TalkMaths : A Speech User Interface for Dictating Mathematical Expressions into Electronic Documents

Angela M. Wigmore, Gordon J.A. Hunter, Eckhard Pflügel and James Denholm-Price

Faculty of Computing, Information Systems and Mathematics,
Kingston University, KT1 2EE, U.K.
{k0330947, G.Hunter, E.Pfluegel, j.denholm-price}@kingston.ac.uk

Abstract

We describe the development of a speech-driven user interface system, *TalkMaths*, which enables the dictation of mathematical expressions into electronic documents without the user needing extensive knowledge of any specialized markup language. This system should be of value to many students and teachers, particularly those with disabilities - for whom typing mathematical text is currently very difficult.

1. Introduction

Knowledge of, and skill in using, mathematics and mathematical notation, at least at an elementary to intermediate level, are fundamental to achieving success in many disciplines – including most of the natural sciences, engineering, psychology, economics and commercial subjects. Typing, formatting and editing mathematical expressions in documents are becoming more important for a wide range of students and professionals in many of these areas, in addition to mathematical specialists. However, the process of doing these tasks is rather tedious, slow and error-prone for even quite experienced users of the relevant I.T. systems. Many disabled students – especially the blind and visually impaired, but also those who have lost (or lost use of) their arms or hands – are at particular disadvantage when both learning and using conventional mathematical notation. Braille notation and devices are not particularly well-suited for use with mathematics, and most common systems for entering and editing mathematical text into electronic documents rely on the user being able to use a keyboard and mouse, and visually inspect the output produced.

In this paper, we describe the development of *TalkMaths*, a user interface system which allows elementary mathematical expressions (up to the level of material in the GCSE examinations, normally taken by students in the U.K. around the age of 16) to be dictated into electronic documents. The aim is that our system should complement others giving synthetic speech feedback and/or Braille output to the users, and thus be a major aid for disabled students (and others) learning and using mathematical notation. This should also help them in learning other disciplines, including those noted above.

2. Overview and Background

We start, in Section 3, by reviewing existing systems, and specifying what we aim for our system to do, and its relation to and dependency on other technologies. We then briefly describe how our system functions, and then investigate how features such as the way people speak mathematical expressions, both in terms of the words they use and the way they say them, could be exploited in our system through statistical language models and use of prosodic and timing information. We conclude by summarizing the current system, its achievements and limitations, and proposing future developments.

2.1. Spoken language technology

Automatic Speech Recognition (ASR) technology has now reached a level where real applications – such as controlling devices – beyond the dictation of simple text is now possible. Reviews of the ASR process can be found in [1, 2, 3]. Many ASR systems make use of Statistical Language Models (SLMs) [2, 3] and/or Context Free Grammars (CFGs) [4]. The simplest type of SLMs are *n*-gram models, using statistics relating to sequences of *n* successive words [3] found in "training data". In Section 4.1, we try to develop specialised *n*-gram models for spoken mathematical language with a view of incorporating these into our system.

A *context-free grammar* (CFG) is a set of *recursive rules* or *productions* used to generate or analyse patterns of strings [4]. A CFG consists of a finite set of variables (*non-terminals*) which are determined by the *production rules*.

A *production* is a *rule* which associates strings that are formed by concatenating non-terminals and possibly other primitive symbols called *terminals*. A *rule* is of the form $S=W$ where S is a non-terminal and W is either a set of terminals and/or non-terminals or can be "empty".

All mathematical expressions – even the most complicated, containing operations such as differentiation, roots, etc – can be uniquely represented as a *parse tree* which matches a context-free grammar. Figure 1 shows an example of a parse tree for the expression $\sqrt{x+2y}$.

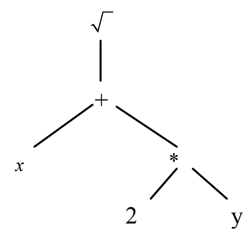

Figure 1: Sample Parse Tree showing the Structure of the Expression $\sqrt{x+2y}$.

