SpatialMaths: a Library for Conveying Content and Structure of Equations

Colin Fitzsimons[1], Emma Murphy[1], Catherine Mulwa[1] and Donal Fitzpatrick[1]

Dublin City University Glasnevin, Dublin 9, Ireland
colin.fitzsimons3@mail.dcu.ie, emma.murphy@dcu.ie, cmulwa@computing.dcu.ie, donal.fitzpatrick@dcu.ie

Abstract

Inaccessibility of STEM content is a well-known and established problem facing many students, researchers and those in industry. Diagrams, mathematical notation and formulae are most often expressed through visual structures, leading to a number of problems for visually impaired users. As more information is digitised there is a greater need to support universally accessible content. The web has seen a growth in the use of MathML, (Mathematical Mark-up Language) which is used to better represent notation. The SpatialMaths library aims to utilise this growing technology to enable non-visual access of mathematical content in both an educational and industrial setting. This paper is an early-stage report on the current state of the library, the testing that has been carried out to date and proposed further research. The focus falls primarily on the documentation of the techniques being employed as well as any shortcomings revealed through technical evaluation.

1 Introduction

Mathematics is a subject with considerable accessibility barriers for visually impaired students, academics and teachers. The spatial presentation of equations makes it difficult to render in a non-visual manner through the linear structures of Braille and synthetic speech. There are certain common problems that can cause ambiguity in spoken maths, whether communicated through human or synthesised speech. In algebra, for instance, fractions are a notable illustration of expressions where ambiguity can arise due to confusion of grouping. For example the spoken phrase "one over x plus 4" does not indicate whether the fraction incorporates "1 over x" as a single grouping or the entire phrase is part of the fraction. If we consider the visual representation of the following equations: $\frac{1}{x} + 4$ and $\frac{1}{x+4}$ we can see that such ambiguity simply does not arise.

In this paper, a library to generate auditory renderings of mathematical expressions is presented which uses a lexical structure in conjunction with a set of spatialised earcons and modified speech, to disambiguate the structure of mathematical formulae. The work presented here is at a very early stage of development, and has been carried out as part of a 12 week internship funded under the auspices of the Faculty of Engineering and Computing Intern Fund. The information given here is based on previously published work in this area [12, 5, 2] and as such should be seen in the context of this research. The aim of this paper is to provide information on this ongoing effort, with a view to fostering interest and possible collaboration. The remainder of the paper is structured as follows:

- Section 2 describes previous research which underpins the SpatialMaths library;
- Section 3 shows how various mathematical structures are conveyed in the library;

- Section4 describes the structure of the library;
- Section 5 describes a **very early stage evaluation** of the effectiveness of the library, and highlights some deficiencies;
- Section6 outlines some thoughts on the direction of future work.

2 Background

Karshmer et al. [10] highlight the two-dimensional nature of visual or printed mathematic equations, which is difficult to convey through both linear systems of synthesised speech and Braille. The spatial representation of equations can encode essential semantic information necessary for understanding the mathematical construct. Previous studies have investigated the possibility of retaining this spatial information through non-visual interface solutions to enhance accessibility for visually impaired users. Solutions have included lexical structures [14], spatial displays (audio and haptic) [9], non-speech audio cues [16] active browsing [9, 16, 15]and prosody [7, 5]. Various studies have indicated that the use of non-speech audio can help improve access to non-visual user interfaces, by reducing the burden of information through speech alone [3, 13]. Brewster et al. [4] describe how musical parameters such as timbre, register pitch and duration can be manipulated to create earcons and convey information to the user through sound. Where earcons are sonic messages that have abstract musical qualities, auditory icons are based on sounds sampled from real-world environments. Auditory icons are based on the attributes of source events (such as size, material and force) rather than in terms of physical attributes of the sounds themselves [6]. While it is beyond the current scope of this paper to present a detailed literature review on previous approaches to creating an accessible platform to mathematics it is important to consider that there have been previous attempts to enhance spoken mathematics with non-speech sound in addition to prosody. Notably the Mathtalk system [16] uses musical earcons to indicate structural delimiters (such as a pattern of 3 short ascending and descending notes to indicate the beginning and end of an expression respectively) and also provide an abstract overview of the entire equation [16]. The significant drawback to this approach is that a complex and highly specific musical grammar must be learned prior to using the system. In addition, remembering and decoding musical patterns may be quite difficult for non-musicians and again, the additional cognitive effort required to decode each pattern could detract from the processing of the mathematical content. This is also symptomatic of a wider issue in the field of auditory display design. At present, most sound design methods for non-speech sounds in auditory interfaces are based on empirical knowledge. This often results in the creation of sounds derived from random selection or the personal preference of the sound designer. Such ad hoc approaches to interface design are often part of a simple implement and evaluate iterative cycle, without a proper design phase.

