Navigable, Customizable TTS for Algebra

Lois Frankel, Beth Brownstein,

Educational Testing Service

lfrankel@ets.org bbrownstein@ets.org

Neil Soiffer

Design Science, Inc.

neils@dessci.com

Abstract

To improve accessibility of math expressions for students with visual impairments, our project employs MathML and the products MathType and MathPlayer to provide an improved style of synthetic speech called ClearSpeak. This paper describes the project's background, the essential features of ClearSpeak, and results of the first feedback study with students that show for the tested notations, ClearSpeak was superior to the two pre-existing speech styles in MathPlayer (SimpleSpeak and MathSpeak).

Keywords

Math, MathML, Audio, Text-to-speech, Blind, Algebra

Introduction

Our IES funded development grant R324A100355, *Expanding Audio Access to Mathematics Expressions by Students with Visual Impairments via MathML*, addresses the math accessibility gap for students with visual impairments. We are developing improvements in text-to-speech assistive technology to make math expressions encountered in secondary school Algebra, as well as instructional material and assessments containing such expressions, accessible via synthetic speech and interactive navigation. This type of audio access can be used along with on-screen visual access to the document or with printed or brailled versions of the document, or it can provide previously-unavailable access for those who are using neither braille nor vision.

The IES development grant enhances the widely used MathType math authoring tool and the MathPlayer visual and audio rendering plug-in for Internet Explorer. The project has four main components:

1. Create a semantically rich speech style (ClearSpeak) for secondary-school algebra-level math expressions that uses language that is familiar to students to speak the expressions
2. Enhance authoring tools within MathType to allow customizable speech for math expressions
3. Enable speech of mathematical expressions from within Microsoft Word
4. Develop interactive within-expression navigation

The design of the ClearSpeak speech style and the way navigation works will be guided by several feedback studies and the final result will be the subject of a pilot study. This paper discusses the results of the first feedback study comparing ClearSpeak to the two pre-existing speech styles, SimpleSpeak and MathSpeak.

Background

Currently, math expressions are notoriously difficult to access using any form of audio (Bouck and Meyer 44). Among the reasons for this are their often-complex structure and lack of consistency in audio rendering. Moreover, math's non-linear two dimensional presentation in print provides important information about an expression’s structure that is more difficult to convey using linearly-presented speech. For example, $\frac{x}{3}+\frac{3}{x}$ shows the cancellation possibilities more clearly than do $x/3 + 3/x$ or the words “x over three, plus three over x.” Significant problems for audio math, including middle and high school algebra, include detecting boundaries (e.g., where fractions, exponents, roots, etc., begin or end) and tracking expressions involving nested parentheses.

Although teachers frequently develop electronic text-only documents for student use, mathematical material is usually inaccessible because math expressions and figures are typically represented by images. In addition, although over the last few years MathML has begun replacing images to represent math expressions in some documents, screen readers don’t currently know how to read MathML directly. They depend upon either pre-existing, custom written translations of math expressions to words (“alt text”) or Design Science's MathPlayer plug-in for Internet Explorer.

The first part of the project is aimed at developing a new, consistent, and familiar semantically rich speech style for MathPlayer that we call ClearSpeak. The development of ClearSpeak is informed by the results of earlier projects aimed at developing rules for human- or computer-spoken math. In the 1950s, Nemeth began developing a way of uniquely linearizing a mathematical expression for speech that mapped one-to-one to the Nemeth Braille Code he also developed. He initially developed this speech style for working with aides/readers so that he could write braille immediately as the reader spoke each part of the expression. Many years later, Nemeth worked with gh, LLC to formalize this into the MathSpeak™ speech style.

