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. The objective of this paper is to propose techniques which may unambiguously render the equations in audio. It provides a subjective analysis on the performance of each of the proposed techniques.

Section 1.1 talks about the significance of the problem and section 1.2 gives an overview of some of the existing methods relevant to audio rendering of mathematical content.

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 subtracting 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. To effectively do this, we must map information presented through visual cues such as spatialisation to their auditory equivalent. Mathematics, in its visual form, gives the reader a very high level granularity in perceiving the equation. The advent of markups like MathML give way to programatically identify the visual cues and the semantics of a mathematical expression. The ability to identify the visual cues can be used to enable text to speech systems to speak mathematics with minimum or no ambiguity. Rendering mathematics in audio can also be very advantageous to people with various print disabilities like visual impairment, dislexia, etc.

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 alternative modes to vision. Efforts dating back to 1946 have been made to formulate standards for presenting math through Braille and speech. Nemeth Code[1] 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. Dr T.V Raman has developed an audio system for technical readings (ASTER)[2]. 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 that displays and speaks out mathematical content marked up in MathML[3].

There have been attempts to form a set of guidelines to effectively speak mathematics in audio. The handbook for spoken mathematics [4] gives an account of such an attempt. An article on how to speak math also describes the challenges in speaking mathematics to and by a computer [5].

Section 2 explains the motivation for the proposed ideas. Sections 3.1 through 3.5 explain the proposed techniques. Section 4 gives the criterion behind selecting the equations. Section 5 explains the experiment and 5.1 outline the evaluation criteria. Section 6 shows the results.

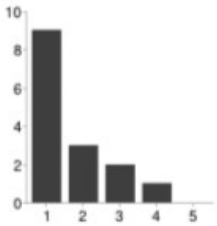
2. BASIS FOR THE PROPOSED TECHNIQUES

A set of equations(A) were formed and were recorded by people trained in teaching math to visually challenged students. We noticed a difference in the speaking pattern while certain parts of the mathematical expressions were spoken. The variations used by the speakers to effectively convey the equations are:

- Pauses.
- Intonation Variations.
- Pitch Variations.

A subjective analysis was performed taking the listening effort and the comprehendibility of the equation into consideration. The results from this experiment are shown in the figure below. It was observed that the highest listening effort was encountered when the equations had the following issues :

- Quantification.
- Superscripting and subscripting.
- Handling fractions.



Listening Effort

Fig: Evaluation Results of equations spoken manually

We now provide more details on the rational behind the idea of enhancing these aspects of a mathematical expression while rendering it in audio.

2.1. QUANTIFICATION

Most of mathematical equations contain expressions in parentheses. for instance, consider the equation $(A + B) * (C + D) + E$ 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 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.

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^{2^5}, the TTS reads it as five hundredtwenty 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 spatial 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.

The proposed techniques are aimed at enhancing the above mentioned aspects in the expressions using the cues

intuitively employed by humans to speak math.

3. 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.

3.1. 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.

3.2. TECHNIQUE 2 : 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 getting a glance 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. We make an attempt to provide the equation such that a person gets a similar advantage when he listens to it. We make use of special sounds or ear cons while presenting the equations. It may seem obvious that text to speech is the easiest way to convey equations and speaking every detail provides direct information. However, listening to speech results in more mental stress. For instance, We can not converse when we are listening to speech from an other source[6]. Replacing speech with sounds is not the most effective way to tackle the problem of presenting mathematic equations in audio. We make 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 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 .

Table 1. Pitch and Rate Variations

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

- 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 sounds that would be pleasant to the ear and that are passively noticed by a listener. 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 listener is not paying complete attention and expecting a sound. We used sounds such as the ding and a few variations of the sound.

3.3. TECHNIQUE 3: Equations with Pitch and Rate Variations

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 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 used 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 shows the pitch and rate variation(in percentage) that is applied to the mathematical equation. The variation is with respect to the base pitch and rate of the TTS in general.

3.4. TECHNIQUE 4: Equations with Audio Spatialisation

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) [7]. Table 2 shows the sets

Table 2. 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

of angles chosen for the different parts of the equation such as superscript, etc.

3.5. 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.

4. 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 was asked to reproduce(write) the equation after making him/her listen to it. The listener will be given equations recorded by special educators. The rationale 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 hand picked a set of 12 equations(Appendix B) 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. It was ensured that the selected equations were semantically unrelated.

5. EVALUATION

Each participant was made to listen to 18 equations and his responses were recorded. Equations 1-3 were recorded by people familiar with reading math to people with print disabilities. Equations 3 to 6 are rendered with the methods outlined in section 3.1, equations 6 to 9 are rendered using methods outlined in 3.2, equations 10 to 12 using methods from 3.3, equations 13 to 15 with 3.4 and 16 to 18 with 3.5.

