EAR-Math: Evaluation of Audio Rendered Mathematics

Hernisa Kacorri[1,2], Paraskevi Riga[1], and Georgios Kouroupetroglou[1]

[1] National and Kapodistrian University of Athens
Department of Informatics and Telecommunications,
Panepistimiopolis, Ilisia, 15784 Athens, Greece
[2] City University of New York (CUNY)
Doctoral Program in Computer Science, The Graduate Center,
365 Fifth Ave, New York, NY 10016 USA
{c.katsori,p.riga,koupe}@di.uoa.gr,
hkacorri@gc.cuny.edu

Abstract. Audio rendering of mathematical expressions has accessibility benefits for people with visual impairment. Seeking a systematic way to measure participants' perception of the rendered formulae with audio cues, we investigate the design of performance metrics to capture the distance between reference and perceived math expressions. We propose EAR-Math, a methodological approach for user-based evaluation of math rendering against a baseline. EAR-Math measures systems' performance using three fine-grained error rates based on the structural elements, arithmetic operators, numbers and identifiers in a formula. The proposed methodology and metrics were successfully applied in a pilot study, where 5 sighted and 2 blind participants evaluated 39 stimuli rendered by MathPlayer in Greek. In the obtained results, we observed that structural elements had the highest mean and variance of errors, which improved from 18% in the first attempt to 10% and 7% in two following attempts.

Keywords: mathematics, audio rendering, visually impaired, blind, evaluation, user study.

1 Introduction

The World Health Organization estimates a total of 285 million people with visual impairment of which 39 million are blind [1]. One considerable accessibility barrier is the comprehension of mathematical concepts and formulae through audio or haptic modality, which require significant additional cognitive processing [2]. Braille, linear in nature, faces difficulties in keeping up with complex mathematical expressions represented in two dimensions. Moreover, as 6-dot braille is limited to 64 characters, it results in complex notations to represent the vast number of mathematical structures and symbols. Usually this is performed through escape sequences, which map more than one meaning to the same braille character depending on the context.

Audio, a popular output medium for the visually impaired, is one approach favored by researchers to create an accessible platform for mathematics [3]. Similar to braille, audio rendering of a mathematical formula, either by a trained reader following a

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2014, Part II, LNCS 8514, pp. 111–120, 2014.

spoken structure (e.g. [4-5]) or by synthetic speech driven by a rule-based system such as [6-10], [24], communicates the expression using lexical and prosodic cues in a linear way. Researchers have investigated the possibility of alternating between multiple audio views such as summarization and detailed description [7], user customization of the rendering rules [7-8][10], and the use of non-speech audio cues or spatial audio to indicate structural delimiters within the formula [11-13].

Blind students may avoid adopting approaches for accessing mathematics that require the use of a language or representation of mathematics which differs from the one used by their teachers or peers. An adaptation of lexical cues, which indicate the structural information in the student's native language, is often required.

In the past few years, the Speech and Accessibility Lab, University of Athens, has focused on accessible mathematics [14-15] and the support of the Greek language for MathPlayer [8]. To construct the audio rendering rules we typically consult blind students during the initial design and the development phase. The challenges during these phases are: achieving a delicate balance between resolving ambiguity and reducing verbosity, and fine-tuning other non-lexical parameters such as pauses, pitch, and volume. When selecting between differently parameterized rendering styles, we are interested in quantitative results obtained from the users to measure the performance of our system. Despite advances in the audio rendering of mathematics, there are no widely accepted methodologies or benchmarks.

In this paper, we propose EAR-Math, a methodological approach along with a set of metrics, designed to evaluate and compare audio rendering systems for math expressions. Furthermore, we present results from an application of the proposed methodology in a pilot study evaluating the Greek audio rendering rules of MathPlayer.

2 Related Work

Relatively few researchers have evaluated the performance of math audio-access approaches. We reviewed prior studies evaluating math audio rendering systems based on perception by visually impaired and sighted participants. We present in Table 1 the relative methodologies based on the groups of participants recruited, selected stimuli, and the type of feedback obtained from the participants.

