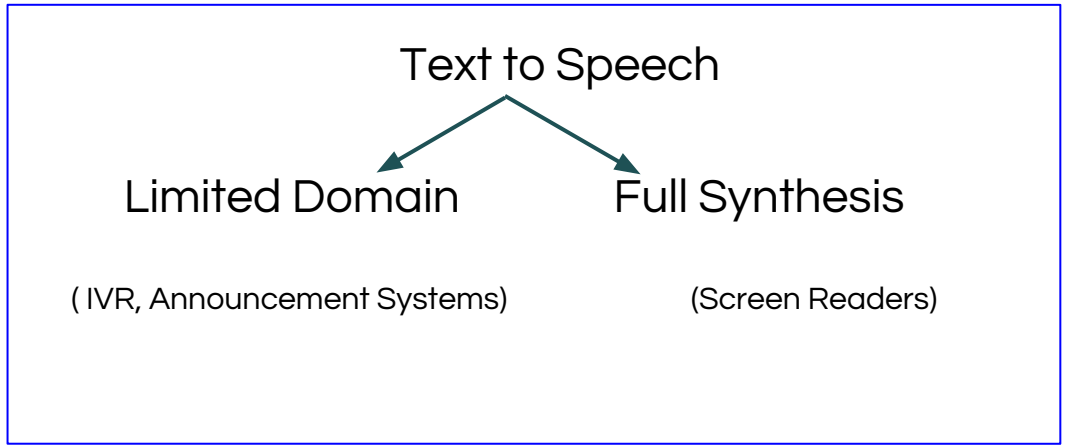
Synthesis of Mathematical Content with Audio Cues

Venkatesh Potluri, Sai Krishna Rallabandi, Priyanka Srivastava,
Kishore Prahallad,
Speech and Vision Lab,
IIIT- Hyderabad

Introduction

Text to speech systems convert a text into spoken form.

Used in a wide variety of applications.



Our context of discussion : Screen Readers

Screen readers are software that enable users to navigate and listen to the content on screen.
examples: JAWS, NVDA, VoiceOVer

Problem

Mathematical content has a lot of information present in visual cues.

Speaking a mathematical equation like a traditional english sentence will not suffice to convey its meaning.

Challenge:

Translating these visual cues to their auditory equivalents

Key Issues in Audio Rendering

Quantification

(A+B)*(C+D)

Superscripting and Subscripting

e^(x+1)+e^(x)+1

Fractions

(3x + 7)/5

Quantification

Tough to identify different parts of an expression enclosed in parentheses
example:
(A+B)*(C+D)

Possible spoken outcomes:

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

"A plus B times C plus D"

Superscripts and Subscripts

Difficult to identify the beginning and end.

Expression:
1+P^Q+K

Spoken form:

"1 plus P Q plus K"
"1 plus P power Q + K"

Superscripts and Subscripts

More Examples

Expression:
A^B
Spoken form:
"ab"

Expression:
2^2^3
Spoken form:
"2 2 3"
"two twenty three"

Fractions

Difficult to identify the beginning and end of fraction.

Expression:
1+A/B+2
Spoken form:
"1 plus A over B plus 2"

Expression:
P+1/Q +1
Spoken form:
"P plus 1 over Q plus 1 "

Current methods followed by visually challenged individuals

Cited assistance

The abacus is used to perform basic arithmetics

Tailor frame is a device used to perform basic algebra and statistics.

Shortcomings:
These devices do not facilitate storage and require additional skill set.

Attempts to render Mathematical content in Braille

Nemeth code is a form of 6-dot braille used by Dr Nemeth to write mathematical content.

Dr Nemeth used this to have math spoken to him.

<insert Nemth code example here.>

[1] Craig, Ruth H. *Learning the Nemeth Braille Code: A Manual for Teachers*. Brigham Young University Press, 1979.

Attempts to render Mathematical content in alternate forms

ASTER by Dr T.V.Raman

MathPlayer by design science

ChromeVox project by google

Difference:

We emphasise on effectively using paralinguistic cues to render mathematical content in audio

[2] Raman, T. V. "AsTeR: Audio system for technical readings." *Information Technology and Disabilities* 1.4 (1994).
[3] Soiffer, Neil. "MathPlayer: web-based math accessibility." *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2005.

Significance of Cues

Cues help in identifying demarcations such as Bracketing, subscripting and superscripting, etc.