In our system, a customized CFG for mathematical expressions is used to produce a parse tree, which is then converted into appropriate mark-up for the current expression (see Section 5.1).

3. Review of Existing Systems

Existing systems that enable the dictation of mathematics into electronic documents include *MathTalk*™ [5] and *CamMath* [6]. However, these are dependent on commercially available mathematics editors, for example Scientific Word, and are not freely available. However, they both support a wide range of mathematics. *Math Speak & Write* [7] also supports the dictation of mathematics, uses Microsoft's® *Speech Application Programming Interface* (SAPI) but has a small set of mathematical vocabulary. Bernareggi and Brigatti's [8] system (and documentation on it) is currently not yet freely available, and it only supports dictation of mathematics in Italian. However, their system is compatible with the commonly used equation editing system *MathType*. Hanakovič and Nagy [9] propose a system that is designed to be a formula writer for visually impaired people. Their system uses the *XHTML+Voice* (X+V) technology and is restricted to the *Opera* web browser.

4. Speaking and Interpreting Mathematical Expressions

Mathematics is a difficult subject for many people, let alone a blind or visually impaired person. Understanding a mathematical expression that is being described/read-out to such a person can cause ambiguity if important details are omitted or left unclear.

There are no explicit, universally accepted standards on how to articulate mathematical expressions in speech. However there have been several attempts at developing such standards or guidelines. Two notable blind mathematicians, Larry Chang [10] and T.V. Raman [11], have tried to create specifications on how to communicate mathematics "clearly". Fateman [12], Schleppenbach [13] and Fitzpatrick [14] have also made suggestions on this topic but their recommendations have not been widely adopted. In addition, David Hall [15] issued guidance to aid teachers in communicating and understanding mathematics. Despite these attempts, there remain many ways in which people say the same mathematical expression. In Sections 4.1 and 4.2 below, we describe and discuss evidence of the variety in how people actually "speak mathematics" in British English.

4.1. Statistical evidence of word frequencies

The British National Corpus (BNC) [16] is a collection (exceeding 100 million words) of text material and transcriptions of speech from a variety of sources, intended to represent modern British English, both written and spoken. For our study we selected a subset of the BNC dialogue material (15 files - 123745 words in total), composed of transcriptions of audio recordings from school, college and university mathematics lessons. The transcribed speech was extracted from the BNC files and analysed. Table 1 shows the "unigram statistics" for the ten notable mathematical words (excluding common numbers "one" to "six" – "one" occurring the most : 1718 times), which were the most commonly-spoken in this subset. The trend shown by the common numbers was broadly in line with that observed in other studies [17].

Table 1: Unigram statistics for notable mathematical keywords within our subset of the BNC data

Occurrences	Mathematical Keyword
588	X
488	A
483	hundred
449	minus
387	times
352	twenty
328	take
328	add
326	ten
312	plus

It should be noted that these "mathematical keywords" can have various meanings. For example, "A" could mean an algebraic symbol "A" or be part of (say) "A hundred". We have also compiled comprehensive statistics of all bigrams and trigrams containing mathematical keywords in the data set, with a view to making use of these to improve our current system. This showed that the word pairs "a hundred" (192 times) and "hundred and" (172 times) appeared together the most. Similarly the most common trigram was "a hundred and", appearing 120 times.

4.2. Speech timing and prosody

When reading-out mathematical expressions, people tend not to include parentheses and/or lexical indicators such as "begin square root" and "end fraction", because omitting these seems more natural and saves them time and effort. However if such an expression is to be read-out to a blind or visually impaired person, these cues will need to be included so that the expression is not ambiguous. There is evidence [11] that prosodic information – pitch and timing – can be used to distinguish between similar but non-identical forms.

