AUDIO RENDERING OF MATHEMATICAL EQUATIONS

Venkatesh Potluri, Saikrishna Rallabandi, Kishore Prahallad

speech and vision lab, International Institute of Information Technology, Hyderabad

ABSTRACT

Text to speech (TTS) systems hold promise as an information access tool for literate and illiterate including visually challenged. Current TTS systems can convert a typical text into a natural sounding speech. However, auditory rendering of mathematical content, specifically equation reading is not a trivial task. 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. Earlier works have attempted to use pauses as acoustic cues to indicate some of the semantics associated with the mathematical symbols. In this paper, we first perform an experiment to measure the effectiveness of mathematical equations synthesised using a traditional TTS system. We then analyse the acoustic cues which human-beings employ while speaking the mathematical content to (visually challenged) listeners and then propose four techniques which render the observed patterns in a text-to- speech system. The evaluation considered eight aspects such as listening effort, con- tent familiarity, accentuation, intonation, etc. Our objective metrics show that one of the proposed techniques could render the mathematical equations using a TTS system as good as that of a human being. To strengthen this observation, we performed a comprehension test for one of the proposed techniques on participants with and without vision impairment.

Index Terms— mathematical equations, multidimensional audio, paralinguistic cues, Text To speech

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. Despite advances in screen reading and text to speech technologies, the problem of speaking 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 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. 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.

1.1. Quantification

Most of mathematical equations contain expressions in parentheses. For instance, considering the equation $(A+B)*(C+D)+E$, it may seem that the equation can just be treated as a general string of text while speaking. However, this will create a confusion in the listener, as there are two ways of expressing.

- ”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 spoken form of the equation should have additional information to solve this ambiguity.

1.2. Superscript and subscript

Today's screen readers and TTS engines do not effectively convey the equations with superscript and subscript content. They often speak out such content continuously, with the rest of the equation. They also treat variables in the baseline and the superscript as an english word. There is a similar problem with numbers. For instance, let us say the expression is E^X. With the currently available technologies, the expression may be rendered as ”EX”. Consider the expression A^B. The TTS may speak it as ”ab”. In case of numbers, the expression 5^{2^5} may be spoken as ”five hundred twenty five” or

”five two five”. We come across the same issues while trying to render subscript text. If such equations are to be rendered effectively, the listener should be able to differentiate between superscript, subscript and the rest of the equation. This can be achieved using cues.

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

There have been several attempts to present mathematical content through alternative modes to vision. Efforts 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. This code could also be used to speak mathematical content. 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 [3] that displays and speaks out mathematical content marked up in MathML [4]. The handbook for spoken mathematics [5] is an attempt to form a set of guidelines to effectively speak mathematics in audio. An article on how to speak math [6] also describes the challenges in speaking mathematics to and by a computer.

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.

The objective of this paper 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 to map the visual cues to their auditory equivalents to programatically and unambiguously render the mathematical content in audio using a text-to-speech system. We then evaluate these techniques and perform a comprehensive evaluation on one of the techniques.

Section 2 discusses the basis for the study and section 3 gives the results and sets the tone for the idea behind the research. Section 4 discusses the proposed ideas and presents the analysis of the qualitative study performed. section 5 explains the comprehension test and section **??** gives the observations from the test.

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 equations spoken by human beings.

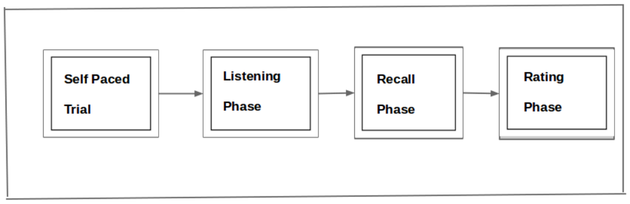


Fig: Evaluation procedure

A set of 15 participants were made to listen to the synthesised 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. A similar procedure was followed with equations recorded by trained human speakers. The participant evaluated the spoken equation based on eight parameters, i.e., perform objective analysis. We arrived at these parameters partly by following the listening test procedures followed in the Blizzard challenges [7] and our own analysis.

The parameters considered for evaluation include 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), Number of repetitions of each equation and Effectiveness of additional cues such as sounds, pitch and rate variations, change in direction, etc. (1 = hardly noticeable and 5 = very helpful). In case of content familiarity, 1 indicates that the user is not acquainted to the terminology used in the equation. The participants’ response for that particular equation can not be considered 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.

Table 1. Evaluation of Spoken Math vs TTS

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

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. Each of the equations is semantically unrelated. 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 or synthesis. The reason behind choosing such equations is to ensure that the listener's prior knowledge does not influence the ability to recall the equation. This ensures that the user's recall is based on the synthesised equation, not his prior knowledge.

3. INFERENCES FROM THE LISTENING TESTS

The results of this experiment, shown in the Table 1 indicate 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) handling 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.

4. PROPOSED TECHNIQUES

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.