MathTalk [9] incorporates a set of rules to insert prosodic cues into spoken expressions. In a user study, participants wrote their recall of the formulae once the rendering was over using either question marks or ellipses to denote any missing objects. Responses were graded for comprehension of the structure and retention of content. A correct answer required over 75% of an expression's content and the major structural features.

TechRead's [16] used prosody to indicate nested structure. Sighted participants were asked to choose among 4 answers matching the audio stimuli. Of 16 rendered math expressions, 3 were in training while the remaining were played in 3 speech modes.

Gellenbeck et al. [17] conducted a study to assess whether insertion of pauses inside spoken mathematical expressions reduces ambiguity between similar algebraic expressions. Participants heard stimuli and rated side-by-side formulae, using a 0-10 Likert-scale, on how well they thought the visual expression matched the audio. Before the study, participants gained familiarity through four ungraded example tasks.

Murphy et al. [13] used a mixture of non-speech auditory cues, modified speech, and binaural spatialization for disambiguation. To evaluate initial cues, the authors conducted a four-part accessible online survey with no training phase. The first part tested word recognition and lexical language of spoken mathematics, the second measured understanding of spoken equations with non-speech sound elements, the third assessed intuitiveness of spatial attributes, while the fourth was purely qualitative. The listeners chose among 3 alternatives rendered both graphically and acoustically.

I-Math [18] was evaluated by participants with varying visual ability. Audio stimuli were graded on speech quality, comprehension effort, and transcription. Transcriptions were compared to original textual expressions and correct, missing, and incorrect words were counted. The number of correct words and positions were evaluated using precision, recall, and F-score.

Table 1. User studies evaluating approaches for audio rendering of mathematics

	MathTalk	TechRead	Gellenbeck et al [10]	Murphy et al [14]	I-Math
# Participants	24 (sighted)	20 (sighted)	16 (sighted)	35 (sighted) 21 (blind)	35 (sighted) 6 (blind), 4 (v.i.)
Participants' Age	university students	university students	university students ~22,7	~38	students and teachers
Knowledge/ Use of Math	daily-infrequent use (qualification > 'O' Level)	good knowledge of math constructs in stimuli	N/A	10 years since last exam (mean)	N/A
# Stimuli	2 matched sets of 12 and 3 samples	16 (3 samples)	30 min. session	13 (6 partial)	35
Stimuli Categories	fractions, parenthesised subexpressions, superscript	simple fractions, radicals, sum, limits, integrals, trigonometry	selected from the general area of algebra	N/A	fractions, vectors, superscripts, log, radicals, lim, sum., trigonometry, integrals
Training	explanation and synthetic speech familiarization	explanation and examples	description and examples	-	N/A
Language	English	English	English	N/A	Thai (tonal)
Participants' Answers	written math expression	4 multiple-choice solutions	rating of 2 alternatives	3 multiple-choice solutions	transcribed text
Baseline/ Comparison	lexical vs. prosodic cues, lexical vs. neither lexical nor prosodic cues	natural vs. enhanced vs. unenhanced speech	pauses vs. no pauses	sighted vs. blind participants	-
Quantitative	accuracy	accuracy	average rating	accuracy intuitiveness, confusion	intelligibility, speech quality, understanding effort
Qualitative	-	-	-	Alternatives, comments	usefulness, ease of use

3 Evaluation Methodology

Evaluation of Audio Rendered Math (EAR-Math) is proposed as an experimental methodology for user-based performance evaluation of mathematical expressions described by a rule-based system with audio output. The accuracy scores achieved over the stimuli in a user study depend not only on the system output, but also on the difficulty of the mathematical expressions chosen and on the group of participants. Thus, EAR includes a baseline in the studies to make the results more meaningful. This baseline may typically be a trained reader following a spoken structure to read mathematical expressions aloud. The baseline could be a previous version of the system for researchers who wish to evaluate improvements within their system. Similarly, EAR-Math allows for direct comparison between two or more alternative systems.