Another early set of human-speech rules was developed by Lawrence Chang. A practicing blind scientist, he developed the speech guidelines (“Larry’s SpeakEasy”) for his aides/readers. In the spirit of Chang’s rules, but enhanced for the requirements of high-stakes testing, ETS test developers developed a mathematical speech style for scripting tests and test preparation materials for recordings and for use with live test readers. This speech style uses pauses plus familiar language and syntax so that examinees are not forced to learn a new “language” at testing time. In the 1990s, development of computer generated math speech began with ASTER (Raman). ASTER uses prosodic cues (i.e., pauses plus changes to pitch, rate, volume, and voice) among other parameters to speak expressions in a more concise manner than MathSpeak. Design Science developed its MathPlayer plugin to speak with the user's choice of its simplified ("SimpleSpeak") style or a version of MathSpeak. It uses some prosodic speech cues, but screen reader support for these has been very limited to date. Others have also worked on computerized speech (MathTalk (Stevens, Edwards, and Harling 47-92), MathGenie (Karshmer Bledsoe, and Stanley 614-19) and auditory cues (Murphy, Bates, and Fitzpatrick 75-82)), but none of those systems were widely used.

Speech Styles

A speech style governs computer-generation of speech. Design Science’s MathPlayer interprets MathML and, since it stores its rules and vocabulary externally as plain text files, it can work with any speech style that can be expressed using its rule-based pattern matching language. Prior to the current project, MathPlayer supported two speech styles:

- SimpleSpeak – common, simple expressions such as $\sqrt{x}$ are spoken simply; begin/end bracketing is used for disambiguating the start and end of more complicated expressions
- MathSpeak – maps 1-1 to Nemeth code with exceptions for some common expressions such as “x squared”

ClearSpeak is designed to improve on the existing speech styles by speaking math in a way that will be familiar to most students--the way teachers and other content providers speak math to them--but with adjustments to that customary speech to make sure the spoken math is unambiguous. Unfortunately, adding words to disambiguate expressions tends to increase the verbal and memory load and so tends to make spoken expressions harder to understand than the print versions of those expressions. ClearSpeak seeks to minimize this tendency by extensive use of pauses (which do not increase verbal load), by using language that integrates well with the expression, and by “translating” print conventions into their mathematical equivalents if a sighted reader of the print version would do so instantaneously (e.g., reading an exponent as “to the power” rather than as “superscript”)--thus removing an obstacle to understanding not encountered by sighted readers. For example, ClearSpeak refers by default to “power” rather than to “super,” “superscript” or “baseline.” Finally, since in different contexts or for different audiences it is desirable to have the same math structure speak differently, ClearSpeak provides three mechanisms for speaking them in a way that is appropriate for the intended purpose: default speech rules, author preferences, and exact speech.

Rules

Speech rules are the standard or default way for speaking math structures. The speech is based on information that can be parsed from the MathML representation of the expression/object. MathPlayer analyzes expressions based on types of structures present in the expression (fractions, exponents, parenthetical expressions, radicals, etc.), the level of complexity, the subject area, and the user's stated level of expertise. Based on that analysis, rules are invoked for speaking the expression, and the resulting speech information (including prosodic information) is supplied to the speech engine. In addition, rules include specifications of how various symbols (raised dot, set membership, etc.) are to be spoken. For expressions currently lacking a ClearSpeak specific implementation, the SimpleSpeak rules are applied.

Example: Simple fractions, defined as those with a single integer or variable in the numerator and in the denominator, fall into two cases:

1. Speak as a common fraction (e.g., "three quarters") if the numerator is a positive integer less than or equal to 19 and the denominator is a positive integer less than or equal to 10. E.g., $\frac{5}{7}$ is spoken as "five sevenths".
2. Otherwise, speak as '[numerator] over [denominator]'. E.g., $\frac{6}{21}$ is spoken as "six over twenty one," and $\frac{x}{8}$ as "x over eight."

Preferences

Sometimes, identical expressions may need to be spoken differently because of contextual differences, to resolve ambiguities, or for instructional or assessment purposes, so rules can and will not generate the appropriate speech in all cases. We provide a system of pre-defined preferences that authors of math documents can set to produce commonly encountered alternatives to the speech generated by the rules.

