SpeechMatE:
A Speech-driven Maths Editor for Motor-Impaired People

Cristian Bernareggi[1], Mattia Ducci[1], Alessandro Mazzei[3], Dragan Ahmetovic[2], Tiziana Armano[1], Anna Capietto[1], Sandro Coriasco[1], and Nadir Murru[1]
[1] Università degli Studi di Torino, Department of Mathematics G. Peano, Italy
[2] Università degli Studi di Milano, Department of Computer Science, Italy
[3] Università degli Studi di Torino, Department of Computer Science, Italy

Introduction

All over the world, millions of people have severe motor disabilities (e.g., diplegia, tetraplegia), temporary or permanent motor impairments (e.g., tendinitis, arthritis, musculoskeletal diseases), and motor disorders (e.g., developmental coordination, stereotyped movements), which hamper many of their everyday activities [17]. In particular, people with upper limb disabilities, amputations or hand disorders run into difficulty when required to write either with a pen on paper or with text-entry through keyboard, use a pointing device or operate a touchscreen.

So far a number of solutions have been developed to facilitate text-entry through techniques which do not require finger coordination, dexterity or fine motor skills (e.g. eye-tracking, speech recognition, adaptable onscreen keyboards, switches). [21, 22]. These solutions are effective when entering and editing text (i.e. characters, words and sentences), but they are not designed for typing and editing maths. Even though some of these assistive tools can actually be used for typing maths, they require a considerable amount both of cognitive load and physical effort [5]. For instance, a person with tetraplegia can use a head-mounted pointing device or an eye-tracking input system to select and enter maths symbols in an off-the-shelf editor, but the time and cognitive load required to perform these tasks are extremely high in comparison to average speed and effort for typing maths by pen on paper or through mouse and keyboard. As a consequence, concentration is drawn away from mathematical concepts and simplification rules towards the typing technique [16, 6, 5]. Recent development of speech-to-text (STT) technologies has paved the way for the design and development of new speech-based interaction paradigms which are to introduce motor-impaired people (MIP) to hands-free typing and editing of maths.

This work presents *SpeechMatE*, a prototype application designed to type and edit maths through speech input in the Italian language. To the best of our knowledge, it is the first tool that enables input of mathematics in Italian. At this stage of development, *SpeechMatE* can be used to type arithmetic and elementary algebra taught in the first year of Italian secondary schools.

Related Work

Research on techniques to support people with motor impairments in typing has investigated a number of different solutions. This section, at first introduces text-entry systems not specifically designed to type maths expressions, but which can be used also to enter math notation in conjunction with a mainstream maths editor. Afterwards, it examines systems designed to type mathematics via speech input.

Text-Entry Systems for People with Motor Impairments

Text-entry techniques for MIP can be classified in five categories: direct selection, scanning, pointing, gestures and speech-to-text [21].

Through direct selection MIP can select one out of a limited set of keys. This technique is used in chording [11], ambiguous [12] and encoding [19] keyboards. Through scanning, items in a list (e.g., characters, words) are automatically highlighted sequentially until one is selected [14, 13]. Text-entry based on pointing is performed by selecting characters through a pointing device controlled by a body part such as head [10], eyes [15], nose [20], foot [18] etc. Typing based on gestures is achieved through the recognition of a set of gestures (e.g., through a camera or a touchscreen), representing characters or words[2] [26].

These solutions present three main limitations: low typing speed, error proneness and a steep learning curve [21, 22], [9]. By contrast, thanks to recent development of speech recognition techniques, STT enables fast and effective dictation of entire sentences at a time. *SpeechMatE* adopts STT to facilitate maths typing.

Speech-to-Text Systems for Writing Mathematics

This section introduces the systems which have been developed to write mathematics through speech input. TalkMaths [24, 25] is a prototype application which translates arithmetic, elementary algebra and trigonometry from spoken English into LaTeX or MathML. It adopts Dragon Naturally Speaking[3] (DNS) as speech recognition system. The translation rules are defined only for English and they are based on pauses, which slow down the dictation process [8].

Mathifier [1] is an open source software module which converts maths expressions from English into LaTeX. It combines a dictionary, a language model and an acoustic model to recognize mathematical English utterances. It uses Sphinx-4 [23] to recognize speech. This system has been designed for recognizing maths only in English, and as yet no reliable Italian speech recognition has been enabled.

CamMath [8] is designed to prove the advantages of continuous speech over discrete utterance of mathematical expressions in English. It works as a front-end for Scientific Notebook[4], a commercial maths editing system. It uses Microsoft speech recognition platform which supports recognition of the Italian language only through Microsoft development tools.