3 Audio Depiction of Mathematics

The aim of the SpatialMaths library is to present a mathematical expression through the use of synthesised speech to depict the content, and non-speech audio in the form of earcons and spearcons [19]to convey the structure which is readily visually apparent. The spoken component of the equation is presented using standard stereo, with the utterance being balanced equally between left and right. The non-speech audio, however, is binaurally processed to left or right, to aid in the disambiguation of the structure of the material being presented.

There have been previous attempts to directly map the visual spatial structure of printed mathematics to an equivalent spatialised audible structure in order to reduce the mental effort required by the user to process and solve the equation. However, the spatial resolution of vision is much more accurate than audition and this suggests that accurately replicating the spatial layout of a printed equation with spatial audio is difficult to achieve. In fact, it has been found [8] that the additional mental processing required to determine the spatial trajectory detracts from the processing of content. The vertical layout of mathematical equations would appear to be well matched to a spatial audio presentation of an equation, however, auditory localisation is particularly inaccurate in the vertical dimension and extremely difficult to synthesise, particularly with binaural spatialisation techniques based on the head-related-transfer function, or HRTF.

However spatial audio should not be dismissed as a redundant parameter to enhance spoken mathematics. Previous studies have illustrated that sounds produced from different spatial locations are easier to distinguish, which suggests that if additional sounds are added to the main speech signal, these should be produced from different spatial locations. The externalisation effect which generally occurs when binaural material is presented via headphones has also been found to be much less fatiguing than standard stereo, and this may also be of some benefit. Therefore, the binaural cues in this present system are intended to reinforce the lexical structure of spoken maths while at the same time relieving some of the cognitive burden of speech processing. A left/right binaural spatial position has been implemented to convey the lexical open/close or begin/end.

Spearcons [19]are created by time compressing a spoken phrase so that the resulting short sound is not entirely comprehensible as speech. Spearcons are intended to be perceived somewhere between recognisable speech and abstract sounds such as earcons. It is proposed that spearcons are easier to learn and result in an increase in performance in interactive interface tasks such as menu browsing. The major advantage of spearcons for this present system is that they can potentially function as either a descriptive lexical phrase or an abstract sound depending on its familiarity to the listener, or its function as a structural delimiter. It should be noted that the spearcons generated by Walker et al. and described in [19] were not utilised in this current system. Rather, a bespoke set of spearcons was generated based on the techniques outlined in [19].

The following provides a summary of the auditory cues implemented in the system:

3.1 Brackets

In this current system, brackets are indicated by a short beep-like earcon of 20ms, positioned to the left and with a rapid upward frequency glissando indicates an opening bracket, to the right and with a downward glissando indicates a closing bracket. To provide users with a design choice, an alternative bracket sound was also designed in contrast with the short beep earcon. This ?noise? earcon consists of synthesised white noise which was dynamically filtered to become broadband for an opening bracket, and narrowband for closing. This sound was designed with a long attack and release to make it as unobtrusive as possible and had a duration of 400ms.

3.2 Fractions

Fractions are depicted using the spearcon ?frac?, played from the left to indicate the beginning of a fraction, and from the right to mark the end.

3.3 Superscripts/Subscripts

Superscripts are indicated using the spearcon ?sup? and an EQ change which reduces bass frequencies and increases the treble. This is also implemented using a left and right ?sup? spearcon like the other spearcons. Subscripts are similarly indicated using the spearcon ?sub? and an EQ change that reduces the treble and increases the bass frequencies.

3.4 Nested Layers

Nested layers have been implemented by adjusting the spatial position and pitch of the beep-like bracket earcons. Nested layers are also indicated by adjusting the speaking rate by 6

4 Library Structure

The library functions as three separate modules, each with a clearly defined and self-contained purpose. Firstly, the rendering layer handles taking MathML content in a raw format and creates a series of strings to represent the information. This process converts the raw MathML into format which may be more readily transformed into an auditory format, retaining the structure and integrity of the information. The rendering interface is intended to provide a set of function headers which can be used by developers to create new rendering rulesets, which may be used interchangeably. Once the renderer has finished parsing the raw MathML content, the results are sent to the synthesiser layer. This layer controls how the content is translated into synthesised speech, which synthesiser is used and whether or not earcons and spearcons are added. The synthesiser layer also has a common interface which can be extended by developers who wish to incorporate their own speech synthesiser into the system. Lastly, the 3D Audio layer handles taking the content that has been synthesised and mapping it in a 3D aural space. Regardless of which speech synthesiser or renderer was used, this layer operates in the same way. Through the use of the irrKlang library [1], the equation is read from the material output by the synthesiser layer and presented aurally, replacing certain mathematical artefacts with earcons or spearcons.