Examples of Cues:

- Pauses
- Intonation and Pitch Variations
- Verbal Cues

Asked Humans to record the equations

Details of the Experiment

Equations were from the domain of the participants' knowledge.

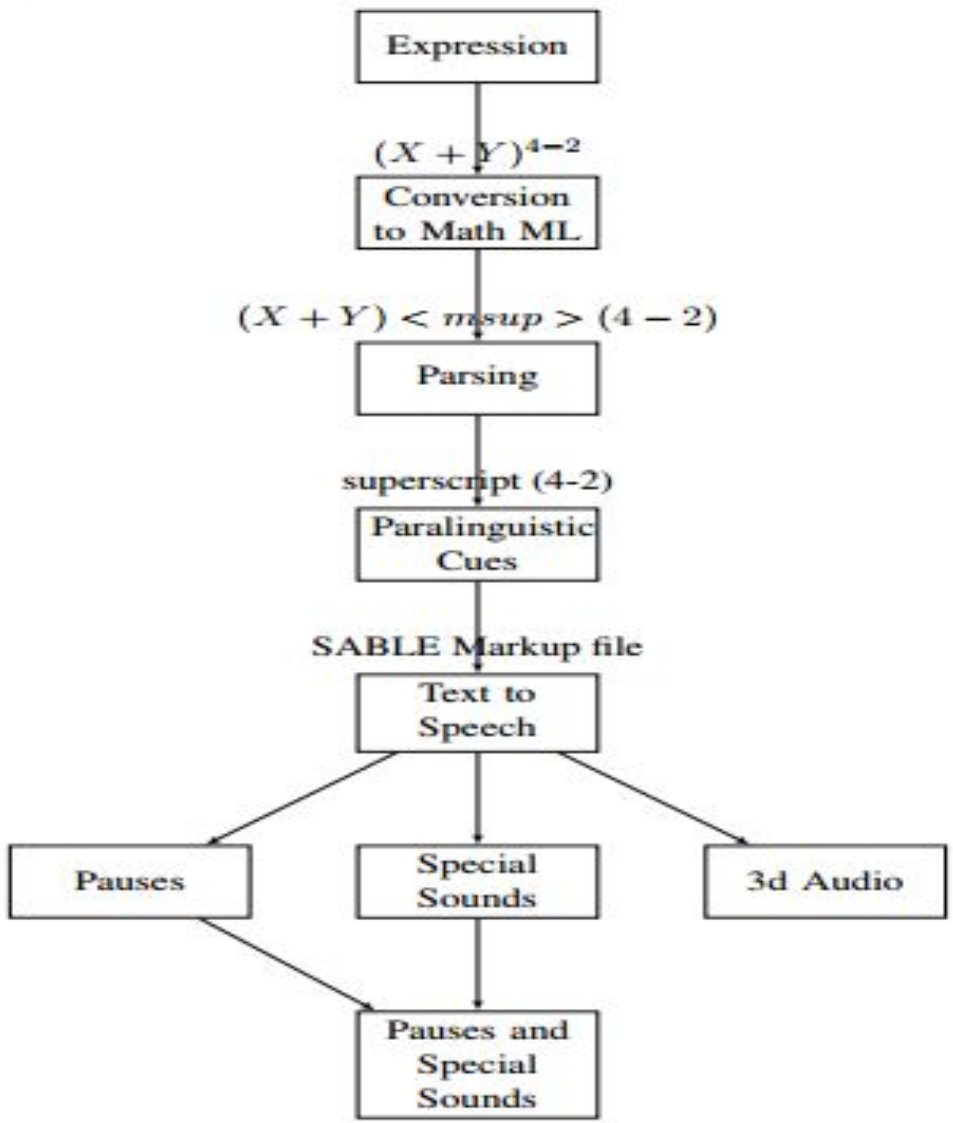
Chose 4 speakers (formally trained and trained by experience)

Had to omit one speaker due to heavy native language accentuation.

Participants were made to listen to equations from both sets

<insert results table

System Design



Rendering Equations using Sounds

Use Sounds to alert the user for demarcations.

Use pauses to separate subexpressions and certain demarcations.

Rendering Equations using Variations:

Use of Pitch and Rate Variations

Increase or Decrease according to the base parameters

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

Table : Pitch and Rate Variations

Rendering Equations using 3D Audio:

Different parts of an expression are rendered such that the listener perceives the audio from different directions.

Used the Head Related Transfer Function technique to render the desired segments of the equations in 3D.