Bernareggi et al. [3] investigate speech input of mathematics in the Italian language for the LAMBDA editor [2], a commercial maths editor for visually impaired people. Speech is recognized through DNS. A dictionary of symbols and commands which are valid for the LAMBDA editor is defined. Hence, the dictation process is limited to the LAMBDA editor and it is far from dictation in natural language. Metroplex MathTalk[5] is a commercial application that provides speech input of arithmetic, algebra, calculus and statistics in English. Recognition is based on DNS and it implements a dictation model based on pauses, which slow down the dictation process.

EquatIO[6] enables dictation of simple maths expressions in English in MS Word and in GSuite applications.

SpeechMatE, similarly to CamMath and Mathifier, adopts continuous speaking interaction and it includes software components which can be used for free. Unlike all the previous solutions, it is designed for the Italian language.

---

[2] https://www.click4all.com/it/
[3] https://www.nuance.com/index.html
[4] http://www.mackichan.com/
[5] https://www.metroplexvoice.com/
[6] https://www.texthelp.com/en-us/products/equatio/

SpeechMatE

System Description

SpeechMatE consists of five main functional components: a speech recognizer, a parser, a LaTeX editor, a LaTeX compiler, and a PDF viewer supporting hot reload when open files change. The user interacts with the system through speech commands, or the dictation of text or maths expressions. Speech flow is processed by the speech recognizer which generates the corresponding text in the Italian language. This text is parsed by the parser that converts mathematical expressions into LaTeX and links speech commands to a corresponding script in the editor. Afterwards, the result of the parser is sent to the editor which updates the PDF document or executes the script.

Two videos[7] show how *SpeechMatE* works.

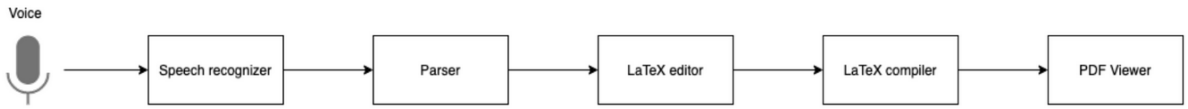

Fig. 1: A high level representation of the components of SpeechMatE

Design Challenges

The design process has gone through three main challenges: (i) the design of a speech interaction model which defines how to dictate text and maths in the Italian language so that they can be parsed and properly input and edited in a document; (ii) the selection of a maths editor which can enable the editing and displaying of maths in real time while writing; (iii) the identification of an automatic speech recognition service that is reliable for the Italian language and which can be used for free.

The Speech Interaction Model

SpeechMatE enables full speech interaction, which means that all the operations available in *SpeechMatE* must be enabled by speech. To this purpose, three environments are defined: the text environment, wherein speech flow is recognized as text according to the rules of the Italian language, the maths environment, wherein only mathematics can be dictated and the commands environment for entering all the editor commands (e.g. editing, open, save). Three reserved words (i.e. testo, matematica, comandi) are used to activate the desired environment. So far, this work has defined the maths environment for arithmetic and elementary algebra. There is not a reference book for spoken mathematics in Italian, so the rules to read maths expressions in Italian are defined by localizing the rules for English introduced in the Handbook for spoken mathematics [4]. Mathematical expressions that can be dictated are defined as sequences of:

Atomic symbols including for example Latin letters, Greek letters, numbers, parentheses

Operators that are applied to one or more mathematical expressions that can be separated by reserved words and ended by the stop word fine.

Here are some examples of expressions (in Italian) which can be dictated with *SpeechMatE*: *"aperta quadra 12 più aperta tonda 13 meno 4 chiusa tonda diviso 3 chiusa quadra diviso 4"* "open square bracket 12 plus open round bracket 13 minus 4 close round bracket divided by 3 close square bracket divided by 4"

$$[12 + (13 - 4) \div 3] \div 4 \qquad \textit{Expression 1}$$

[7] http://www.integr-abile.unito.it/icchp2020/frazione.mp4 and http://www.integr-abile.unito.it/icchp2020/equazione.mp4