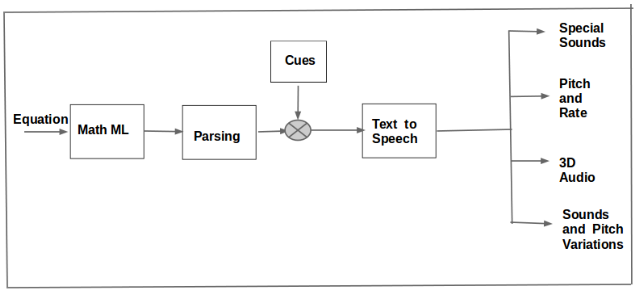


Fig: Overall framework for the proposed techniques

An example depicting the workflow of the entire algorithm is shown in the Figure 1. For the sake of illustration, a simple expression , $(X + Y)^{4-2}$ was taken :

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.

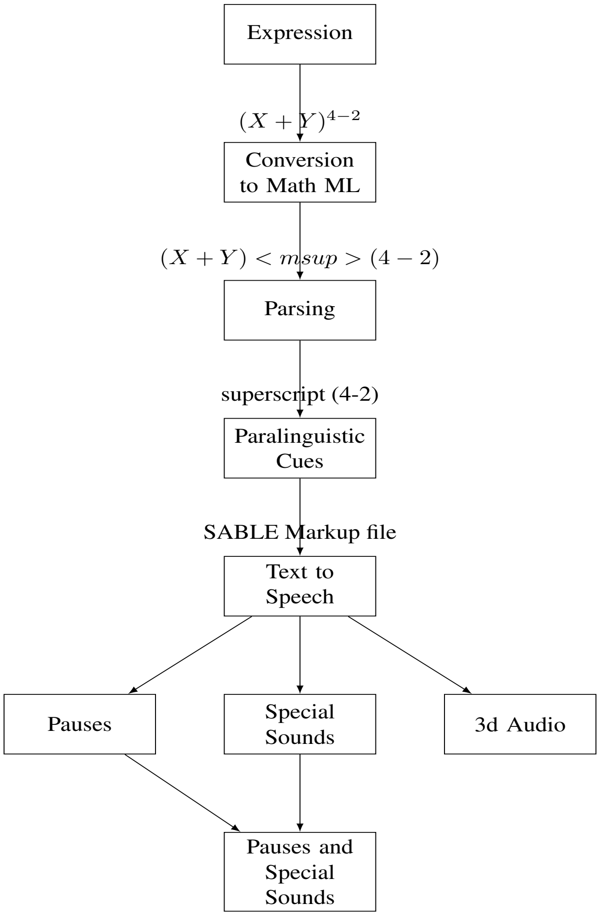
- Subscripts and superscripts
- Fractions
- Square root terms
- Overscripts and underscript

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 markup language [8]. The SABLE file is input to the speech synthesis system which generates the audio form of the equation with specified pauses and intonation. We have generated the audio files using the Festival Speech Synthesis System[9]. Sections 4.1 through 4.4 discuss each of the four proposed techniques.

4.1. 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 getting a glance at the equation. A person has the advantage of perceiving a

Fig. 1. Example Synthesis using a simple expression



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 additional paralinguistic cues such as **Pauses** to convey certain parts of an equation and **Sounds** to indicate certain symbols and mathematical operations. These pauses are mainly used to separate the parts of mathematical expressions. Consider $(A+B)^2$. It would sound more natural and intuitive if the expression is spoken as "the quantity A + B pause superscript 2 " 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.

4.2. Technique 2 : Rendering 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 3, we noticed 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 is employed to properly render fractions. The numerator is spoken in a higher pitch and the denominator is spoken in a lower pitch. Quantities in a root are spoken at a faster rate. The variation is with respect to the base pitch and rate of the TTS.

4.3. Technique 3: Rendering 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 in print. 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) [10]. Table 2 shows the sets of angles chosen for the different parts of the equation such as superscript, etc.

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

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.

Table 3. Evaluation of the proposed techniques

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	

4.4. 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 4.2 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. The sounds used here are the same as the ones mentioned in section 4.1. The Pitch and rate variations that are introduced are the same as those used in section 4.2.

4.5. evaluation metrics of the 4 techniques

5. EVALUATING THE COMPREHENSION

From the evaluation of the techniques presented in table 3, we could observe that people were able to get a clearer understanding of the equations synthesised using the technique explained in 4.3. To strengthen this observation further, We perform another experiment to evaluate the comprehend ability of the equations when they are rendered in audio with spacial orientation. We perform this experiment on participants with and without vision loss.

5.1. the experiment

The experiment contains equations rendered using a traditional TTS and the idea explained in section 4.3. The equations contain numbers and basic mathematical operations and are designed such that the participants will be able to solve them mentally. The participant will be made to listen to equations of both types in random order and he or she will have to enter the response for each of the equations.

6. RESULTS AND CONCLUSION

The results will be put here.

6.1. conclusion

Conclusion goes here.

7. 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] Patrick Ion and Robert Miner, "Mathematical markup language," *Internet document http://www. w3. org/TR/WD-math*, 1998.

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

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

[7] 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.

[8] Richard Sproat, Andrew Hunt, Mari Ostendorf, Paul Taylor, Alan Black, Kevin Lenzo, and Mike Edgington, "Sable: A standard for tts markup," in *The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*, 1998.

[9] Alan W Black, Paul Taylor, Richard Caley, and Rob Clark, "The festival speech synthesis system," *University of Edinburgh*, vol. 1, 2002.

[10] 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.