

日本語音声認識と GUI によるマルチモーダル数式入力方式の試作

石丸 琴海[†] 妹尾 美侑[†] 白井 詩沙香[‡] 福井哲夫[†]武庫川女子大学 生活環境学部 情報メディア学科[†] 大阪大学 サイバーメディアセンター[‡]

1. はじめに

GIGA スクール構想により，初等中等教育における端末整備が進められるなか，デジタル教科書をはじめとするデジタル教材の活用が高まっている。文部科学省はデジタル教科書の利点の一つに，音声読み上げ機能を挙げており[1]，多様な学習・ニーズに対応できることが期待される。一方，近年推進されている STEAM 教育では概念理解のために数式が必要となるが，数式のデジタル的な取り扱いは複雑であり，生徒が入力・編集するには煩わしく，難しいのが現状である。我々はこれまで数式入力の負担を軽減する目的で，曖昧な数式用文字列を AI が正しい数式に予測し，キーボード・GUI ベースで構築できる数式入力インタフェース MathTOUCH を開発してきた。また，多様な入力方法に対応するために，音声による数式入力の実現に向けて，拡張を進めてきた[2]。音声による数式入力が可能になれば，数式入力の取り扱いが容易になり，生徒が STEAM 教育を受けやすくなると期待できる。

本研究では，音声認識と GUI タッチ操作などを組み合わせたマルチモーダルな数式入力方式を試作した。数式は複数の英数字・特殊記号で構成されており，音声だけで数式を確定するのは困難である。そこで我々は GUI と併用した修正機能によって補う方式を提案する。試作では中学校レベルを対象としている。

2. 音声による数式入力の試作システム

日本語音声認識技術について，W3C が仕様策定している Web Speech API は無料で利用でき，ブラウザで扱いやすいことから本研究ではこれを採用した。2019 年に白井・福井[2]によって，簡単な多項式の認識に実験的に成功した。しかし，入力できる数式分野を拡大する必要があり，実用性のためには，音声による曖昧性や誤認の修正機能が必要である。そこで本研究では，MathTOUCH 既存のキーボードおよび GUI ベースのタッチ入力機能に音声入力機能を加える形で実現する。本研究の対象は中学校で扱う数式レベルまでとし，「等式」「不等式」「比例式」「分数」「平方根」「多項式」「方程式」「関数」「直角三角形の角度」「球の表面積・体積」の基本数式を音声入力可能にする。

Prototype of Multimodal Math Input Method by Japanese Speech Recognition and GUI

[†] Kotomi ISHIMARU, Miyu SENO, Tetsuo FUKUI
Mukogawa Women's University

[‡] Shizuka SHIRAI
Osaka University

2.1 試作システムの概要

実際の使い方は簡単で図 1 のように MathTOUCH を呼び出し，UI パレットの「音声認識開始ボタン」(図1-①)を押し，数式を発音すれば，線形文字列入力エリアに入力され，従来の MathTOUCH と同様に AI が候補を提示するので，所望の数式を選択して数式を構築できる。従来のハード/ソフト(GUI)キーボード操作も使用可能なため，音声誤認や音声入力ミスにも簡単に修正・編集操作が可能である。

2.2 方法

日本語認識エンジンをそのまま利用し，数式の読み上げ音声を一旦日本語として認識させ，その認識文字列を数式線形文字列に変換し，MathTOUCH の数式予測 AI に連携出力する方法をとる。

MathTOUCH による入力は図1のように①音声またはキーボード UI を使い数式線形文字列を入力し，②AI によって提示された候補から所望の数式を選択・確定することによって，③数学アプリ(文書エディタや関数グラフツール)へ完成数式を挿入する。MathTOUCH の採用する数式線形文字列は数式を読む順番に要素のキーワードを並べるように定義されている。特殊記号やギリシャ文字などのキーワードは LaTeX に準拠しており，主にその頭文字だけで入力できるようになっている。特に x^2 は $x2$ となり，表記されない演算記号や括弧などは省略できる。線形文字列の詳細は文献[2]に示されている。この読む順番にキーを並べる特徴が，数式を読み上げた時の音声認識文字列と親和性が高く，重要着想である。

しかし，数式の音声認識は困難な課題が多く，音声の同音異義語に加え，アルファベットの a と e や c と t など似たような発音が多い上に，外部のノイズ音や人の喋り方・アクセント・間などの癖の違いによっても認識結果が異なるため，認識精度が下がる。

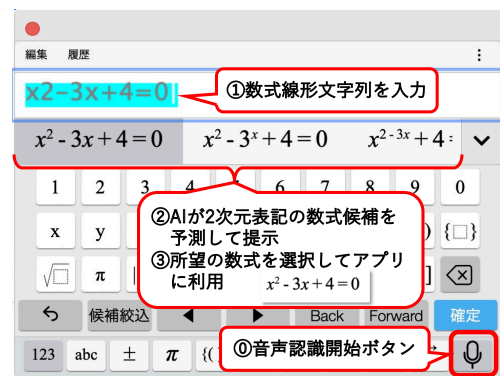


図 1 MathTOUCH の音声による数式入力例

そのためには、声だけで完全に数式を構築することは困難で、誤認識・エラー部分の修正のしやすさが重要となる。それゆえ、音声操作に加えて、ハード／ソフトキーボードを併用したマルチモーダルな入力方式を提案する。MathTOUCH を利用する利点は、(a)数式のかち入力が可能なこと、(b)途中数式部分のカーソル移動、追加・削除や操作を一つ前に戻す undo 機能など編集機能が充実していることである。従って音声入力においても効果を発揮できる。