3.1 Experiment Design

Since it is not feasible to exhaustively examine the mathematical expressions likely to be rendered by a system, the design of a smaller representative subset is required. Stimuli should cover an extended number of mathematical structures (e.g. fractions, integrals, roots, subscripts, and arrays). They should be engineered to reveal potential ambiguity in the system output and have a realistic length. The number of stimuli should also allow incorporation in an experimental session lasting less than 2 hours. We considered the publicly available set of formulae for ASTeR demonstration [19] as a potential candidate.

The EAR-Math user study consists of three phases. As a preliminary phase, participants are introduced to the system's rules with sample examples to gain familiarity with the generated output. Symbols naming, a list that can extend over 2.000 Unicode characters, is mentioned only for a small number of characters, which are included in the stimuli and are assigned a non-familiar name.

In phase 2, participants listen to the description of the math formula (rendered either by the system being evaluated or the baseline) and are prompted in parallel to keep notes of the formula. Notes may be taken on paper for sighted participants in Braille for blind participants, in LaTeX, or in any other format and means the participant is familiar with. Stimuli are played upon request two more times and the participants are allowed to correct their initial guess.

In phase 3 of the study, participants hear two sequential versions of the same mathematical expression rendered by the system and baseline (with alternating sequence). The participants can hear any of the outputs as many times as they wish. They are then asked to respond on a 1-to-10 Likert-scale question about how sure they are that the expressions are identical. For each of the two versions, participants were asked two 1-to-10 Likert-scale questions about the understandability and naturalness of the rendered formula.

3.2 Performance Metrics

EAR-Math proposes a performance metric to measure the distance between the intended math expression and the actual one delivered by the system as perceived by sighted or blind users. Our metric is tailored to account for both content and structure. Specifically, the performance of a math audio rendering system is defined as a triplet of fine-grained error rates on the structural elements, arithmetic operators, and numbers and identifiers of the math formula:

$$Structure\ Error\ Rate\ (SER) = \frac{\#ins + \#del + \#sub}{\#structuralElem}$$

$$Operator\ Error\ Rate\ (OER) = \frac{\#ins + \#del + \#sub}{\#operators}$$

$$Identifier/Number\ Error\ Rate\ (INER) = \frac{\#ins + \#del + \#sub}{\#identifiers + \#numericalValues}$$

The proposed metrics are derived by first aligning the perceived mathematical expression with the original expression to be rendered by the system. Alignment is performed between the syntax trees of both expressions. Figure 1 illustrates an example demonstrating the process. Our metrics can then be computed as error rates; number of insertions, deletions, and substitutions in the perceived expression compared to the reference over the total number of elements in the reference expression for each of the three categories of elements. Structure Error Rate (SER) involves structural components of a math formula such as fractions, roots, and arrays. Operator Error Rate (OER) is focused on mathematical operators e.g. plus, minus, and times. And last, Identifier and Number Error Rate (INER) represent the number of errors for identifiers and numerical values within the expression. For example, given a reference and the perceived expression pair, we get the syntax trees and alignment of Fig. 1.

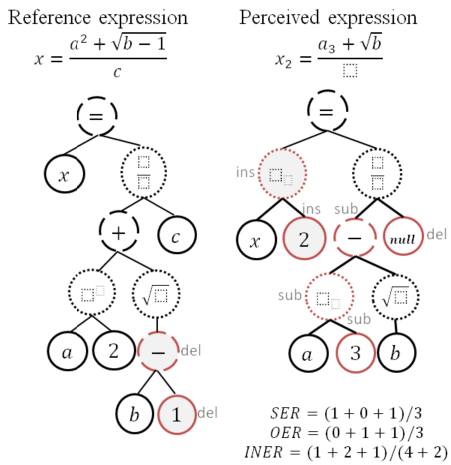

Fig. 1. An example of alignment and SER, OER, and INER calculation

Word Error Rate (WER), commonly used in the field of speech recognition, was adopted for the design of the proposed performance metrics. To group the components that comprise a mathematical formula, we were inspired by the token and layout elements in Presentation MathML, such as <mo> for operators, <mi> for identifiers, and <mn> for numbers. This grouping simplified incorporating user's feedback towards the improvement of our Greek audio rendering rules for MathML expressions in MathPlayer.