Example: The rules would speak $\frac{12}{15}$ as “twelve over fifteen,” but an author can invoke a preference to force it to speak as “twelve fifteenths,” as “the fraction twelve over fifteen (pause) end fraction,” or as “the fraction with numerator twelve and denominator fifteen.”

Exact Speech

For cases in which the developed rules and preferences do not produce the desired speech, authors can insert exact wording, plus additional pauses or changes to pitch, volume, or speech rate, if desired. Exact speech can be applied to the entire expressions or to parts of it.

Example: $\frac{x^2 + y}{3x} + 4$

The rules would speak this as “the fraction with numerator x squared plus y and denominator 3x (pause) plus 4.” If the teacher wants to lengthen the pause before “plus 4” to make it even clearer that the 4 is not part of the denominator of the fraction, exact speech could be used to lengthen the pause or else to insert “the integer” before “4”.

Initial ClearSpeak Evaluation

Setup

Our first feedback study, conducted in March 2012, compared the initial implementation of ClearSpeak with the two styles already built into MathPlayer (MathSpeak and SimpleSpeak). The 15 high school students who completed the study were blind (7) or had low vision (8), and were taking Algebra 1 or a subsequent math class. A questionnaire gathered background information such as how they accessed math (screen reader, human reader, etc.) and their proficiency with various mathematical concepts relevant to the study (fractions, exponents, parentheses). The study itself consisted of instructions and samples, plus seven math expressions. A slightly different version of each expression was created in each of the three speech styles (ClearSpeak, MathSpeak, and SimpleSpeak), and the expressions in each speech style were combined into a section, so that each student received one section for each speech style.

The math expressions were provided as pre-recorded text-to-speech (TTS) audio using Microsoft Anna. After each expression a set of math and behavioral/affective questions was developed to assess how clear, familiar, and helpful the speech was in understanding the math expression and whether the student was able to understand what the expression was, *not* whether the student was able to simplify the expression or solve a problem based on the expression. Additional feedback was requested asking for suggestions to improve the speech and how much the student liked or disliked each style. A sample expression and math question are below:

Sample Expression: $16x^9 + 12x^{13} + 5x^7 + 1$

Math Question for the Sample Expression: “What is the biggest exponent in the math statement?”

Quantitative Results

Table 1 summarizes average scores on four measures: students’ comfort with the speech (favorability), confidence that they understood what they heard (confidence) and that they answered the math question correctly (math confidence); and actual correct answers to math questions (correct math response). The averages shown are based on each student’s response to several questions related to each math statement.

Table 1: Results of speech style tests

	ClearSpeak			SimpleSpeak			MathSpeak		
Measure (range)	All	Blind	Low Vision	All	Blind	Low Vision	All	Blind	Low Vision
Favorability (0 to 6)	4.75	4.92	4.61	3.25	3.96	2.62	2.96	3.65	2.36
Confidence (0 to 3)	2.64	2.53	2.73	1.87	2.04	1.72	1.64	1.85	1.46
Math Confidence (0 to 5)	3.92	3.80	4.03	2.68	3.45	2.00	1.88	2.86	0.95
Correct Math Responses (0 to 1)	0.89	0.91	0.88	0.70	0.78	0.64	0.56	0.67	0.46

As Table 1 indicates, ClearSpeak outperformed the other styles on all measures. One-way ANOVA found significant differences among speech styles on each of the measures: favorability [$F(2, 312)=32.21$, $p<.01$], confidence [$F(2, 312)=20.51$, $p<.01$]; math confidence [$F(2,177)=19.36$, $p<.01$]; and correct math response [$F(2, 44)=9.17$, $p<.01$]. Post hoc Bonferroni comparisons indicated that ClearSpeak performed significantly higher ($p<.01$) than both MathSpeak and SimpleSpeak on all four measures. Additionally, it was found that achieving a higher number of correct answers is significantly associated ($p<.05$) with higher favorability, confidence, and math confidence scores.

