

Mathtalk: The Design of an Interface for Reading Algebra Using Speech

Robert Stevens and Alistair Edwards

Department of Computer Science, University of York, York, UK YO1 5DD

Abstract. The problems that a visually disabled person has with reading, writing and manipulating standard algebra notation are characterized in terms of *speed*, *control* and *external memory*. The Mathtalk program has been developed to enable the listener to read algebra notation in a quick and active manner, that overcomes some of these problems. Prosody has been added to the synthetic voice output to resolve grouping ambiguities, decrease mental workload and improve memory for the presented notation. Browsing functions and the associated command language allow the listening reader to shift his or her attention to any part of an expression. To make most effective use of the *speed* and *control* afforded by Mathtalk, an *audio glance* is provided that should allow planning of the reading process. The development and evaluation of Mathtalk has led to the proposal of a set of design principles that should facilitate the production of other, similar user interfaces for the reading of structured information. Further work includes the development of Mathtalk in the TIDE Maths project. As well as reading, the problems of writing and manipulation need to be tackled.

1 Introduction

This paper describes the design and evaluation of the interface to Mathtalk. This is a PC based system that aims to enable blind people to gain a quick and active reading of standard algebra notation, using speech and non-speech audio. Algebra and other related notations are the most common form of expression in both mathematics and sciences, which are both subjects of great importance. The ability to use such notations in a quick and effective manner has the potential to enhance the employment and educational prospects of many visually disabled people. This work will be extended to cover the writing and manipulation of a wider set of Mathematics in both speech and Braille as part of the European Tide Maths project.

Standard algebra forms the core of most notations used in mathematical, scientific and technical disciplines. It is used both to manipulate and to communicate mathematical concepts. Printed algebra notation gives a persistent record of these ideas, thus making many mathematical tasks easier by affording an *external memory*. The external memory allows a sighted reader to *control* the information flow with speed and accuracy. This control makes the sighted reader the *active* partner in the reading process. As well as enhancing memory

for an individual, this external memory can support communication between mathematicians, as the spoken word may in other disciplines.

A blind user does not have the luxury of a piece of paper and must use either braille or recorded speech. A tape recorder is a poor interactive tool [1]. Using a tape recorder is a slow and inaccurate process. Mathtalk is an attempt to improve the usability of spoken algebra. Spoken mathematics is often ambiguous and symbol names are inconsistent. The speech signal is transient and thus great demands are made upon human memory. In addition, the listener is passive and frequently loses concentration, making the reading task more difficult. The listener is not truly a reader: to be so the listener must be the active party in the interaction and become what Simpson [13] calls the *listening reader*.

The Mathtalk program seeks to provide both a useful external memory and the means to control information flow to make the listening reader an active reader. The design, implementation and evaluation of the Mathtalk program has given rise to a set of design principles that can be used to enhance the reading of complex information by blind people. The following sections describe these design principles in more detail.

2 How to Speak Algebra notation

The first question that needs to be answered is ‘what information to present?’ Printed algebra notation presents only the syntactic elements of an expression. The syntax primarily presents the grouping within an expression. It is the reader, using his or her knowledge of mathematics, algebra and any other surrounding text, who parses and makes any mathematical interpretation. This is the first principle for the design for an aid to read algebra: The aid should only present the elements of the algebra notation, it is for the reader to make any mathematical interpretation.

2.1 Resolving Grouping Ambiguities

After deciding what information to present, a decision must be made on how to present that information. The spatial cues within a printed expression group items together and help to indicate how they should be parsed [9]. In print that the fact that $x^n + 1$ is parsed differently to x^{n+1} is obvious. The eyes can be used to move through an expression in such a way that these cues can be used to facilitate parsing. As a result of the transient speech form providing no external memory and control being poor, the utterance ‘x to the n plus 1’ is potentially ambiguous in that the scope of the superscript is not made explicit. The auditory display must be enhanced to avoid grouping ambiguities and facilitate parsing in a manner analogous to printed algebra.

Using Lexical Cues One method of preventing grouping ambiguity is to use special words or phrases (*lexical cues*) inserted into the utterance to delimit the

scope of syntactic groups. Chang [7] proposes set of rules for speaking mathematics, both consistently and unambiguously. Chang offers several levels of mathematical information in the types of lexical cue he uses. A Mathtalk option applies a subset of Chang's rules that are consistent with the minimal level of syntactic interpretation offered to the reader.