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

Table: HRTF Angle Variations

Algazi, V. Ralph, Carlos Avendano, and Richard O. Duda. "Elevation localization and head-related transfer function analysis at low frequencies." *The Journal of the Acoustical Society of America* 109.3 (2001): 1110-1122.

Rendering Equations using Pauses and Special Sounds:

Use of Pitch and Rate Variations

Increase or Decrease according to the base parameters

Use Sounds to alert the user for demarcations.

Use pauses to separate subexpressions and certain demarcations.

Results

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

Comprehensive Evaluation

Picked the best performing technique, technique 3.

Generated a set of equations.

Asked visually impaired and sighted participants to solve these equations.

Used equations with small computable numbers.

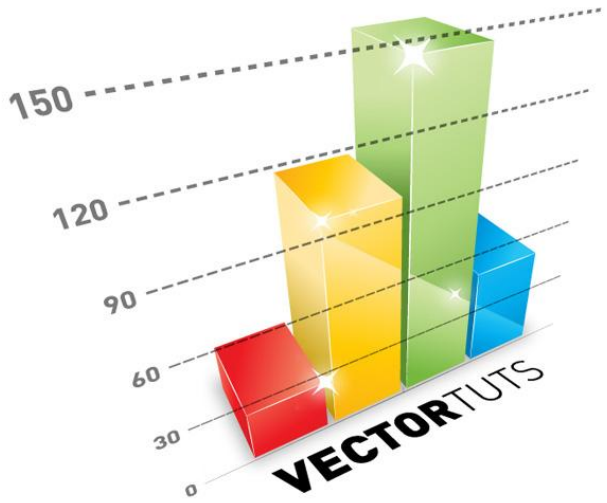
Allowed participants to use calculators and abacus
<insert screen shot>

Audio Rendering of Statistical Data

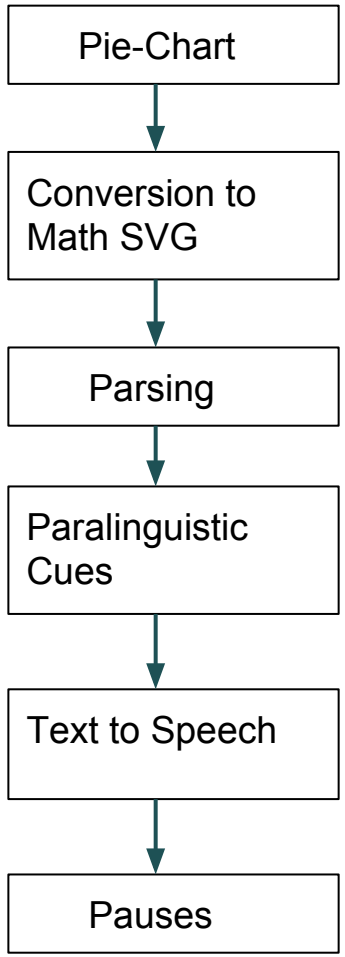
Statistical information — pictorially presented

Screen readers are incapable of speaking

Speaking with no tones would make it difficult to remember



System Design - Statistical Data Rendering



Experiment

Performed an experiment similar to the comprehension test.

Rendered pie charts with no cues and with pitch variations.

Person	Proposed Technique	Standard TTS (Traditional TTS)
Normal	96	73
Visually Challenged	95 .7	29.7

Conclusion

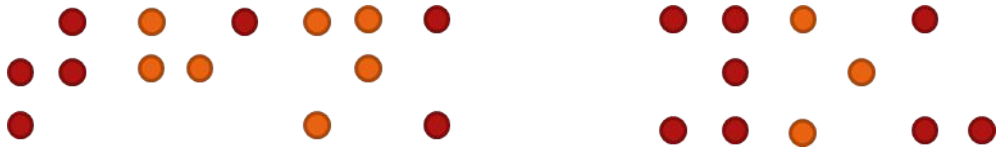
- Use of paralinguistic cues can help in effectively rendering equations in audio.
- A combination of the mentioned techniques can be more effective.
- Can be used in virtual assistants.
- Can be very beneficial for the Print Disabled.

Publications and Presentations

Accepted for publication in the International Conference on Natural Language Processing.

Accepted for presentation at the Annual International Technology and Persons with Disabilities Conference, San Diego, USA.

Thank you!



Questions?

We thank Microsoft Research for the travel grant and L.V. Prasad Eye Institute for providing participants for our research.