After the audio was recorded and testing began, a bug that caused some extra words to be used in some MathSpeak expressions was discovered. This likely contributed to students' difficulty with MathSpeak: several students mentioned the faulty phrasing as being confusing. However, those complaints were also consistent with objections to several other aspects of MathSpeak, so it is unlikely that the bug had a significant effect on the results.

Qualitative Results

Although we have seen an overall preference for and better performance with ClearSpeak, students had varying opinions on pauses, speed, terminology, parentheses, and the verbalization of "times" in expressions where the "times" is implied by parentheses (e.g.,

$2(x+3)$). They seemed more united on a preference for the language of “powers” rather than that of “super,” “superscript,” or “baseline.” Eight students explicitly objected to “super” or “superscript”, and four objected to “baseline.” Two mentioned “to the” and three mentioned “power” positively. Some did not know what these terms meant. Many did not care for the TTS voice used to record the speech: Microsoft Anna. Anna is tuned to speak regular text and has prosodic and pronunciation difficulties with math. Future studies will make use of the Eloquence voices that are commonly used by Window-Eyes and JAWS users. These do not try to be as natural sounding and do not make as many assumptions about word flow as does Anna and hence do a better job speaking math expressions. Screen reader users tend to prefer these voices because they can be used at high speech rates and still be understood.

Conclusions and Future Work

The first study showed that for the tested notations, ClearSpeak was an improvement over SimpleSpeak and the MathPlayer implementation of MathSpeak. As a side benefit, based on the results of the study, several features of ClearSpeak such as the use of “power” instead of “super” were added to SimpleSpeak after the first study, and the MathSpeak bug was corrected.

Having concluded that we were on the right track with ClearSpeak, our second study will focus on prosody and speaking nested parentheses. Initial work has narrowed our prosody changes to pausing and rate changes. These will be compared against the speech that uses words to indicate the start and end of fractions, roots, or other structures. For expressions with nested parentheses, we are experimenting with rate changes, pausing, and using different words for open and close parentheses to help students determine and remember which close parenthesis matches which opening parenthesis. GW Micro, a project partner, is integrating this work into their Window-Eyes product; many other companies have expressed interest in integrating their products with this work when the development is mature.

Acknowledgements

We are grateful to Nan Kong and Eric Hansen for their work in applying statistical data analysis techniques and for their helpful suggestions regarding research design.

Works Cited

Bouck, Emily C., and Nancy K. Meyer. "eText, Mathematics, and Students with Visual Impairments: What Teachers Need to Know." *Teaching Exceptional Children* 45.2. (2012): 42-49. Print.

Chang, Lawrence A. *Handbook for Spoken Mathematics (Larry's SpeakEasy)*. Livermore: Lawrence Livermore National Laboratory, 1983. Print.

Karshmer, Arthur, Chris Bledsoe, and Paul Stanley. "The Architecture of a Comprehensive Equation Browser for the Print Impaired." *Proc 2005 IEEE Symposium on Visual Languages and Human-Centric Computing (Dallas, TX, September 2005)*. IEEE Computer Society Washington, DC: 319-20. Print.

Murphy, Emma, Enda Bates, and Dónal Fitzpatrick. "Designing Auditory Cues to Enhance Spoken Mathematics for Visually Impaired Users." *Proc 12th international ACM SIGACCESS conference on Computers and accessibility* (ASSETS '10): Orlando: ACM, 2010. 75-82. Print.

Raman, T. V. *Audio System for Technical Readings*. Ithaca: Cornell University, 1994. Print.

Stevens, Robert, Alistair Edwards, and Philip Harling. "Access to Mathematics for Visually Disabled Students through Multi-Modal Interaction." *Human-Computer Interaction (Special issue on Multimodal Interfaces)* 12 (1997): 47-92. Print.