An example that covers many of the syntactic types used by Mathtalk is shown below:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (1)$$

This is spoken as:

x equals the fraction, numerator negative b plus or minus the square root of the quantity b squared minus four a c denominator two a.

The notion of *simple* and *complex* syntax can be introduced to guide the speaking of an algebra expression. A complex item is one in which more than one term is grouped together by an explicit parsing mark or spatial grouping in the printed form. For example, parenthesised sub-expressions are complex items. A complex fraction has more than one term in either numerator or denominator. Superscripts, like fractions, can be either simple or complex. It is to delimit complex items that lexical cues are used. This notion of simple and complex will arise throughout the discussion of the Mathtalk program.

Thus, the use of lexical cues can avoid any grouping ambiguity in a spoken algebra expression. Evaluation of this style of presentation revealed some problems. The addition of lexical cues makes an utterance much more verbose. The large amount of spoken material is difficult to parse, retain and integrate. The use of lexical cues often invoked the suffix effect [2], causing the content of complex items to be lost. Evaluation of mental workload [8] showed this style of presentation to be very demanding and frustrating [16].

Using Prosody in Spoken Algebra Synthetic speech is a much impoverished signal compared to human speech. The speech component known as prosody is often missing. Prosody can be thought of as the non-verbal information content of speech [10]. An example of the latter phenomenon may be seen in the following sentences: "The last time we met / Robert was horrible" and "The last time we met Robert / was horrible". The slash indicates where a prosodic cue, usually a pause, can be inserted to completely change the meaning of the utterance. As the grouping structure of an algebra expression is known, appropriate prosodic cues could be ascribed to the syntactic junctures to indicate grouping instead of lexical cues.

For the Mathtalk program we derived a relatively simple set of prosodic rules from the analysis of a series of spoken algebra expressions. These prosodic cues agreed with and enhanced those derived from earlier research [17, 11]. These rules cover the core of algebra notation presented by Mathtalk, and should be capable of being extended. A simple expression such as $3x + 4 = 7$ is divided

into terms by pauses; the start of each new term is also indicated by a pitch rise; multiplication is indicated by verbal juxtaposition and the imminent end of the expression is indicated by a sharp pitch fall. The complex part of the expression $3(x + 4) = 7$ is grouped together by a pitch fall, increased speed and no internal pausing. So, despite having the same lexical content, the prosodic structures are radically different.

These rules have been implemented in the Mathtalk program. Evaluation of the use of prosodic cues in comparison with lexical cues reveals several interesting improvements [16]. Listeners indicate an overall preference for the prosodic cues. Using prosodic cues instead of lexical significantly improved both apprehension of syntactic structure and retention of content. The absence of lexical cues probably reduces problems due to the suffix effect. The prosodic cues, particularly the pauses, chunk the utterance into meaningful subunits facilitating parsing, integration and retention in short-term memory. The prosodic cues can be thought to have the same effect as the spatial parsing cues present in print (see [9]).

This use of prosodic cues to make the spoken presentation more usable is one of the design principles of Mathtalk. However, enhancing the auditory display and making the presentation more usable is not enough to enable active reading by a listener. With both prosodic and lexical cues, recovery of structure and retention of content fails with greater *complexity* and length of expression. A full utterance will often overwhelm a listener and more importantly the listener is still the passive agent in the reading process.

3 Active Reading by Browsing Algebra

Visual reading is an active process because the reader has fast and accurate control over the information flow. This level of control needs to be given to the listening reader to make him or her a truly active reader. Mathtalk provides a series of browsing functions and the means to control them, that should enable this active reading.

A default reading strategy is provided. This is a left-right movement through the expression, revealing a term at a time, under the control of the reader. In each term the simple items are spoken in full. Any complex items are folded or hidden and simply referred to by their type. When a complex item is encountered, the next stage of the default strategy is to move into that item and successively reveal the contents.

The principle behind *folding* means that complex items are hidden from view and referred to only by their type. moving through the expression $3(x + 4) = 7$ at the top-level reveals the items 3, 'a quantity', = and 7. The sub-expression is referred to by its type label: 'A quantity'. Other reading moves can be used to explore the sub-expression or to reveal the whole item without exploration. This folding of complex items helps the reader maintain control of the information flow, aiding integration and retention.