"frazione numeratore 2 più 4 denominatore 5 fine meno 6" "fraction numerator 2 plus 4 denominator 5 end minus 6"

$$\frac{2+4}{5}-6$$

Expression 2

"frazione numeratore x alla 4 fine meno 1 denominatore frazione numeratore x più 1 denominatore x meno 1 fine meno 3 fine" "fraction numerator x to the power of 4 end minus 1 denominator fraction numerator x plus 1 denominator x minus 1 end minus 3 end"

$$\frac{x^4-1}{\frac{x+1}{x-1}-3}$$

Expression 3

The Maths Editor

The functional role of the maths editor is to enable entering text and mathematics as well as the editing and the displaying of mathematical expressions in visual notation while typing. Initially, the features of the editor were defined while typing. Initially, the features of the editor were defined, then a number of editors were compared in order to choose the most adequate one. First, since LaTeX is the most complete system to compose scientific documents, in view of future extensions, the editor must include tools to assist the writing LaTeX documents (e.g., syntax highlighting). Second, in order to automate operations through speech commands, the editor must include a scripting language. Third, for the expressions to be displayed in visual notation in real time, the editor must be interfaced with an external LaTeX compiler, which in turn is interfaced with a PDF viewer with hot reload capabilities.

A number of LaTeX editors were compared to select the most adequate one, including, in particular Glimpsee [7], Overleaf[8] and TeXStudio[9]. In the end, TeXStudio was chosen for its being an open source, cross-platform and fully-featured LaTeX editor. Potentially, it can be extended through QTScript[10], a JavaScript-like scripting language and it can be interfaced with a LaTeX compiler (e.g. latexmk on Mac OS or LaTeXdaemon for Windows), which is in turn interfaced with a PDF viewer (e.g. Skim for Mac OS or SumatraPDF for Windows). Every time the document is saved, the PDF document is updated and viewed.

Automatic Speech Recognition for Italian

Many different speech recognition systems were evaluated with respect to the ability of recognizing the Italian language. They can be divided in two groups: online and offline recognizers. For what concerns offline recognizers, Sphinx, Kaldi and DeepSpeech were analysed and it emerged that, at the present stage of development, they are not accurate enough for recognizing Italian language.

Among online recognizers, Google Cloud Service[11], Microsoft Bing[12], Houndify[13] and IBM Watson[14] are commercial solutions that recognize Italian very precisely. Nonetheless, the pricing scheme based on the number of requests or on the number of processed minutes leads quickly to high costs when applied to continuous speech recognition.

[8] www.overleaf.com
[9] www.texstudio.org
[10] doc.qt.io/archives/qsa-1.1.2/language.html
[11] cloud.google.com/speech-to-text/
[12] azure.microsoft.com/en-us/pricing/details/cognitive-services/speech-services/
[13] www.houndify.com
[14] www.ibm.com

In the end, Google Cloud Service was chosen, which is available for free with some restrictions through Google Chrome browser.

Implementation

A JavaScript application listens to the continuous speech input that is sent to the recognizer via Web Speech API[15]. The recognizer transforms speech into text based on grammar rules specified in JSpeech Grammar Format[16] (JSGF). These rules force the recognizer to properly recognize Latin letters, Greek letters and numbers. The text generated is sent to a multi-thread parser that transforms the maths expressions into LaTeX in real time. The output of the parser is sent to the document in TeXStudio. If a command is sent, TeXStudio will run a script. If text or LaTeX is sent, TeXStudio will save the document and the LaTeX compiler will generate the PDF document to be displayed by the PDF viewer.

Conclusions and Future Work

This work has introduced *SpeechMatE*, a prototype system which can be used by MIP to type maths through speech input in the Italian language. This prototype is the starting point for future development along four main directions: (i) adoption of an open source speech recognizer which properly recognizes the Italian language; (ii) extension to further mathematical notations including in particular trigonometry, calculus and set theory; (iii) design and development of speech- driven navigation and editing of maths expressions; (iv) usability evaluation with MIP.

