

音声認識による数式入力UIの開発と課題

武庫川女子大学・社会情報学部 福井 哲夫

Tetsuo Fukui, School of Social Informatics, Mukogawa Women's University

大阪大学・サイバーメディアセンター 白井 詩沙香

Shizuka Shirai, Cybermedia Center, Osaka University

1 はじめに

2020年に文部科学省が提唱したGIGAスクール構想[1]により、生徒一人に一台の端末環境が整うようになってきた。また、デジタル教科書が全面導入され、紙の教科書にはない付加価値である、デジタルならではの機能を活かす試みが広がり始めている。例えば、教科書の一部を音声で解説するなどが挙げられている[2]。

しかし、理数系教育においては、デジタル端末上で、生徒が数式を直接入力編集する環境が統一されておらず、操作も面倒なものが多く、課題とされている。

それを改善するために、福井は、数式を曖昧な文字列で表し、かな漢字変換のように数式候補を提示して構築するような、新たな数式入力方式を2012年に提案した[3, 4]。2015年には、生徒による提案システムを使った数式入力におけるパフォーマンス実験において、CAS文法やLaTeX形式によるテキストベースの入力方式に比べて1.2倍、GUIベースの従来方式と比べて1.6倍効率よく入力でき、満足度も高いことが示された[5]。その後、AI技術を導入し、数式候補をAIが予測するようインテリジェント化を実現した[6]。

しかし、現状の提案システムは主にキーボードベースの入力方式であり、タブレット端末用の独自のソフトキーボードにも対応した[7, 8]が、小学校や中学校などのITスキルが十分でない初学者にとっては向いていないと考えられる。

そこで、本研究のねらいは、多様な入力方式の選択肢を実現するため、音声による数式入力を行うことにより、数式入力の取扱いを容易にすることで生徒がSTEAM教育を受けやすい環境を構築することである。2019年には、音声による簡単な数式の入力システムの開発を試みた[9]。数式の音声認識を発展させるためには、日本語による数式の読み上げルールを定める必要があり、先行して、音声合成技術による数式の音声読み上げに関する研究に取り組んだ[10]。2023年3月には、中学校数学のレベルの数式まで認識できるようになった[11]。

しかし、認識精度に問題があり、中等教育に活用するためには、さらに高校数学で扱う数式まで範囲を広げる必要がある。そこで、本研究の目的は、高校数学I,Aレベルに拡張し、その際の課題を報告し、解決に向けた提案を行うことにある。

2 先行研究と課題

ここでは、数式のキーボード入力による新たな方式を使った改善の取り組みと、音声入力へ発展させるための先行研究の取り組み、および残された課題について解説する。

2.1 数式のキーボード入力改善の取り組み

第1章で述べたように、我々はこれまでにインテリジェントな数式入力インタフェース MathTOUCH を開発してきた。提案の入力方式は、数式を曖昧な文字列で表し、AI が数式候補を提示して構築するもので、以下では「数式線形文字列変換方式」と呼ぶ。本来数式は $\frac{1}{2}$ や x^2 のように2次元的に表記されるが、我々はこれを独自に定めた「数式線形文字列」で表す。「数式線形文字列」のルールはキーボードに無い特殊記号の場合は $\text{\LaTeX}$ に準拠したキーワードを割り当て、数式表記を読む順番に、その構成するキーワード文字を並べるもので、 $\frac{1}{2}$ の場合は、英語式の “1/2” の他に、「2 分の 1」と読む場合を想定して “2\1” と表す。また、 x^2 の場合は “x2” と表す。このように、分数命令や指数命令の表記されない区切りのためのカッコや演算記号は入力しなくてもよいようになっている。容易に想像できるように、このルールでは数式表現が曖昧で、所望の数式表記は一意的に決まらない。そこで、AI に優先順位の高い数式候補を複数提示し、所望の数式を選んで構築できる方式を採用した。

この AI は高校数学の教科書 [14] に現れる約 4000 個の数式を学習しており、精度テストでは候補のトップ 10 位以内に正解式が含まれる割合が約 85 % となっている [6]。

本研究の重要な着想の一つは、この「数式線形文字列」と数式の音声読み上げ言葉との親和性が非常に高いことにあり、既存の音声認識エンジンと「数式線形文字列変換方式」が容易に連携できることが想定される。これについては第4章で報告する。