We have conducted a study into how 10 people (5 students and 5 staff) at Kingston University, with various levels of mathematical expertise, read-out some mathematical expressions. Table 2 shows a comparison of the timing information from two similar expressions which, when read-out, could potentially be "ambiguous". We found that, from analyzing this data, regardless of their mathematical experience, all the participants included pauses to indicate the grouping of the terms in the expression.

Table 2: A comparison of the pauses (in seconds) taken between "a" and "plus" across speakers.

	Regular Pause taken between "minus" and "B"	$R=\frac{\sqrt{a+b}}{a-b}$	$S=\frac{\sqrt{a}+b}{a-b}$
Min	0.05	0.01	0.09
Mean (SD)	0.08 (0.02)	0.07 (0.03)	0.33 (0.25)
Max	0.11	0.13	0.85

Applying a paired sample t-test to this data showed that there was a highly significant difference ($t=3.17$, $df=9$, $p\approx0.01$) between the length of pauses in expressions R and S, but a non-significant difference ($t=1.12$, $df=9$, $p\approx0.29$) between the "regular pause" and that in expression R. This shows that speech timing information has potential to help resolve ambiguities in such dictated expressions.

5. *TalkMaths*

TalkMaths is a speech user interface that enables the user to dictate mathematical expressions using relatively natural language, and translates them into electronic document formats for display in a text document or webpage. The *TalkMaths* project builds on the previous *KU-TALK* system [18, 19]. The system is designed for use by anyone, particularly students, with a good or reasonable understanding of mathematics. It aims to help both physically disabled (and, to some extent, visually impaired) and able-bodied users convert spoken mathematical expressions into *MathML* [20] and *LaTeX* [21], without them needing to have extensive knowledge of the syntax of either "mark-up" language.

The system uses *Dragon® NaturallySpeaking®* (*DNS*) [22] as the ASR "front-end" and an open source C++ library created by Joel Gould called *NatLink* [23]. This is used to extend *DNS* through *Python* [24] macros and a *Voice Command Language*, *Vocola* [25].

5.1. How *TalkMaths* works

To begin dictating mathematics the user opens the "preview window" by saying "start *TalkMaths"* or clicking the relevant icon, which activates the mathematical speech grammar. The utterance *DNS* has tried to identify is passed to *NatLink*, which will then look for a match in the grammar and execute the corresponding procedures. The utterance is then converted into equivalent notation, e.g. "one plus two" is converted into "1 + 2". For precision, the alphabetic characters (A-Z) are dictated using the *NATO phonetic alphabet* names [26].

The expression is then parsed using *YAPP* ("Yet Another Python Parser"), version 2 [27], which ensures the expression is syntactically correct and returns a *parse tree* representing the structure of the expression. This *tree* is then converted into *MathML presentation mark-up* and displayed in the *TalkMaths Preview Window* (see Figure 2), which is based on the open source *XULRunner* (upon which the browser Firefox version 3 is based). The expression is then stored in memory and can be translated into *LaTeX* or *MathML* for use in other applications by saying "export to *LaTeX*" (or similarly for *MathML*) whereupon the relevant mark-up is transferred to the current window. The expression is retained until *TalkMaths* is closed or the user says "clear expression".

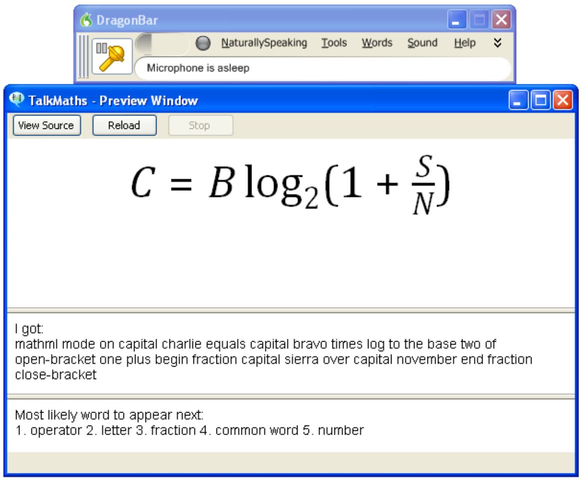

Figure 2: *TalkMaths Preview Window*.

