AUDIO RENDERING OF MATHEMATICAL EQUATIONS

Venkatesh Potluri, Sai Krishna Rallabandi, Kishore Prahallad

International Institute of Information Technology, Hyderabad

ABSTRACT

Text to speech (TTS) systems hold promise as an information access tool for people with learning and print disabilities. However, audio rendering of mathematical equations using TTS is not very effective till date. In this paper, we address this problem by proposing five different techniques which exploit the paralinguistic cues such as pauses, special sounds, pitch variations and spatialization of speech. A subjective evaluation was performed on each technique. The evaluation considered 10 aspects such as listening effort, content familiarity, accentuation, intonation, etc. The work provides analysis on the different possibilities that can be employed to effectively render mathematics through audio.

Index Terms— Audio Rendering, Paralinguistic cues, ScreenReader, MathML, Spatialization

1. INTRODUCTION

Mathematical equations comprise of different types of visual cues to convey their semantic meaning. Some of these visual cues are superscripts, subscripts, parentheses,etc. Conveying the proper meaning and the information containing the semantics of a math expression to a listener without ambiguity and the loss of information (meaning) is a major challenge. The main objective of this paper is to form techniques which may unambiguously render the equations in audio.

1.1. Significance

Despite advances in screen reading and text to speech technologies, the problem of reading complex math remains majorly unsolved. Speaking the equation just as any other string of text, a line, or a sentence will not suffice to effectively render mathematics in speech. For instance, $e^{x+1} + 1$ denotes that the value e should be multiplied x+1 times before adding 1 to 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.

The major problem in rendering equations in audio is that we are making an attempt to convert information that is trivially perceived through one input mode, the vision to another input mode, audio. In order to succeed in efficiently accomplishing this conversion, we must map each of the visual elements that convey the information about various aspects of a mathematical equation to an auditory equivalent. The equations, when presented to a person in a visual form, say, on a paper, are available for him to refer at any point. On the other hand, when the person has to listen to the equation, he is presented with a holistic auditory view of the equation. Though there could be some level of granularity in presenting auditory information to the user, the challenge lies in matching the granularity of information in the audio rendering with that offered by visual representation of a mathematical equation.

Until recent times, mathematical equations were presented in documents and the web in the form of GIF and other image formats[1]. In some cases, equations presented in a document were pictures taken of hand written equations. Due to this, they were not accessible and TTS based softwares could not get any information from them. In addition to accessibility issues, presenting equations in documents using images caused problems while saving documents. In case of the web, presenting equations as images resulted in problems relating to loading time and saving the webpage containing the equation[1]. Fortunately, the advent of languages like LaTeX and MathML made it easier to write equations in documents and the web. Moreover, MathML and LaTeX contain information about the equation in their syntax.

One way to solve the problem of rendering mathematical equations in audio is to add additional information while speaking out the equation.

1.2. Review of the Existing Methods

There have been several attempts to present mathematical content through other modes than vision. Efforts have been made to formulate standards for presenting math through braille and speech. These efforts date back to 1946. Nemeth Code is a special type of Braille used for math and science notations. It was developed by Dr. Abraham Nemeth as part of his doctoral studies in mathematics. In 1952, the Braille Authority of North America (BANA) accepted Nemeth Code as the standard code for representing math and science expressions in Braille. With Nemeth Code, one can render all mathematical and technical documents into six-dot Braille [2] [3].Dr T.V Raman has developed an audio system for technical readings (ASTER). 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. Aster is not restricted to a specific markup. every thing that is required for it to recognise other markups is a recogniser for the system [4]. A more recent attempt has been made by a company called design science. They developed an internet explorer plugin called MathPlayer that displays and speaks out mathematical content marked up in MathML[5].

2. BASIS FOR THE PROPOSED TECHNIQUES

We recorded a set of equations spoken by people trained in teaching math to visually challenged students. The observations from these recordings helped us understand that the problems can be categorised as follows:

- Quantification problem
- Superscripting and subscripting problem
- Handling fractions

2.1. QUANTIFICATION

Most of mathematical equations contain expressions in parentheses. for instance, consider the equation (A+B)*(C+D)+E+F*G = K It may seem that the equation can just be treated as a general string of text while rendering it in audio. However, this will create a confusion in the listener. the equation is spoken as left parenthesis A plus B right parenthesis times left parenthesis C plus D right parenthesis plus E plus F times G equals K or A plus B times C plus D plus E plus F times G equals K. In the former case, the user will have to keep a track of all the parentheses when he listens to the equation. This becomes a hectic task for bigger equations and also results in deviating the listeners attention from concentrating on the actual contents of the equation. On the other hand, in the latter case, the user gets an ambiguous representation of the equation. He may interpret the spoken form of the equation as A+(B*C)+D+E+F*G = K or (A+B)*(C+D+E+F)*G = K or (A+B)*(C+D+E+F*G) = K We will have to add additional information to the equation to solve this ambiguity.

2.2. SUPERSCRIPTING AND SUBSCRIPTING:

Todays 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 often 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 would 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 treats the expression as a complete word. consider the expression A^B. The TTS speaks it as ab. In case of numbers, say we have an expression 5^25, the TTS reads it as five hundred twenty five or five two five. We come across the same issues while trying to render subscript text. In addition to these problems, the real challenge lies in effectively conveying the spacial orientation of the different parts of the equation. That is, the equation, rendered in audio must give the listener a view of what content is in the superscript and the subscript. The listener must also be notified when the part of a mathematical expression in the superscript or subscript ends; 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. We must provide the user with different cues for superscript and subscript content.