4 Pilot User Study

To assess feasibility, time, and statistical variability required for evaluating MathPlayer's Greek audio rendering rules, we conducted a pilot user study. This helped to test, adjust and obtain valuable insights in the design of EAR-Math and particularly its metrics and is shown as an applied example of the proposed methodology.

For the purposes of this study, the system was not evaluated against a baseline. Therefore, only the first two phases of EAR-Math were employed. Stimuli were based on the set of formulae for ASTeR demonstration [19] rendered through the Dimitris of Acapela Text-to-Speech [20] driven by MathPlayer with lexical and prosodic cues. The voice, volume, and speed for all stimuli were identical. Mathematical expressions were prerecorded, thus there was no navigation and no control over the speech velocity.

Of the 7 participants recruited for the study: 2 were congenitally blind and 5 were sighted. Participant attended a Greek university whose entrance required a high level of mathematics according to the Greek educational system (6 were majoring in computer science and 1 in law). Thus, all participants had been exposed to more complex mathematical expressions than the stimuli. There were 5 men and 2 women of ages 20-34 (average age 25.9). While using a sighted comparison group wouldn't be either fair or appropriate, we included people both with and without visual impairment. The rationale behind this decision is that audio rendering of mathematics has universal benefits (e.g. to temporary print disabled people). In addition, recruitment of a homogeneous group of people with visual impairment and similar performance in mathematic assessments from a geographically dispersed population is difficult. These challenges may set user studies investigating multiple improvements within the same system with adequate power of visually impaired participants impossible.

During the study, participants would listen to mathematical expressions and write down the perceived formulae. They were allowed to make changes to their initial guess two more times. We collected all three perceived versions for each of the expressions. Sighted participants wrote down their answers using a pen-and-paper experimental packet while blind participants used formats and technology that were most familiar to them. One participant used Nemeth [21-22]. The other used a variant math notation he had developed and used throughout the years. After the experimental session, blind participants would read their notes and describe their answers to a sighted member of the team who would then visualize the expression in two-dimensions.

Short mathematical expressions described in the design of Greek audio rendering rules for math [23] served as sample expressions to familiarize the participants with the experiment and the output of the system.

5 Results

In our study, MathPlayer's Greek audio rendering of math is evaluated with a set of 49 stimuli with 190 structural elements, 113 operators, and 295 identifiers or numbers. While not compared against a baseline, this evaluation demonstrates the EAR-Math performance metrics within the proposed experimental setup.

In our calculations, all reference expressions from the stimuli and perceived expressions were drawn as syntax trees. A computer science PhD student 'manually' aligned the perceived formulae syntax tree to the reference tree and recorded the errors. While identifying the operator, identifier, and number errors were a straightforward process, structural elements errors were more challenging, especially when the correct structure was improperly positioned in the tree. An improper position of the structural element was calculated as a delete and re-insert.

The number of mathematical elements for each stimuli varied by context: 0-13 structural (mean 5), 1-10 operators (mean 2.97), and 3-18 identifiers and numbers (mean 7.76). Therefore the error rates were calculated for the aggregated elements among all stimuli, as shown in Table 2.

Table 2. Overall error rates in the stimuli set

	SER	OER	INER
1st Attempt	18%	12%	11%
2nd Attempt	10%	6%	4%
3rd Attempt	7%	4%	2%

Figure 2 shows the distribution of the SER, OER, and INER in the perceived mathematical expressions from the participants as boxplots with whiskers at the 1.5 IQR (inter-quartile range). To aid the comparison, mean values, illustrated with a star, are added as labels at the top of each plot. For the first attempt, we observe that both the structural error rate (SER) and the error rate for identifiers and numbers (INER) have higher variance than the operators' error rate (OER). We speculate this is due to the inherent dependence of the INER on SER. For example, when participants do not understand a structure and omit it, they often omit the identifiers and numbers within the structure. We also observe that participants tend to improve their errors the second and the third time they hear the mathematical expression. This suggests that the audio rendering might have been accurate, but other factors (such as audio memory and familiarity with the system) may have an effect on the results and should be taken into account when designing the experiment.