The system also displays the recognised utterance from *DNS* and tries to predict what type of word will be dictated next. The data obtained from our study outlined in Sections 4.1 and 4.2 have been used to create an algorithm that looks two tokens before the last word and assigns "likelihoods" to what might appear next. The results are displayed in the preview window, listed in order of how likely each type of word is to appear next. However, this functionality is currently somewhat limited and development of it is still work in progress.

6. Future Work and Conclusions

TalkMaths currently supports mathematical expressions up to approximately the level found in the British GCSE (General Certificate of Secondary Education) examinations, which are a set of qualifications taken by secondary school students in England, Wales and Northern Ireland, normally around the age of 16, that include elementary algebra and trigonometry.

TalkMaths has the ability, albeit limited, to predict what type of word will appear next, e.g. a number, letter, operator, etc. It is planned to make the system more "intelligent" and "adaptive" by using our specialised mathematical SLMs to enable it to predict the words which are most likely to appear next (similar to "predictive text" for SMS messages). Using the statistical, timing, and prosody evidence, the system should provide "option lists" for ambiguous expressions. The system already provides some feedback via synthetic speech. However, it is planned to extend this to complement other systems which produce mathematical output in a mode accessible to blind people (e.g. Braille and more extensive feedback via synthetic speech), such as those resulting from the *Lambda Projects* [28, 29].

Our system has already proved of benefit to staff and students at Kingston University, but is still under development. It should be of value, particularly to students with disabilities, in the wider education community. Further details of the system can be found at http://talkmaths.sourceforge.net.

7. Acknowledgements

Angela Wigmore is supported by a Doctoral Training Award from the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom. The initial phases of this project were partially funded by Kingston University's Disability Stewardship Fund. We would also like to thank Dr. Mark Huckvale of University College London for advice regarding the BNC.

8. References

[1] Delgado, R.L. and Araki, M., *Spoken, Multilingual and Multimodal Dialogue Systems: Development and Assessment*, John Wiley & Sons Ltd, West Sussex, 2005.

[2] Rosenfeld, R., "Two Decades of Statistical Language Modeling: where do we go from here?", *Proceedings of the IEEE*, **88(**8), 1270-1278, 2000.

[3] Young, S., "Large Vocabulary Speech Recognition: A Review", *IEEE Signal Processing Magazine*, 13(5), 45-57, 1996.

[4] Hopcroft, J.E. and Ullman, J.D., *Introduction to Automata Theory, Languages and Computation*, Addison-Wesley Publishing company, Inc., Reading, MA, 1979.

[5] Metroplex Voice Computing, Inc., "Tech Notes". [Online].Available:http://www.mathtalk.com/tech_notes .htm.[Accessed: May 8, 2008].

[6] Elliot, C. and Bilmes, J.A., "Computer Based Mathematics using Continuous Speech Recognition", *Striking a C[h]ord: Vocal Interaction in Assistive Technologies, Games and More*, CHI 2007 workshop on non-verbal acoustic interaction, San Jose, CA, 2007.

[7] Guy, C., Jurka, M., Stanek, S., and Fateman, R., *Math Speak & Write, a Computer Program to Read and Hear Mathematical Input*, University of California, Berkeley, Electrical Engineering and Computer Sciences Department Technical Report, 2004.

[8] Bernareggi, C. and Brigatti, V., "Writing Mathematics by Speech: A Case Study for Visually Impaired", *Proceedings of ICCHP 2008 (11th International Conference on Computers Helping People with Special Needs),* 879-882, 2008.

[9] Hanakovič, T. and Nagy, M., "Speech Recognition Helps Visually Impaired People Writing Mathematical Formulas", *Proceedings of ICCHP 2006 (10th International Conference on Computers Helping People with Special Needs)*, 1231-1234, 2006.