2.3. HANDLING 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 user 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. The audio equivalent of the equation should effectively be able to convey nested fractions in addition to the regular fractions to the listener.

3. SELECTION OF THE EQUATIONS

Selection of suitable equations is a critical component to evaluate the proposed systems. The goal of this experiment is to identify the impact the length of the equation and the number of the variables in the equation have on the listeners ability to remember the equation. The listener is asked to reproduce(write) the equation after making him/her listen to it. The listener will be given equations recorded by special educators. The rational behind using the equation recorded by a real human is that the listeners response should not be effected by the audio quality (voice, accent, etc).

We have identified certain parameters that would affect the listener such as:

- Number of terms in the equation.
- Length of the audio utterance of the equation.
- Number of different variables in the equation.
- Number of times each variable occurs in the equation.
- Number of times the user repeats listening to the equation.

We hand picked a set of 15 equations and rendered them in audio. The equations consist of different combinations of mathematical attributes that must be rendered differently. The

equations also vary in length and the number of variables. Longer equations with less number of variables can be remembered easily where as equations with more number of variables can be difficult to remember. We made sure that the selected equations are semantically unrelated.

4. PROPOSED SYSTEM AND TECHNIQUES

Based on the audio recordings from special educators and teachers, We could form a few rules and techniques to render mathematics in audio. We developed systems that incorporated these rules and techniques to render math in audio. The system takes presentation markup MathML as the input format for math and outputs audio in wav format. We made use of festival for speech synthesis. The equations are converted to SABLE markup and are given to the festival speech synthesis system for speech generation. Some of the sounds used in the generation of math audio are all licensed under creative commons.

4.1. TECHNIQUE 1 : General system with Verbal Cues

We developed a system that would render mathematical equations in audio just as any other text to speech system would do. The only addition is that the audio contains spoken form of some of the visual elements such as superscript, subscript, etc. The system converts the mathML into text and speaks it out.

4.2. TECHNIQUE 2 : Equations with Pauses and Special Sounds

We make use of additional paralinguistic cues to render the mathematical equation. These cues include; Pauses to convey certain parts of an equation. Sounds to indicate certain symbols and mathematical operations. E.g.: Sounds to indicate superscripts, subscripts and parentheses. We chose sounds that would be pleasant to the ear and would be passively be noticed by a user. That is, the sounds will not be too obtrusive to the listener if he is paying more attention to what is being spoken. At the same time, The sounds will not go completely unnoticed even if the user is not paying complete attention and expecting a sound. We used sounds such as the ding and a few variations of the sound.

4.3. TECHNIQUE 3: Equations with Pitch and Rate Variations

We use pitch and rate changes to denote certain mathematical attributes. The pitch and rate increase while speaking out the superscript text and decrease while speaking the subscript text. The system does the same with 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 1. Sets of HRTF angles for audio spatialization

Term	Elevation Angle	Azimuth Angle
Superscript	50	837
Subscript	47	877
Fraction	31	25
Underscript	35	144
Overscript		

4.4. TECHNIQUE 4: Equations with Audio Spatialization

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 gets 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) [] . Table 1 shows the sets of angles chosen for the different parts of the equation such as superscript, etc.

4.5. TECHNIQUE 5: Equations with Pitch variations and Special tones

We rendered the equations in audio by varying the pitch, adding pauses, emphasising the speech and added sounds at required parts of the math equation. As explained in technique 3, We have made pitch and rate manipulation while rendering superscripts, subscripts, fractions, under scripts and overscripts. In addition to the paralinguistic cues, 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 used in rendering math using technique 2.

5. EVALUATION

We perform listening tests to evaluate the proposed techniques. Each participant is made to listen to 6 equations and his responses are recorded. The participent is made to listen to 3 equations from 1 system. We test 2 systems with each participant. We do a subjective evaluation of the system based on the participents responses. The participant will have to reproduce the equation he listens to. In addition to the equation, the listener will have to evaluate the system based on a few other parameters. We arrived at these parameters partly by following the listening test procedures followed in blizzard challenges and our own analysis.

5.1. PARAMETERS

We record the participants reproduced version of the equation. We have three cases: correct, partially correct and wrong. We

Table 2. Nonlinear Model Results

Parameter	Human	Technique#1	Technique#2	Technique#3	Technique#4	Technique#5
Intonation Variation	50	837	970			
Pitch Variation	47	877	230			
Pauses	31	25	415			
Listening Effort	35	144	2356			
Content Familiarity						
Effectiveness of additional cues						
Accentuation						
Number of repititions						
Overall Comfort	45	300	556			

get the intelligibility of the audio rendered by our systems. On a scale of 1-5, we ask the user to evaluate our systems on the following

- listening effort.
- intonation(1=ineffective and 5 = very effective)
- acceptance(1=poor, 5=good).
- speech pauses(1=not noticeable and 5=very prominent) 5 means that the pauses are having a negative impact on the participants ability to understand the equation in audio.
- accentuation(1=poor and 5=very prominent). [6] Similar to speech pauses, 5 indicates that the speech is too accentuated.
- content familiarity.(1=totally new concept and 5 = very familiar). 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 = disturbing).
- Number of repetitions of each equation.

6. RESULTS AND CONCLUSION

7. REFERENCES

List and number all bibliographical references at the end of the paper. The references can be numbered in alphabetic order or in order of appearance in the document. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [**?**].