2.2 数式の音声入力への取り組み

2019 年に「数式曖昧入力変換方式による音声入力機能の開発（白井・福井）」[9]において、音声認識エンジンに WebSpeechAPI を採用し、簡単な多項式の認識に成功した。

しかし、日本語による数式の音声認識には、数式の音声読み上げルールが統一されていないという問題がある。そこで、数式の音声読み上げルールを検討するため、音声合成技術を使った数式の音声読み上げシステムの開発に取り組んだ [10]。本研究で定めた数式の音声読み上げルールの詳細については、3.1 節で報告する。

それを踏まえて音声認識改善に取り組み、2023 年 3 月の情報処理学会全国大会にて「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・妹尾・白井・福井）[11]において、音声認識のための辞書テーブルを中学校数学レベルに対応できるよう拡張し、55 パターンの数式に対する認識動作の確認を行なった。その結果、評価サンプル式の 96 % は認識に成功した。

本研究の中学校数学レベルとは、「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。具体例は 4.2 節で解説する。

しかし、複数の英数字・特殊記号で構成されている数式は、音声だけで数式を確定するのは困難であることがうかがえる。例えば、a（エー）とe（イー）やc（シー）と4（し）などの似たような発音が多く、判別は非常に困難であった。その上に、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違いため、認識精度が下がる。したがって、100 %の認識率を目指すには限界がある。

むしろ、実用的 UI を目指すためには、音声のみによる入力ではなく、キーボードやソフトキーボードタッチなど、入力手段の選択肢を増やしたマルチモーダル化を進め、ユーザの状況に応じて切り替えられる、あるいは音声入力操作の一部を補えるようにする必要がある。

3 方法

音声認識のための辞書テーブルの対応できる数学分野の情報を拡大していけば、高校数学レベルの入力は可能になると思われる。これについては、第4章で報告する。

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。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。

3.1 数式読み上げルール

本研究における数式認識のための音声読み上げルールに関しては、可能な限り教科書通りの自然な読み方に対応できるようにする。英語の数式の読み上げルールはある程度文献によって統一されている [12, 13]。一方、日本語による数式の読み上げルールは確定したものがないのが実情である。特に、括弧の始まりと終わりや絶対値記号の始まりと終わりに関しては通常、始まりは発音しても終わりを省略することが多いが、認識精度を高めるために、終わり部分を明示的に発音するルールを独自に採用した。具体的には特殊記号も含め、表1のように定める。本研究で扱う数学の範囲を中学数学と高校数学 I, A とし、表1の2列目が扱う数学記号で、3列目が対応する読み上げ言葉である。外来語はカタカナで表記している。また、表1の3列目、4列目にそれぞれ記号の使用例とその読み上げ例を示した。

4 開発システム

4.1 音声認識による数式入力 UI 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（図1-②）、従来の MathTOUCH と同様に AI が候補を提示する（図1-③）ので、所望の数式を選択して数式を構築できる（図1-④）。

表 1: 数式読み上げルール

No.	数式記号	音声読み上げ文字列	数式例	読み上げ例
1	=	イコール	$x = 2$	エックス イコール 2
2	+	プラス	$x + 2$	エックス プラス 2
3	-	マイナス	$x - 2$	エックス マイナス 2
4	±	プラスマイナス	$x \pm 2$	エックス プラスマイナス 2
5	×	かける	$x \times 2$	エックス かける 2
6	·	かける	$x \cdot 2$	エックス かける 2
7	/	わる, スラッシュ	$x/2$	エックス わる 2
8	÷	わる	$x \div 2$	エックス わる 2
9	分数	ぶんの	$\frac{1}{2}$	2 ぶんの 1
10	べき乗	じょう	x^2	エックス の 2 じょう
11	>	だいな	$x > 2$	エックス だいな 2
12	≥	だいなイコール	$x \geq 2$	エックス だいなイコール 2
13	<	しょうな	$x < 2$	エックス しょうな 2
14	≤	しょうなイコール	$x \leq 2$	エックス しょうなイコール 2
15	√	ルート	$\sqrt{2}$	ルート 2
16	:	たい	$a : b$	エー たい बी
17		へいこう	$a \parallel b$	エー へいこう बी
18	,	コンマ	a, b	エー コンマ बी
19	(	かっこ	(x)	かっこ エックス かっこことじる
20	)	かっこことじる		
21	{	ちゅうかっこ	$\{x\}$	ちゅうかっこ エックス ちゅうかっこことじる
22	}	ちゅうかっこことじる		
23	[	だいかっこ	$[x]$	だいかっこ エックス だいかっこことじる
24	]	だいかっこことじる		
25	'	ダッシュ	x'	エックス ダッシュ
26	∠	かく	$\angle a$	かく エー
27	°	ど	60°	60 ど
28	π	パイ	πr^2	パイ アール 2 じょう
29	sin	サイン	$y = \sin x$	ワイ イコール サイン エックス
30	cos	コサイン	$y = \cos x$	ワイイコール コサイン エックス
31	tan	タンジェント	$y = \tan x$	ワイ イコール タンジェント エックス
32	∩	かつ	$A \cap B$	エー かつ बी
33	∪	または	$A \cup B$	エー または बी
34	!	かいじょう	$n!$	エヌのかいじょう
35		ぜったいち	$ x $	ぜったいち エックス ぜったいちことじる
36	→	ぜったいちことじる		
37		ベクトル	$\vec{a}$	ベクトル エー
38			$\vec{0}$	ゼロベクトル