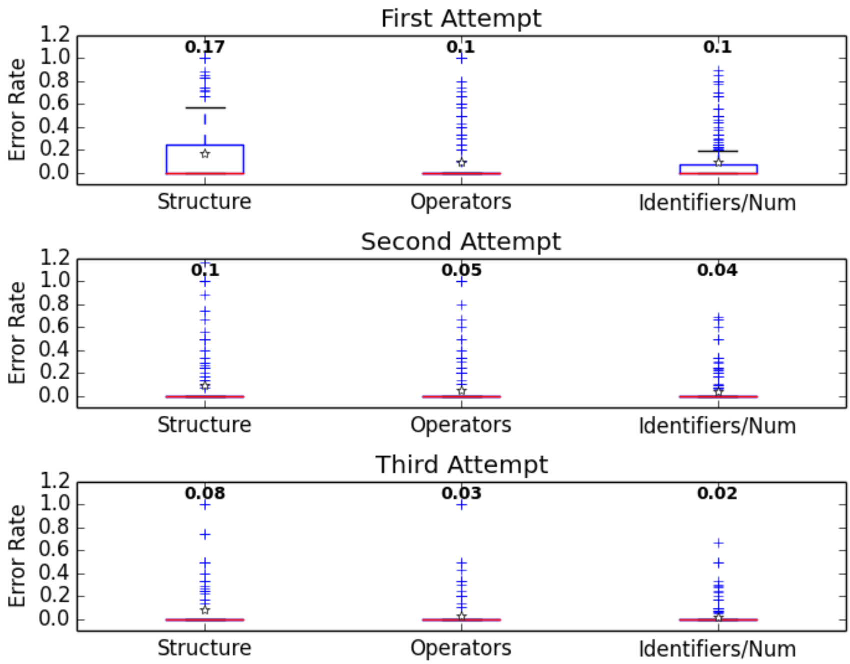

Fig. 2. Error rate distribution in each of the three attempts

(12) $$1+\cfrac{\chi}{1+\cfrac{\chi}{1+\cfrac{\chi}{1+\cfrac{\chi}{1+\cfrac{\chi}{\ddots}}}}}$$ (39) $$e^{\left(e^{e^{\chi}}+e^{\chi}+\chi\right)}$$

Fig. 3. Stimuli with the Highest Structure Error Rate in the User Study

We analyzed the data and found that the two most misperceived stimuli at a structural level (12 and 39) had SER improving in each of the attempts from 93% to 69% and 38% for expression (12) and from 14% to 3% and 3% for expression (39).

6 Discussion and Future Work

This paper has described EAR-Math, a methodological approach for the evaluation of math audio rendering systems with users and a set of performance metrics that capture the distance between the reference and the perceived math expressions. In particular, EAR-Math metrics are defined as error rates (Structure Error Rate, Operator Error Rate, and Identifier and Number Error Rate) calculated by comparing the syntax trees of the reference and perceived mathematical expressions. Our methodology and metrics are demonstrated in a pilot user study evaluating the Greek audio rendering rules of MathPlayer with 7 participants and 39 stimuli. Obtained results, show that Structure Error Rate had higher mean and variance than the other two metrics, which was improved the second and third time participants heard the stimuli.

This research has two key contributions. First, it provides guidance for researchers conducting user-based evaluation studies to measure the performance of math audio rendering systems against a baseline, compare alternative systems, or iteratively evaluate improvements/styles. Second, it provides results from a pilot study to assess feasibility, time, and statistical variability required for a case study.

In future work we want to explore alternative ways to automate the calculation of the proposed metrics without human intervention. This would allow for more robust results with less human errors during the data processing steps.

Acknowledgments. We would like to thank Neil Soffer from DesignScience for his support during the development of the Greek rules in MathPlayer. This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) under the Research Funding Project: “THALIS-University of Macedonia-KAIKOS: Audio and Tactile Access to Knowledge for Individuals with Visual Impairments”, MIS 380442.