References

1. S. N. Batlouni, H. S. Karaki, F. A. Zaraket, and F. N. Karameh. Mathifier - speech recognition of math equations. In 18th International Conference on Electronics, Circuits, and Systems, pages 301–304. IEEE, 2011.
2. Bernareggi C. (2010) Non-sequential Mathematical Notations in the LAMBDA System. In: Miesenberger K., Klaus J., Zagler W., Karshmer A. (eds) Computers Helping People with Special Needs. ICCHP 2010. Lecture Notes in Computer Science, vol 6180. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-14100-3_58
3. Bernareggi C., Brigatti V. (2008) Writing Mathematics by Speech: A Case Study for Visually Impaired. In: Miesenberger K., Klaus J., Zagler W., Karshmer A. (eds) Computers Helping People with Special Needs. ICCHP 2008. Lecture Notes in Computer Science, vol 5105. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-70540-6_131
4. L. A. Chang et al. Handbook for Spoken Mathematics:(Larry's Speakeasy). ERIC, 1983.
5. K.-S. Choi and T.-Y. Chan. Facilitating mathematics learning for students with upper extremity disabilities using touch-input system. Disability and Rehabilitation, 10(2):170–180, 2015.
6. F. Curatelli and C. Martinengo. Design criteria for educational tools to overcome mathematics learning difficulties. Procedia Computer Science, 15:92–102, 2012.
7. P. Dragicevic, S. Huot, and F. Chevalier. Gliimpse: Animating from markup code to rendered documents and vice versa. In 24th annual ACM symposium on User interface software and technology, pages 257–262, 2011.
8. C. Elliott and J. Bilmes. Computer based mathematics using continuous speech recognition. Vocal Interaction in Assistive Technologies, Games and More, 2007.

[15] www.w3.org/community/speech-api
[16] www.w3.org/TR/2000/NOTE-jsgf-20000605/

9. J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard. A fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display. In Workshop on Communication by Gaze Interaction, pages 1–5, 2018.
10. Javanovic R., MacKenzie I.S. (2010) MarkerMouse: Mouse Cursor Control Using a Head-Mounted Marker. In: Miesenberger K., Klaus J., Zagler W., Karshmer A. (eds) Computers Helping People with Special Needs. ICCHP 2010. Lecture Notes in Computer Science, vol 6180. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-14100-3_9
11. K. Lyons, D. Plaisted, and T. Starner. Expert chording text entry on the twiddler one-handed keyboard. In 8th International Symposium on Wearable Computers, volume 1, pages 94–101. IEEE, 2004.
12. S. MacKenzie. The one-key challenge: searching for a fast one-key text entry method. In 11th international ACM SIGACCESS conference on Computers and accessibility, pages 91–98, 2009.
13. S. MacKenzie and B. Ashtiani. Blinkwrite: efficient text entry using eye blinks. UAIS, 10(1):69–80, 2011.
14. S. Mackenzie and T. Felzer. Sak: Scanning ambiguous keyboard for efficient one-key text entry. TOCHI, 17(3):1–39, 2010.
15. P. Majaranta and K.-J. Räihä. Text entry by gaze: Utilizing eye-tracking. Text entry systems, pages 175–187, 2007. 8 Author et al.
16. C. Martinengo and F. Curatelli. Educational methodologies and ict to help learning in motor-impaired children. Italian Journal of Educational Technology, 19(1):36–42, 2011.
17. W. H. Organization et al. World report on disability 2011. World Health Organization, 2011.
18. D. Pedrosa and M. d. G. C. Pimentel. Text entry using a foot for severely motor-impaired individuals. In 29th Annual ACM Symposium on Applied Computing, pages 957–963, 2014.
19. C. G. Pinheiro, E. L. Naves, P. Pino, E. Losson, A. O. Andrade, and G. Bourhis. Alternative communication systems for people with severe motor disabilities: a survey. Biomedical engineering online, 10(1):31, 2011.
20. O. Polacek, T. Grill, and M. Tscheligi. Nosetapping: What else can you do with your nose? In 12th International Conference on Mobile and Ubiquitous Multimedia, MUM '13, New York, USA, 2013. ACM.
21. O. Polacek, A. J. Sporka, and P. Slavik. Text input for motor-impaired people. UAIS, 16(1):51–72, 2017.
22. S. Trewin and J. Arnott. Text entry when movement is impaired. Text Entry Systems, pages 289–304, 2007.
23. W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel. Sphinx-4: A flexible open source framework for speech recognition, 2004.
24. A. Wigmore, G. Hunter, E. Pfl ̈ugel, J. Denholm-Price, and V. Binelli. Using automatic speech recognition to dictate mathematical expressions: The development of the "talkmaths" application at kingston university. Journal of Computers in Mathematics and Science Teaching, 28(2):177–189, 2009.
25. M. Wigmore, E. Pflugel, G. J. Hunter, J. Denholm-Price, and M. Colbert. Talkmaths better! evaluating and improving an intelligent interface for creating and editing mathematical text. In 6th International Conference on Intelligent Environments, pages 307–310. IEEE, 2010.
26. J. O. Wobbrock, J. Rubinstein, M. Sawyer, and A. T. Duchowski. Not typing but writing: Eye-based text entry using letter-like gestures. In COGAIN, pages 61–64, 2007.