ただし、図 1-①において、音声誤認や音声入力ミスによって、読み上げた数式が正しい線形文字列に変換されるとは限らない。その場合は、誤認部分の文字列を削除して、再度音声入力を繰り返す。場合によっては、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、キー入力に切り替えて入力を継続したり、キー操作により簡単に修正・編集が可能である。このことが本研究のもう一つの提案でもある。

4.2 システム設計

開発したシステムの構成を図 2 に示す。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付け、Step 1.~6. のように処理するよう設計した：

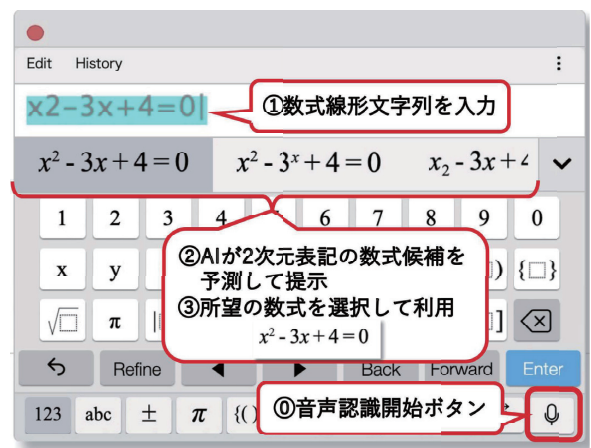


図 1: MathTOUCH の音声入力手順

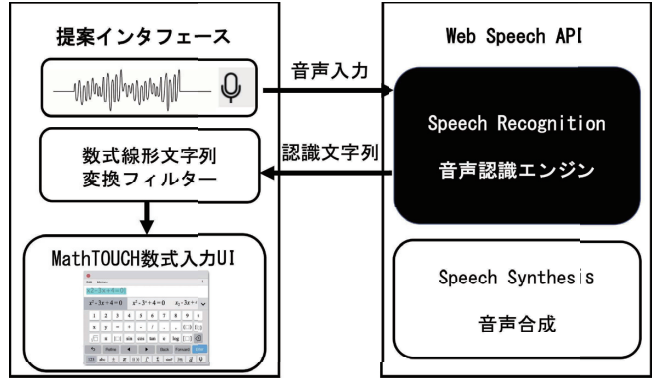


図 2: システムの構成

- 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 2 に示す。括弧に関しては、3.1

表 2: 中学数学数式用の変換フィルターの検証例

No.	数式例	音声読み上げ文字列	API による音声認識結果	線形文字列
1	$a = b$	エー イコール ビー	A=B	a=b
2	$a > b$	エー だいなり ビー	A > ビー	a>b
3	$a : b$	エー たい ビー	永代ビー	a:b
4	$\frac{1}{2}$	2 ぶんの 1	2 分の 1	2\1
5	$\sqrt{2}$	ルート 2	$\sqrt{2}$	root2
6	$x^2 - a^2$	エックスにじょう マイナス エー にじょう	エックス 2 乗 - A 2 乗	x2-a2
7	$(x - a)^2$	かっこエックス マイナス エー かっこことじ にじょう	(エックス - A) 2 乗	(x-a)2
8	$(x - a)(x - b) = 0$	かっこエックス マイナス エー かっこことじかっこエックス マイナス ビーかっこことじ イコールぜろ	(エックス - A) (エックス - B) = 0	(x-a)(x-b)=0
9	$x = a, b$	エックス イコール エー こんま ビー	エックス = a, ビー	x=a,b
10	$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$	マイナス ビー プラスマイナス ルート ビーにじょうマイナス よんエー シーわるにエー	マイナスビー± √ ビー 2 乗 マイナス 4 AC ÷ 2 えー	x=-b+-rootb2-4ac/2a
11	$\angle A = 90^\circ$	かくエー イコールきゅうじゅうど	角 A = 90 度	<a=90'
12	$V = \frac{4}{3}\pi r^3$	バイ イコールさんぶんのよん パイ アールさんじょう	V = 3 分の 4 πアール 3 乗	v=3\4pr3