2.3 数式の音声読み上げ認識データ

音声認識エンジンにはWeb Speech API を使用し、日本の生徒・学生が利用できるように日本語として音声読み上げを受け付け、MathTOUCH 形式の数式線形文字列へ変換する独自のフィルタを開発した(図2)。この変換フィルタではまず、API による音声認識結果の文字列を受け付け、Step1)~6)のように処理するよう設計した：

- Step 1) 前処理 (不要スペース削除・小文字に統一)
- Step 2) 漢字や特殊記号コードを「ひらがな」に統一
- Step 3) 演算子キーワードを線形文字キーに変換
- Step 4) ローマ字キーワードを対応ローマ字に置換
- Step 5) 「じょう」などの省略文字を削除
- Step 6) その線形文字列を MathTOUCH の UI 処理に接続

このために数式用の変換辞書テーブルを作成し、処理にはキーワード文字数の長い順に置換するようにした。数式入力の検証結果の一部を表1に示す。括弧に関しては、今回は明示的に「かっこことじ」と発音するルールを採用した。テスト数式 55 件の内、2 件だけがどうしても正しく入力できなかった。特に注意を要する点は、表1の比例式の例のように「えーたいびー」が「永代 B」と認識され、日本語特有の問題で、アルファベットと記号の組み合わせが思わぬ熟語に認識されるケースがいくつか見受けられた。本システムではこのような場合も変換テーブルに追加し対処した。

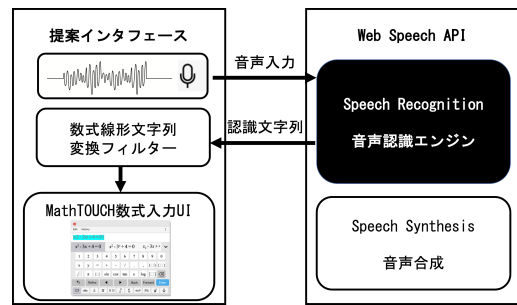


図2 試作システムの構成

3. まとめ

本研究の新規性は以下のようになる。

- MathTOUCH 数式線形文字列と音声認識文字列との親和性から数式音声認識のローコスト実現
- マルチモーダル方式による修正のしやすさ
- 中学レベルの数式に対応
- 数学文書エディタやグラフ作成ツールの数式入力UIとして実装し、実用化に成功

提案システムの課題としては、アルファベットの大文字と小文字の区別ができないことや、c と 4 など極めて判別が困難な英数字の問題が残された。これらは今後の課題としたい。

謝辞

本研究は JSPS 科研費 19H01733 および 20K12117 の助成を受けたものである。

参考文献

- [1] 文部科学省公式サイト：「3. デジタル教科書の制度化について」, http://www.mext.go.jp/a_menu/shotou/kyoukasho/gaiyou/04060901/1349317.htm (2018. 10. 11 参照).
- [2] 白井詩沙香, 福井哲夫：「数式曖昧入力変換方式による音声入力機能の開発」, 教育学習支援情報システム研究報告, 2019-CLE-27, Vol. 19, pp. 1-4, 2019.

表1 数式の音声入力検証結果の例

分類	数式	音声	Web Speech APIによる音声認識文字列	変換フィルタによる数式線形文字列
等式	$a=b$	えーいこーるびー	a=b	a=b
不等式	$a>b$	えーだいなりびー	A > ビー	a>b
比例式	$a:b$	えーたいびー	永代B	a:b
分数	$\frac{1}{2}$	にぶんのいち	2分の1	2\1
平方根	$\sqrt{2}$	るーとに	ルート2	root2
	$\sqrt{a^2b}$	るーとえーにじょうびー	\sqrt{A} 2乗B	roota2b
多項式	$x^2 - a^2$	えつくすにじょうまいなすえーにじょう	エックス2乗- A 2乗	x2-a2
	$(x-a)^2$	かっこえつくすまいなすえーかっこことじにじょう	(X - A) 2乗	(x-a)2
	$(x-a)(x-b)=0$	かっこえつくすまいなすえーかっこことじかっこえつくすまいなすびーかっこことじいこーるせろ	(X - A) (エックス- B) = 0	(x-a) (x-b)=0
方程式	$x=a, b$	えつくすいこーるえーこんまびー	エックス= A, B	x=a, b
	$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$	まいなすびーぶらすまいなするーとびーにじょうまいなすよんえーしーわるにえー	マイナスビー± √ ビー2乗マイナス4 AC ÷ 2えー	$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
直角三角形	$\angle a = 90^\circ$	かくえーいこーるきゅうじゅうど	角A = 90度	<a=90°
球の体積	$V = \frac{4}{3}\pi r^3$	ぶいいいこーるさんぶんのよんばいあーるさんじょう	V = 3分の4 πアール3乗	$v = \frac{4}{3} \pi r^3$