There are two vital components needed to make successful use of the browsing: Planning and directing their use. How Mathtalk enables the listening reader to plan his or her reading strategy is dealt with in the next section. As well as the default reading style, access to any part of an expression must be given in an unconstrained manner, that allows review of the expression and modification of the reading style. The reader controls these browsing functions with a command language. The nature of the notation and associated browsing leads to a necessarily complex interface. Yet, The reading process is of primary importance and any method of controlling the information flow that is mediated externally, unlike visual reading, must intrude into the reading process as little as possible. The means of controlling the reading process, issuing commands and receiving feedback about moves made must be as transparent as possible so that the reading process is supported rather than disrupted.

Mathtalk's command language is a combination of **actions** (e.g. current, next and previous) and a series of syntactic **targets** (e.g. expression, term and item). An action word is combined with a target word to form a command. At present a mnemonic mapping to the key-board is used to issue commands. For example, to move to the next term the action **next** and target **term** are issued as the keyboard command **nt**. This makes the issuing of commands swift. Together with the wide coverage of the browsing gives the reader both fast and accurate control of the information flow.

By adding browsing functions to the user interface the reader may become the active agent in the reading process. The reader's method for directing the browsing must allow for adequate control in the reading process. Yet, the method of control must not intrude into the reading process itself. The command language uses a relatively small number of *action* and *target* words that can be combined to cover the large range of possible browsing moves. The speed and accuracy given by such a command language should go some way to mitigating the effects of the transience of the speech signal. Browsing may be more effective to refresh a reader's memory by giving easy access to any part of an expression, rather than relying on remembering the material recently heard.

The commands fall naturally into a spoken form and thus fit neatly into the speech based nature of the interface. The command language will be mapped to input by voice recognition. In the near future the control aspect of Mathtalk will be evaluated for ease of use, learning, predictability and effectiveness.

4 Planning the Reading Process

As well as allowing control in the reading process, the external memory allows a sighted reader to plan his or her reading process. To develop an appropriate reading strategy, the listening reader also needs to have some high-level information complexity of the expression. With only browsing, the reader has to read the expression to find out the complexity. Thus planning is not possible. The print memory allows the reader to skim the expression quickly and note

the salient features that determine its complexity. The auditory display needs a similar facility.

Speech was rejected as a means of providing this audio glance. A non-interpretive description of an expression's complexity would be too long. A glance should be a quick overview. Work on *earcons* [3, 5, 4] provided a method to use non-speech audio for this glance, giving a very quick, non-interpretive overview of an expression's structure.

Earcons are abstract, structured sequences of sound that may be used within a computer interface to represent objects or events to the user [3]. The parameters that describe earcons are similar to those that describe the prosodic component of speech. Parallels between guidelines for earcon construction [4] and the rules for algebraic prosody mean there are many similarities between the two forms of information. The concepts of earcon structure and prosodic cues have been combined in *algebra earcons* to provide an audio glance to facilitate planning in reading algebra.

In algebra earcons each syntactic item is represented by a different musical timbre. Prosodic rules are then combined with the algebra syntax to give structure to the sounds. A full set of rules for generating algebra earcons have been developed for Mathtalk [15]. Evaluation of the earcons has shown that listeners can recognise even very complex expression structures from listening to an algebra earcon. Subsequent evaluation has shown that people retain a useful mental representation of an expression after hearing an algebra earcon. This representation may vary from an idea of the complexity to a very detailed view of the target expression. When integrated with the other components of Mathtalk, this representation should be useful in providing information to plan the reading of the expression. A good representation of the expression could be used to help resolve grouping ambiguities in the spoken forms of the expression and provide a cognitive framework for the reading process.

Algebra earcons are capable of providing the information needed to plan reading of an algebra expression. The rôle of algebra earcons could be extended. An audio glance could provide orientation and navigation information for the listening reader. The use of musical timbres to represent syntactic types could be exploited. The same timbres could be used as background sounds, that are played while the reader is within that syntactic environment. The onset of such background sounds is noticed by the listener, but fade into the background [6]. The listener can then sample such sounds to determine the current environment, and again notice the switching off of the sound as he or she leaves the current environment.

5 Discussion

The Mathtalk program gives blind readers a truly active reading interaction with algebraic material. The design principles derived from this work will be applied to other potentially complex information, such as programming languages. This work also has a wider significance for the design of non-visual interfaces. The

ability to read such material in a quick and active manner has the potential to enhance educational prospectus of many blind people.

The design principles derived from the Mathtalk program are based on solving the problems of *control* and *external memory*. Once a decision has been made on what level of information to present, such design decisions can be made. The transience of the speech signal is mitigated by the addition of prosody, to ease retention, and giving quick and accurate access to any part of an expression, via browsing. Prosody can also replace the rôle of spatial grouping in printed form, preventing grouping ambiguity.

