \tag{A.33}$$

$$9+(6-3)^3 \tag{A.34}$$

$$25-(2+1)^3 \tag{A.35}$$

$$(1+\sqrt{4})^2 \tag{A.36}$$

$$\sqrt{(5+3)*(2+2)} \tag{A.37}$$

$$\sqrt[3]{17-8}-6 \tag{A.38}$$

$$(2+1)^{\frac{27}{9}} \tag{A.39}$$

$$\frac{70}{9+5} \tag{A.40}$$

$$\frac{9}{25} + \frac{7}{5} - 4 \tag{A.41}$$

$$(9+6) * (9-5) \tag{A.42}$$

$$(3+1)^{\frac{66}{22}} \tag{A.43}$$

$$\sin(90-45)^2 \tag{A.44}$$

$$\cos(\frac{270}{3}) \tag{A.45}$$

$$\cos(60) \tag{A.46}$$

Bibliography

[1] J. Allen, M. S. Hunnicutt, D. H. Klatt, R. C. Armstrong, and D. B. Pisoni. *From text to speech: The MITalk system*. Cambridge University Press, 1987.

[2] A. W. Black and K. A. Lenzo. Limited domain synthesis. Technical report, DTIC Document, 2000.

[3] A. W. Black, P. Taylor, R. Caley, and R. Clark. The festival speech synthesis system. *University of Edinburgh*, 1, 2002.

[4] B. Caldwell, M. Cooper, L. G. Reid, and G. Vanderheiden. *Web content accessibility guidelines (WCAG) 2.0*, volume 11. W3C, 2008.

[5] L. A. Chang, C. White, and L. Abrahamson. Handbook for spoken mathematics. *Lawrence Livermore National Laboratory*, 1983.

[6] R. Fateman. How can we speak math. *Journal of Symbolic Computation*, 25(2), 1998.

[7] L. Frankel, B. Brownstein, and N. Soiffer. Navigable, customizable tts for algebra. 2014.

[8] M. Geronazzo, S. Spagnol, and F. Avanzini. A head-related transfer function model for real-time customized 3-d sound rendering. In *Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on*, pages 174–179. IEEE, 2011.

[9] F. Hinterleitner, G. Neitzel, S. Möller, and C. Norrenbrock. An evaluation protocol for the subjective assessment of text-to-speech in audiobook reading tasks. In *Proceedings of the Blizzard challenge workshop, Florence, Italy*. Citeseer, 2011.

[10] A. J. Hunt and A. W. Black. Unit selection in a concatenative speech synthesis system using a large speech database. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 1, pages 373–376. IEEE, 1996.

[11] P. Ion and R. Miner. Mathematical markup language. *Internet document http://www. w3. org/TR/WD-math*, 1998.

[12] A. Kuhn and J. Stacey. *Screen histories: a screen reader*. Oxford University Press, USA, 1998.

[13] R. C. Kurzweil, F. Bhathena, and S. R. Baum. Reading machine system for the blind having a dictionary, Mar. 7 2000. US Patent 6,033,224.

[14] L. Lamport. *I1 (\ LaTeX)—A Document*, volume 410. pub-AW, 1985.

[15] A. Nemeth. Mathspeak, 2005.

[16] A. Nemeth, N. B. Association, et al. *The Nemeth Braille Code for mathematics and science notation*. American Print. House for the Blind, 1973.

[17] C. Y. Nolan and C. J. Kederis. Perceptual factors in braille word recognition.(american foundation for the blind. research series no. 20). 1969.

[18] V. Potluri, S. Rallabandi, P. Srivastava, and K. Prahallad. Significance of paralinguistic cues in the synthesis of mathematical equations.

[19] A. Quint. Scalable vector graphics. *IEEE Multimedia*, 10(3):99–102, 2003.

[20] T. Raman. *Audio system for technical readings*. Springer, 1998.

[21] R. N. Schmidt, F. J. Lisy, T. S. Prince, and G. S. Shaw. Refreshable braille display system, Mar. 12 2002. US Patent 6,354,839.

[22] F. Scientific. Jaws-job access with speech, 2010.

[23] N. Soiffer. Mathplayer: web-based math accessibility. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, pages 204–205. ACM, 2005.

[24] C. Weaver. Nemeth code made easy.