Additional helper objects exist to provide common utilities for developers. These include a sanitiser which will check a raw MathML file for errors as well as a series of other tools for debugging and testing. The requirements for running the library are minimal, as such, the library can be deployed on any modern Windows operating environment. Users of the library must have both Microsoft .NET 4.0 installed as well as the Microsoft Visual C++ 2010 redistribution; the former of which is included in any version of Windows above 8 and the latter can be included in a distribution package.

A sample scenario has been provided to aid in describing the library at a high level. In this example a developer has incorporated the library into a UI they designed. Note, the example is presented textually here to aid those who cannot see a diagrammatic representation, and has been done in this fashion to aid the accessibility of this document. The example illustrates how the content is parsed from start to finish, the library pipeline can be viewed as such:

1. A UI layer takes MathML content from a user.

2. The content integrity and validity is checked.

3. Options are provided to both the renderer and synthesiser layers.

4. A renderer and a synthesiser are selected.

5. The raw MathML content is passed to the renderer.
6. The renderer passes a series of strings to the synthesiser layer.
7. The strings are converted to audio objects by the synthesiser.
8. The 3D Audio layer acquires this data and prepares it for use.
9. The user steps through the equation using the UI which calls bindings in the 3D Audio layer.

It should be pointed out that the library itself does not provide facilities to step through the material. Rather, it is presumed that the user will utilise their Assistive Technology of choice to do so. Owing to the highly commendable progress in the navigation strategies offered by Screenreaders such as NVDA, it is envisaged that the SpatialMaths library would be incorporated and used in conjunction with these technologies.

5 Evaluation

The Maths Learning Centre in Dublin City University is an initiative setup to help counteract one of Irelands largest causes of poor student retention [17]. The centre provides additional support to students struggling with third level mathematics through one-on-one sessions during the academic year. This process involves the provision of additional learning materials in the form of worksheets. Although non-visual users can greatly benefit from one-to-one support they cannot avail of the worksheets provided by the centre. Therefore, testing and transcribing a subset of mathematical equations found on the worksheets using the SpatialMaths library enables both a technical evaluation of the library as well as creating a platform and relationship with the Maths Learning Centre for future user testing.

In order to assess the system a total of forty unique equations have been drawn from the available worksheets [18] and a further ten were taken from the Mozilla MathML torture test [11] to see how the library handles complex and difficult mathematical equations. The forty tests were split into a total of eight categories; fractions, superscripts, subscripts, square roots, functions, sets, whole equations and miscellaneous mathematical notation. Each category had five equations taken from the worksheets which contained notation related to their topic. Testing was carried out by first locating suitable equations and creating a repository of MathML files. The intended purpose of preserving the test files is to be able to reuse them to test future versions of the library as well as being able to test the current implementation. The files were separated into folders, grouped by their categories. A simple UI was constructed for the purpose of testing the library. Each equation was fed into the system and synthesised. A comparison was then made between what information was conveyed aurally by the library versus the original equation in its visual form. The test sought to identify any edge cases that may cause the library to fail as well identify any problematic or ambiguous renderings before moving onto user testing. Any discrepancies, ambiguities or errors have been noted and a subset of problematic equations can be seen below.

1. $\sqrt{4y - z} + x^3$
2. $\sum_{\mathbf{i=1}}^{\mathbf{100}} \mathbf{i}$
3. $4^{\frac{1}{3}} \times 4^{\frac{1}{3}} \times 4^{\frac{1}{3}}$

4. $a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4}}}}$

5. $$f(x) = ax + b, a \in \mathbb{N}$$

The system successfully conveyed 84

Certain mathematical notation, such as that in equation five, do not render correctly. Although the information is conveyed, it is done so in a non-natural manner. Ideally, the system would have a special case defined for catching such expressions and rendering them accordingly. Variables and notation such as capital letters are read as ?cap A?, ?cap B? and so on which resulted in a clear method of distinguishing between upper and lower case letters. Other mathematical notation such as set operands function as expected and are both rendered and synthesised as intended.