The control aspect of reading is enhanced with addition of the browsing functions and associated command language. This allows the reader to maintain control over the information flow and shift attention to any part of an expression. This ability will allow completely unambiguous reading of an expression, which is not entirely possible with an uncontrolled full utterance.

To make full use of the potential control a reader needs to plan his or her use of the browsing functions. *Algebra earcons* are provided to give an audio glance at an expression. This glance at least gives the reader an idea of the expression's complexity and in the best case provides a framework into which details can be fitted and that can help resolve grouping ambiguity.

Each major component of the Mathtalk system will have been evaluated for its effectiveness in tackling a particular problem of reading algebra using speech and sound. The whole Mathtalk system will then be evaluated. Such evaluation is necessary for the development of usable interfaces that mediate such complex tasks as the reading of algebra. It is not enough to enable presentation of information, that material has to be able to be read in an active, effective and efficient manner.

The work on the Mathtalk program has fed into the European Tide Maths project. This project seeks to produce an algebra work-station for the reading, writing and manipulation of algebra notation. This takes a multi-media approach, using Braille, enhanced visual display, speech and non-speech audio to give a presentation suitable for many visual conditions and tastes. The work to be carried out at York includes methods for writing and manipulating algebra; use of non-speech audio in navigation and orientation, extending the work on algebra earcons; enhancing the work on algebraic prosody.

Acknowledgements

We would like to thank John Local, Department of Linguistics, University of York for his help in the investigation of algebraic prosody. This work was supported by SERC grant 91308897.

References

1. F. Aldrich and A. Parkin. Improving the retention of aurally presented information. In M. Gruneberg, P. Morris, and R. Sykes, editors, *Practical Aspects of Memory 2: Current Research and Issues*. Chichester, England: Wiley, 1988.

2. A. D. Baddeley. *Human Memory: Theory and Practice*. Lawrence Erlbaum Associates Ltd, 1990.
3. M. Blattner, D. Sumikawa, and R. Greenberg. Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 4(1):11–44, 1989.
4. S. A. Brewster, P. C. Wright, and A. D. N. Edwards. An evaluation of earcons for use in auditory human-computer interfaces. In *INTERCHI'93*, pages 222–227. ACM Press, Addison-Wesley, 1993.
5. S. A. Brewster, P.C. Wright, and A.D.N. Edwards. A detailed investigation into the effectiveness of earcons. In G. Kramer, editor, *Auditory Display: The Proceedings of the First International Conference on Auditory Display*. Addison-Wesley, 1992.
6. W. Buxton, W. Gaver, and S. Bly. Tutorial no. 8: The use of non-speech audio at the interface. In *CHI'91 Conference proceedings, Humantems*. ACM Press: Addison-Wesley, 1991.
7. L. A. Chang. *Handbook for Spoken Mathematics (Larry's Speakeasy)*. Lawrence Livermore Laboratory, The Regents of the University of California, 1983.
8. S. G. Hart and C. Wickens. *Workload assessment and prediction*, pages 257–296. New York: Van Nostrand Reinhold, 1990.
9. D. Kirshner. The visual syntax of algebra. *Journal for Research into Mathematics Education*, 20(3):274–287, 1989.
10. E. Lehiste. *Suprasegmentals*. MIT Press, 1970.
11. M. H. O'Malley, D. R. Kloker, and B. Dara-Abrams. Recovering parentheses from spoken algebraic expressions. *IEEE Transactions on Audio and Electroacoustics*, AU-21:217–220, 1973.
12. K. Rayner and A. Pollatsek. *The Psychology of Reading*. Prentis Hall, 1989.
13. P. J. Simpson. Enhancing text to synthesized speech systems for the blind. In *Proceedings of the IEE Colloquium on Special Needs and the Interface*, January 1993. IEE Digest no. 1993/005.
14. R. D. Stevens. principles for the design of systems for reading complex information by visually disabled people. Internal report, 1993.
15. R. D. Stevens. Rules for the construction of algebra earcons. Internal Report, 1994.
16. R. D. Stevens, P. C. Wright, and A. D. N. Edwards. Prosody improves a speech based interface. Unpublished, 1994.
17. L. A. Streeter. Acoustic determinants of phrase boundary representation. *Journal of the Acoustical Society of America*, 64:1582–15:1 92, 1978.