節で示したように明示的に「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 2 の比例式の例のように「エー たい ビー」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。

さらに、本研究では新たに高校数学 I,A のために変換辞書テーブルを作成し、検証した。その検証例を表 3 に示す。6 番目の 6 のみ小数点表示に認識されうまく入力できていない。それ以外の式は線形文字列への変換は成功している。

5 まとめと今後の課題

本研究で実現できたことを列記すると：

- 1. 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システムのローコスト開発を実現した。
- 2. マルチモーダル方式によって、誤認識時の修正操作のしやすさを提案。
- 3. 中学数学から高校数学 I,A レベルの数式に対応した。

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。音声認識アルゴリズムも見直して、精度を向上させる必要がある。これらは今後の課題としたい。

謝辞

本研究は JSPS 科研費 23H01027 の助成を受けたものであり、加えて国際共同利用・共同研究拠点である京都大学数理解析研究所の支援を受けたものである。

表 3: 高校数学 I,A 数式用の変換フィルターの検証例

No.	数式例	音声読み上げ文字列	API による音声認識結果	線形文字列
1	$A \cup B$	エー または ビー	AA またはビー	acupb
2	$A \cap B$	エー かつ ビー	英勝ビー	acapb
3	$\equiv$	ごうどう	合同, ≡	=
4	$\sim$	そうじ	掃除	~
5	$\ni$	ニアリーイコール	ニアリーイコール, ニ	~=
6*	$\dot{6}$	どっとろく	.6	▲失敗 (正しくは dot6)
7	$5!$	ごのかいじょう	5 の階乗	5!
8	$ x $	ぜったいちエックス ぜったいちとじる	絶対値 X 絶対値 閉じる	[x]
9	$\tan \theta = \frac{a}{b}$	タンジェント シータ イコール ビーぶんのエー	タンジェント θ = B 分の A	tantheta = b \ a
10	$\tan 30^\circ = \frac{1}{\sqrt{3}}$	タンジェントさんじゅうど イコール ルート さんぶんのいち	タンジェント 30 度 = √ 3 分の 1	tan30'=root3 \ 1
11	$\sin \theta = \frac{a}{b}$	サイン シータ イコール ビーぶんのエー	サイン θ = ビー分の A	sintheta = b \ a
12	$\cos \theta = \frac{a}{b}$	コサイン シータ イコール ビーぶんのエー	コサイン θ = ビー分の A	costheta = b \ a
13	$12 \sin 30^\circ = 6$	じゅうにサインさん じゅうど イコールろく	12 サイン 30 度=6	12sin30'=6
14	$12 \cos 30^\circ = 6\sqrt{3}$	じゅうにコサインさんじゅうど イコールろく ルート さん	12 コサイン 30 度=6 √ 3	12cos30'=6root3
15	$\tan \theta = \frac{\sin \theta}{\cos \theta}$	タンジェント シータ イコール コサインシータぶんのサインシータ	タンジェント θ = コサイン θ 分のサイン θ	tantheta = costheta \ sintheta
16	$\sin^2 \theta + \cos^2 \theta = 1$	サインにじょうシータプラス コサインにじょうシータ イコールいち	サイン 2 乗 θ + コサイン 2 乗 θ = 1	sin2theta+cos2theta=1
17	$\sin (90^\circ - \theta) = \cos \theta$	サインかっこきゅうじゅうど マイナス シータかっことじる イコール コサイン シータ	サイン (90 度 - θ) =コサイン θ	sin(90'-theta)=costheta
18	$\sqrt{2} = 1.41421356 \dots$	ルートにイコールいってんよんいち よんにいちさんごろくてんでん	√ 2 = 1.41421356...	root2=1.41421356...
19	$ 2 = 2$	ぜったいちにぜったいち とじる イコールに	絶対値 2 絶対値閉じる = 2	[2] = 2
20	$ -3 = 3$	ぜったいちマイナスさん ぜったいちとじる イコールさん	絶対値 - 3 絶対値閉じる = 3	[-3]=3
21	$ 5-2 = 3$	ぜったいちごマイナスに ぜったいちとじる イコールさん	絶対値 5-2 絶対値閉じる=3	[5-2]=3
22	$ 2-\sqrt{5} = -2+\sqrt{5}$	ぜったいちにマイナスルート ご ぜったいちとじる イコール マイナスにプラスルート ご	絶対値 2 - √ 5 絶対値閉じる = - 2+ √ 5	[2-root5]=-2+root5
23	$\sqrt{(-3)^2} = -3 $	ルート かっこマイナスさんかっこ とじるにじょう イコールぜったいち マイナスさんぜったいちとじる	√ (- 3)2 乗= 絶対値 - 3 絶対値閉じる	root(-3)2=[-3]
24	$\sqrt{7+2\sqrt{10}}$	ルート ななプラスにルート じゅう さんかっけいエービーシーごうどう	√ 7 + 2 √ 10	root7+2root10
25	$\triangle ABC \equiv \triangle DEF$	さんかっけいディイーエフ	三角形 ABC ≡ △ディイーエフ	<abc=<def
26	$\triangle ABC \sim \triangle DEF$	さんかっけいエー ビー シーそうじ さんかっけいディイーエフ	三角形 ABC 相似 △ディイーエフ	<abc~<def
27	$\vec{a}$	ベクトルエー	ベクトルエー	veca
28	$\vec{0}$	ゼロベクトル	0 ベクトル	vec0