6 Future Work

The work of producing the SpatialMaths library has only recently started, so it is certainly not the intention here to give the impression that it is a complete project. Initial findings suggest that the library needs to incorporate a more precise method of conveying ambiguous information such as square roots, subscripts and superscripts. Although methods for disambiguating fractions and summations exist [2] those which are used to depict superscripts and subscripts require further examination to improve their efficacy. The current prosody models, whilst sound in theory, are very much dependent on the implementation by speech synthesiser manufactures. Before proceeding further, investigations need to be carried out to ensure a mapping between the theoretical framework defined in [5] and the practical way in which it can be realised on modern synthesisers.

The intention is to attempt to produce a web app incorporating this library with a view to enabling students (at a certain mathematical level) to upload equations provided by, for example, educators, and have them read back. If the library is to be used by both students and educators, then additional work will need to be completed to prepare the library. Currently, the library only provides assistance to non-visual users. In order for sighted educators to be able to aid students it will be necessary to incorporate a visual renderer that will operate alongside the audio component.

Another interesting avenue is to explore the feasibility of including this library in the freely available NVDA Screenreader, with a view to providing an alternative rendering of the mathematical material.

References

[1] Ambiera. Irrklang. online.

[2] Enda Bates and Donal Fitzpatrick. Spoken mathematics using prosody, earcons and spearcons. In *Computers Helping People with Special Needs, 12th International Conference, ICCHP 2010, Vienna, Austria, July 14-16, 2010, Proceedings, Part II*, pages 407–414, 2010.

[3] SA. Brewster. Using non-speech sound to overcome information overload. In *Displays 1997*, pages 179–189, 1997.

[4] S.A. Brewster, P.C. Wright, and A.D.N. Edwards. Experimentally derived guidelines for the creation of earcons. In *Adjunct Proceedings of HCI'95*, 1995.

[5] D. Fitzpatrick. Mathematics: how and what to speak. In *Proceedings of ICCHP, International Conference on Computers Helping People with Special Needs*, Lecture Notes in Computer Science, pages 1199–1206. Springer Verlag, 2006.

[6] W. W. Gaver. Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2(2):167–177, 1986.

[7] E. Gellenbeck and A. Stefik. Evaluating prosodic cues as a means to disambiguate algebraic expressions: An empirical study. In *ASSETS 2009, the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, 2009.

[8] S. Goose and C. Möller. A 3d audio only interactive web browser: Using spatialization to convey hypermedia document structure. In *In Proceedings of the ACM International Conference on Multimedia*, pages 363–371, 1999.

[9] P. A. Harling, R. Stevens, and A. Edwards. The design of an algebra manipulation tool for visually disabled mathematicians using spatial-sound and manual gestures. Technical report, HCI Group, University of York, 1995.

[10] A.I. Karshmer and D. Farsi. Access to mathematics by blind students: A global problem. In *Proceedings of the WMSCI 2007 Conference*, 2007.

[11] Mozilla. Mathml torture test. Online, July 2015. Retrieved from Mozilla Developer: `https://developer.mozilla.org/en-US/docs/Mozilla/MathML_Project/`.

[12] E. Murphy, E. Bates, and D. Fitzpatrick. Designing auditory cues to enhance spoken mathematics for visually impaired users. In Armando Barreto and Vicki L. Hanson, editors, *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2010, Orlando, FL, USA, October 25 - 27, 2010*. ACM, 2010.

[13] E. Mynatt. Transforming graphical interfaces into auditory interfaces for blind users. *Human-Computer Interaction*, 12(1):7–45, 1997.

[14] A. Nemeth. Mathspeak. Online. Nemeth, A.: Mathspeak. Available: `http://people.rit.edu/easi/easisem/talkmath.htm`.

[15] T. V. Raman. *Audio Systems for Technical Reading*. PhD thesis, Department of Computer Science, Cornell University, Ny, USA., 1994.

[16] R. D. Stevens. *Principles for the Design of Auditory Interfaces to Present Complex Information to Blind Computer Users*. PhD thesis, University of York, UK, 1996.

[17] Dublin City University. Report on the maths learning centre. Online, 2005. Available from http://www.dcu.ie/maths/mlc/MLCReport2004-05.pdf.

[18] Dublin City University. About the maths learning centre. Online, 2014. Available from https://www4.dcu.ie/maths/mlc/index.shtml.

[19] B. Walker, A. Nance, and J. Lindsay. Speech-based earcons improve navigation performance in auditory menus. In *Proceedings of the 12th International Conference on Auditory Display*, 2006.