[10] Chang, L.A., *Handbook for Spoken Mathematics (Larry's Speakeasy)*, Lawrence Livermore National Laboratory, University of California, USA, 1983.

[11] Raman, T.V., *Audio System for Technical Readings*, Springer Verlag, Berlin, 1998.

[12] Fateman, R., "How can we speak math?". [Online]. Available:http://www.cs.berkeley.edu/~fateman/papers/speakmath.pdf. [Accessed: July 13, 2007].

[13] Schleppenbach, D.A., et al. "2004 MathSpeak Initiative_". [Online]. Available: http://www.gh-mathspeak.com/. [Accessed: July 13, 2007].

[14] Fitzpatrick, D., "Mathematics: How and What to Speak", *Computers Helping People with Special Needs*, Springer, Berlin, 4061: 1199-1206, 2006.

[15] Hall, D., *Nucleus English for Science and Technology: Mathematics: Teachers' Notes*, Longman Group Ltd, Great Britain, 1980.

[16] Burnard, L., "User's Reference Guide for the British National Corpus", Oxford University Press, Oxford, U.K. [Online]. Available: http://www.natcorp.ox.ac.uk/. [Accessed: October 2, 2008].

[17] Hunter, G and Huckvale, M., "Is it Appropriate to Model Dialogue in the Same Way as Text ?", *Proceedings of European Modelling Symposium*, London, September 2006

[18] Hunter, G.J.A., Pfluegel, E. and Jalan, S., ""KU-Talk" – A Speech User-Interface for an Intelligent Working, Learning and Teaching Environment", *3rd IET International Conference on Intelligent Environments (IE 07)*, Ulm, Germany, 124-130, 2007.

[19] Denholm-Price, D., Hunter, G., Pfluegel, E. and Jalan, S., Wigmore, A., "The KU-Talk Project at Kingston University", *Proceedings of the Institute of Acoustics*, 30(1), 37-41, 2008.

[20] Carlisle, D., et al., "Mathematical Markup Language (MathML) Version 2.0 (Second Edition)". [Online]. Available: http://www.w3.org/TR/MathML2. [Accessed; August 6, 2008].

[21] LaTeX Project, "LaTeX – A document preparation system". [Online]. Available http://www.latex-project.org/. Available: September 18, 2008.

[22] Nuance Communications, Inc., "Dragon NaturallySpeaking 10 Product Suite". [Online]. Available:http://www.nuance.com/naturallyspeaking/products/default. [Accessed: August 7, 2008].

[23] Gould, J. and Gould, D., "Joel". [Online]. Available: http://www.gouldhome.com/joel/. [Accessed : August 8, 2008].

[24] Python Software Foundation, Python Programming Language. [Online]. Available: http://www.python.org/. [Accessed: October 14, 2008].

[25] Mohr, R., Vocola Information Pages. [Online]. Available http://vocola.net/. [Accessed: October 14, 2008].

[26] Law, G., "Phonetic alphabets (Alpha Bravo Charlie Delta)".[Online].Available:http://faqs.org/faqs/radio/phonetic-alph/full/. [Accessed: October 14, 2008].

[27] Patel, A., "Parsing with Yapps". [Online]. Available: http://theory.stanford.edu/~amitp/yapps/. [Accessed: February 6, 2008].

[28] Edwards, A. D. N., McCartney, H. and Fogarolo, F. "Lambda: A multimodal approach to making mathematics accessible to blind students", *Proceedings of Assets 2006*, Portland, Oregon, ACM, 48-54, 2006.

[29] Schweikhardt, W., Bernareggi, C., Jessel, N., Encelle, B., Gut, M., LAMBDA: A European System to Access Mathematics with Braille and Audio Synthesis, *Proceedings of ICCHP 2006 (*10th *International Conference on Computers Helping People with Special Needs)*, 1223-1230, 2006.