参考文献

- [1] 文部科学省：「GIGA スクール構想の実現について」，
https://www.mext.go.jp/a_menu/other/index_00001.htm (2022.9.15 参照)
- [2] 文部科学省：「学習者用デジタル教科書の効果的な活用の在り方等に関するガイドライン (改)」，<https://www.mext.go.jp/>，2021.
- [3] 福井哲夫：「数式のインテリジェントな線形入力方式」，京都大学数理解析研究所講究録，Vol. 1780，pp.160–171，2012.
- [4] 福井哲夫：「数式のインテリジェントな線形入力方式と評価」，日本数式処理学会数式処理，Vol. 18 No.2，pp.47–50，2012.
- [5] 白井詩沙香，仲村裕子，福井哲夫：「数式自動採点システムにおける数式入力インタフェースの提案と評価」，情報処理学会論文誌 教育とコンピュータ (TCE)，Vol. 1，No.3，pp.11–21，2015.
- [6] Fukui, T., Shirai, S.: 「Predictive Algorithm for Converting Linear Strings to General Mathematical Formulae」，Human Interface and the Management of Information: Supporting Learning, Decision-Making and Collaboration, Lecture Notes in Computer Science, Vol. 10274, Springer, pp.15–28, 2017.
- [7] 白井詩沙香，福井哲夫：「数式曖昧入力変換方式による数式入力インタフェースのスマートデバイス向け UI の試作」，第 43 回教育システム情報学会講演論文集，277–278，2018.
- [8] Shirai, S., Fukui, T.: 「Evaluation of Intelligent Input Interface for Entering Equations on Smartphone」，IUI '19: Proceedings of the 24th International Conference on Intelligent User Interfaces: ACM ,New York, pp.9–10, 2019.
- [9] 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.
- [10] 長谷川陽和，原知鈴，福井哲夫：「数学デジタル文書読み上げツールの開発」，第 20 回情報科学技術フォーラム講演論文集，音声言語処理 E-020，pp.1–2，2021.
- [11] 石丸琴海，妹尾美侑，白井詩沙香，福井哲夫：「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」，情報処理学会第 85 回全国大会報告，1W-03, pp.1–2，2023.
- [12] 銀林浩，銀林純：『基礎から分かる数・数式と図形の英語』，日興企画，1999.
- [13] 鶴沼仁：『知りたいことがすぐわかる数・式・記号の英語』，丸善出版，2003.
- [14] 俣野博，河野俊丈，他 27 名：『数学 I』，『数学 A』，『数学 II』，『数学 B』，『数学 III』，東京書籍，2012.