AUDIO RENDERING OF MATHEMATIC EQUATIONS

ABSTRACT

Text to speech systems are evolving at a great pace in the recent times. Voices are sounding more natural and different kinds of information are becoming accessible. However, a major area in which the current day TTS systems and screen readers still trail is the mathematical and scientific equations. Mathematics has a significant amount of information hidden in the spacial arrangement of numbers and symbols. Reading mathematics like any other English sentence is not the most effective way of auditory communication of math to the listener. This document talks about a few methods we have formulated to effectively convey the meaning of mathematical equations to a listener.

KEYWOARDS

Mathematics, audio rendering, paralinguistic cues, screen reader, MathML

INTRODUCTION:

Mathematical equations, or any scientific equations for that matter, comprise of different types of visual cues to convey the meaning of that equation. Some of these visual cues are superscripts, subscripts, parentheses, etc. For instance, 3^5 denotes that the value 3 must be multiplied with itself 5 times. (a+b)*(c+d) means that a and b, and c and d must be added and the individual results should be multiplied. Conveying the proper meaning and the information including the semantics of a math expression to a listener without ambiguity and the loss of information (meaning) is a major challenge. Despite significant advancements in screen reading and text to speech technologies, the problem of reading complex math still remains unsolved. Until recent times, mathematic 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 about 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 web-page containing the equation. [1]

Fortunately, the advent of languages like LaTeX and Math ML made it easier to write equations in documents and the web. Moreover, Math ML and LaTeX contain information about the equation in their code or syntax. Reading out the equation just as any other string of text, a line, or a sentence will not suffice to effectively communicate the mathematic equations in audio. One way to solve the problem of rendering mathematic equations in audio is to add additional information while speaking out the equation. This document gives an understanding of the problem. It gives an idea about the various scenarios that must be handled and the required type of additional information and cues that must be provided to effectively render mathematic equations. We hand-picked a set of equations and recorded the spoken forms. The equations were made to be spoken by special educators and people experienced in teaching mathematics to visually challenged students. We formed a set of rules based on the observations from these recordings and developed a set of systems that render these equations in

audio using these techniques. These rules and the system will be explained in further sections of this article. We performed tests on listeners and compared the effectiveness of each of our developed techniques.

PREVIOUS WORK

There have been several attempts to present mathematical content through modes other than vision. Efforts have been made to formulate standards for presenting math through brail 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. All that it requires to recognize other markups is a recognizer for the system.[4]. A more recent attempt has been made by a company called design science. They developed an Internet explorer plug-in called Math Player that displays and reads out mathematical content marked up in Math ML. [5]

PROBLEM

The major problem in rendering equations in audio is that we are making an attempt to convert the 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 mathematic 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. If an equation is treated as a regular string of text by a TTS engine, information about the mathematical equation is not conveyed completely. The problems can be categorized as follows:

1. Quantification problem
2. Superscripting and subscripting problem
3. Handling fractions

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 an equation can just be treated as a general string of text while rendering it in audio. However, this will create confusion for 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 listener’s 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.

SUPERSCRIPTING AND SUBSCRIPTING:

Today’s screen readers and TTS engines do not effectively convey equations with superscript and subscript content. They often do not speak out the parts of the equation contained in the superscript and subscript or 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 5^2^5, 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 mathematic expression in the superscript or subscript ends; The listener should understand that anything 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.

HANDLING FRACTIONS

Fractions, like the other mathematical concepts discussed above cannot 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.

THE DATA:

We handpicked 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 whereas equations with more number of variables can be difficult to remember. We also made sure that the selected equations are not familiar to the user. That is, the user may not have come across those equations in any textbook. However, the listener may have come across parts of the equation but will not be able to predict the equation. We selected such equations to make sure that the listener's previous knowledge should not enable him to guess the equation. However, the listener requires domain knowledge in math in general and specific domains like algebra is required to understand the equation. We conducted experiments to determine the impact the length and the number of variables can have on a person's ability to remember and reproduce the equation.

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. We chose presentation markup Math ML as the input format.

TECHNIQUE1:

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 Math ML into text and speaks it out.

TECHNIQUE 2:

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 to obtrusive to the listener if he is paying more attention to what is being spoken. At the same time, the sounds will not get 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.

TECHNIQUE 3:

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.

TECHNIQUE 4:

In this technique, We made an attempt to draw a closer analogy to the spacial positioning of various variables and numbers of a mathematic equation. The listener gets the illusion that the superscript part of a math expression is spoken from above his head and the rest at the usual level. We identified areas that require this 3-dimensional audio manipulation by detecting for pitch changes in the equation. We then calculated the required HRTF function to do the necessary manipulation.

TECHNIQUE 5:

We rendered the equations in audio by varying the pitch, adding pauses, emphasizing 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 over scripts. 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.

THE SYSTEM

The system takes presentation markup Math ML 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.

EXPERIMENTS

EXPERIMENT 1

This experiment is to identify the impact that the length of the equation and the number of the variables in the equation have on the user's ability to remember the equation. We will ask the listener to reproduce (write) the equation after making the user listen to it. The listener will be given equations recorded by special educators. The rational behind making the listener listen to the equation recorded by a real human is that the listener's response should not be effected by the audio quality (voice, accent, etc). If the person is made to listen to outputs from a TTS engine, He may not be familiar with listening to TTS audio and as we know, the current day TTS systems do not read math accurately and that is primarily the problem we are trying to address.

DATA:

We have identified certain parameters that would affect the listener's ability to remember a mathematical equation when it is presented to him in audio. We are interested in the following data with regards to this experiment.

- 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.

EXPERIMENT 2

We perform experiment 2 to evaluate the systems we developed. Each participant is made to listen to 6 equations and his responses are recorded. The participant 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 participant's responses. The participant will have to reproduce the equation he listens to. In addition to the equation, He or she 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.

THE DATA

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