22 participants wer made to perform the subjective analysis of the systems. The participant will have to reproduce the equation he/she 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 [8] and our own analysis.

5.1. PARAMETERS

On a scale of 1-5, the participants were asked to evaluate our systems on the following parameters.

- 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). [8] 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 = very helpful).
- Number of repetitions of each equation.

6. RESULTS AND CONCLUSION

Table 3 contains the normalized scores(1 to 5) calculated over the responses for all the equations. In case of the number of repitions of the equation, the mode value(most occuring value) is presented.

6.1. ANALYSIS

On analysing the data from the experiment(section 5), it is observed that the participants are able to understand the human spoken equations. More over, it can be clearly seen from

Table 3. Evaluation of the proposed systems

Parameter	Human	Technique#1	Technique#2	Technique#3	Technique#4	Technique#5
Intonation Variation	4.26	1.6	2.3	**4.7**	4.32	4.68
Pitch Variation	4.17	1.4	1.4	4.43	**4.82**	4.36
Pauses	3.1	2.15	**4.15**	3.7	3.7	3.87
Listening Effort	2.5	4.4	3.5	**2.3**	2.64	2.47
Content Familiarity	2.7	2.7	2.7	2.7	2.7	2.7
Effectiveness of additional cues	3.2	1.2	1.82	4.32	**4.37**	4.23
Accentuation	**4.3**	2.5	2.3	3.2	3.6	3.47
Number of repititions(Mode)	2	4	3	2	2	2
Mean Opinion Score	4.42	1.89	2.27	4.37	**4.62**	4.35

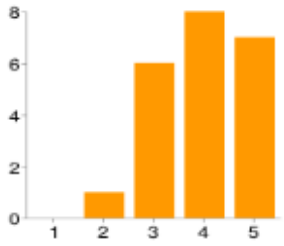


Figure 2:Overall Acceptance of the systems

table 3 that generating spoken forms of mathematical equations without making any enhancements(technique 1, section 3.1) 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 3.2 will not suffice either. Also,it is evident that the pitch and rate changes while rendering certain parts of the mathematical expressions have proven to be helpful to the participants in comprehending the expression. In the method described in section 3.4, the user has been able to draw an annalogy to the print form of mathematics. Table 3 shows that the participants had to put minimum effort in understanding the equation. It has been observed that the method explained in section 3.2 did not prove to be helpful to the listeners. However, from the table 3 and the values corresponding to the technique explained in section 3.5, it is evident that use of cues(pauses and rate variations) in addition to special sounds can be significantly effective in helping a listener. The key difference between technique 3(section 3.3) and technique 5(3.5 is the addition of sounds. Moreover, it can be observed that pauses were more helpful in understanding equations rendered using technique 5 when compared to technique 3. Pauses were more noticeable in technique 3 and when it is combined with technique 2, it enhances the effectiveness of the pauses. From this observation, we can conclude that a combination of these techniques may prove to be more effective in some cases. However, it is difficult to conclude that a specific technique is the ideal way of rendering mathematical equations in audio.

Figure 2 shows the overall acceptance of each of the proposed techniques as rated by the participants. The proposed techniques show an improvement over the output that can be achieved from the traditional TTS systems.

we

A. EQUATIONS RECORDED BY HUMAN VOICE

$$X = Y$$

$$X + Y = z$$

$$\frac{X+Y}{K} = \alpha$$

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

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

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

$$\frac{X+Y}{K} = \alpha$$

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

B. EQUATIONS FOR TESTING THE SYSTEMS

$$1 + 2 + 3 - 5 + 4 + 2 + 3 = (3+2) * (1+1)$$

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

$$\frac{\partial}{\partial x} x^2 y = 2xy$$

$$\frac{\partial u}{\partial t}$$

$$= h^2 - E^{n+1} - 1$$

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

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

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

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

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

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

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

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

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

C. REFERENCES

[1] Abraham Nemeth, National Braille Association, et al., *The Nemeth Braille Code for mathematics and science notation*, American Print. House for the Blind, 1973.

[2] TV Raman, *Audio system for technical readings*, Springer, 1998.

[3] Neil Soiffer, "Mathplayer: web-based math accessibility," in *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2005, pp. 204–205.

[4] Larry A Chang, CM White, and L Abrahamson, "Handbook for spoken mathematics," *Lawrence Livermore National Laboratory*, 1983.

[5] Richard Fateman, "How can we speak math," *Journal of Symbolic Computation*, vol. 25, no. 2, 1998.

[6] Tilman Dingler, Jeffrey Lindsay, Bruce N Walker, et al., "Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech," in *Proceedings of the 14th International Conference on Auditory Display, Paris, France*, 2008, pp. 1–6.

[7] Michele Geronazzo, Simone Spagnol, and Federico 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*. IEEE, 2011, pp. 174–179.

[8] Florian Hinterleitner, Georgina Neitzel, Sebastian Möller, and Christoph 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.