References

1. World Health Organization. Fact Sheet 282: Visual Impairment and Blindness (2013), `http://www.who.int/mediacentre/factsheets/fs282/en/index.html`
2. Dick, T., Kubiak, E.: Issues and aids for teaching mathematics to the blind. The Mathematics Teacher 90(5), 344–349 (1997)
3. Freitas, D., Kouroupetroglou, G.: Speech Technologies for Blind and Low Vision Persons. Technology and Disability 20, 135–156 (2008)
4. Chang, L.A., White, C.M., Abrahamson, L.: Handbook for spoken mathematics. Lawrence Livermore National Laboratory (1983)
5. Nemeth, A.: Tips for Reading Math Out loud to Blind Students, `http://www.rit.edu/easi/easisem/talkmath.htm`
6. Edwards, A.D., McCartney, H., Fogarolo, F.: Lambda: a multimodal approach to making mathematics accessible to blind students. In: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 48–54 (2006)
7. Raman, T.V.: ASTER-Towards modality-independent electronic documents. In: Electronic Multimedia Publishing, pp. 53-63, Springer US (1998)
8. Soiffer, N.: A flexible design for accessible spoken math. In: Stephanidis, C. (ed.) UAHCI 2009, Part III. LNCS, vol. 5616, pp. 130–139. Springer, Heidelberg (2009)
9. Stevens, R.D.: Principles for the design of auditory interfaces to present complex information to blind people. Doctoral dissertation, University of York (1996)
10. Yamaguchi, K., Komada, T., Kawane, F., Suzuki, M.: New features in math accessibility with infty software. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 892–899. Springer, Heidelberg (2008)

11. Bates, E., Fitzpatrick, D.: Spoken mathematics using prosody, earcons and spearcons. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 407–414. Springer, Heidelberg (2010)
12. Brewster, S.A.: Using non-speech sound to overcome information overload. Displays 17(3), 179–189 (1997)
13. Murphy, E., Bates, E., Fitzpatrick, D.: Designing auditory cues to enhance spoken mathematics for visually impaired users. In: Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 75–82 (2010)
14. Kouroupetroglou, G., Kacorri, H.: Deriving Accessible Science Books for the Blind Students of Physics. Proceedings of the American Institute of Physics 1203, 1308–1313 (2010)
15. Tsonos, D., Kacorri, H., Kouroupetroglou, G.: A design-for-all approach towards multimodal accessibility of mathematics. Assistive Technology Research Series 25, 393–397 (2009)
16. Fitzpatrick, D.: Towards Accessible Technical Documents: Production of Speech and Braille Output from Formatted Documents. Doctoral dissertation, Dublin City University (1999)
17. Gellenbeck, E., Stefik, A.: Evaluating prosodic cues as a means to disambiguate algebraic expressions: an empirical study. In: Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 139–146 (2009)
18. Wongkia, W., Naruedomkul, K., Cercone, N.: I-Math: an Intelligent Accessible Mathematics system for People with Visual Impairment. Computational Approaches to Assistive Technologies for People with Disabilities 253, 83–108 (2013)
19. Raman, T.V.: Mathematics for computer generated spoken documents - ASTeR Demonstration, `http://www.cs.cornell.edu/home/raman/aster/demo.html`
20. Acapela text-to-speech, `http://www.acapela-group.com/`
21. Kouroupetroglou, G., Florias, E.: Greek Braille Scientific Notation – Application in Information Systems for the Blind. Education and Rehabilitation Center for the Blind, Athens (2003) ISBN 960-87918-0-4
22. Nemeth, A.: The Nemeth Braille code for mathematics and science notation: 1972 revision. Produced in braille for the Library of Congress, National Library Service for the Blind and Physically Handicapped by the American Printing House for the Blind (1972)
23. Kacorri, H.: Audio Rendering Rules for Mathematical MathML Expressions in Greek, Diploma Thesis, University of Athens (2006)
24. Ferreira, H., Freitas, D.: Audio rendering of mathematical formulae using MathML and AudioMath. In: Stary, C., Stephanidis, C. (eds.) UI4ALL 2004. LNCS, vol. 3196, pp. 391–399. Springer, Heidelberg (2004)