

音声と指差しの統合理解による マルチモーダル幾何問題解答システムの構築*

清原侃太[†]徳島大学大学院社会産業理工学研究部^{†‡}西村良太[‡]北岡教英[§]
豊橋技術科学大学情報・知能工学系[§]

1 はじめに

人は誰かとコミュニケーションをするとき、その時々に合わせて様々なモダリティを使用することで、自分の意図を相手に明確に伝えようとする。そのため、指示語などの曖昧な表現の意図を正確に伝えることは難しい。

本稿では、複数のモダリティによるマルチモーダル認識を用いて、機械が曖昧な表現を認識できるようにすることを考える。曖昧な表現が使用される場面として、幾何問題の解答を説明する過程に注目し、マルチモーダル認識が可能なシステムとして、音声と指差しを用いた幾何問題解答システムの構築と評価を行った。

2 関連研究

これまでに、音声と指差しを含めたジェスチャに関する多くの研究 [1][2] がなされているが、どの研究結果でも、ジェスチャを使用することでタスク達成に少なからず貢献できることが示されている。数学学習タスクにおける音声とジェスチャの関係についての調査 [3] では、数学の専門家ではない人ほどジェスチャを使用する割合が増加することを明らかにしている。また、ジェスチャを併用することで、学習者の認知負荷を軽減し、学習のパフォーマンスを向上させる可能性があることを示している。これらのことから、幾何問題解答タスクにおいても、マルチモーダルな解法を示すことで、学習効率の向上が期待できると考える。

3 システム構築

本章では、提案システムの構築手法について述べる。図1は、提案システムの構成図を示しており、幾何学問題作成段階と問題解答段階の2つの段階から構成されている。次節からは各段階について述べる。

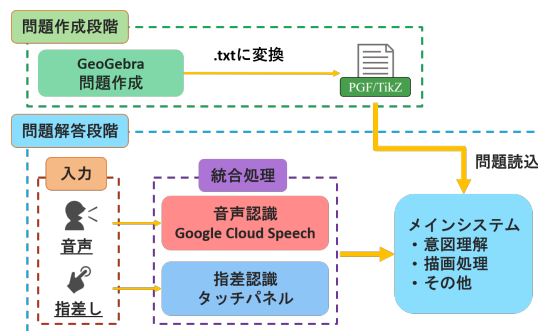


図1 システム構成図

3.1 幾何問題作成

まず、幾何問題作成段階では、動的数学フリーソフトウェアである GeoGebra¹ を使用して幾何問題の作成を行う。GeoGebra は簡単な操作で、幾何問題を含めた多種の問題を作成することができる。この GeoGebra を使用して幾何学問題を作成したものをテキストファイルに変換し、提案システム内で読み込ませることで、教員などのユーザが作成した独自の問題を解答することができる。

3.2 幾何問題解答

提案システムでは、幾何問題を解答するにあたり、ペンを使用して直接書き込むことはせず、音声と指差しを用いて解答していく。

音声認識器は、以前の研究 [4] では、Julius² を用いていた。Julius を使用するにあたり、正規文法を用いて認識できる語彙・文法を制限することで、認識精度を上げていたが、設定した文法しか認識できないため、ユーザが自由な解答を入力できないという問題があった。そこで本研究では、ユーザが自由な解答を入力できるように、リアルタイム音声を高精度かつ高速に認識することできる Google Cloud Speech-to-Text API³ を用いる。また、提案システムにおける指差しは、タッチパネルを用いたタッチ操作を使用している。

次に、音声と指差しの2つのモダリティを統合理解するにあたり、例えば、指示語の発話と対象の指差し動作（“ここ”と発声しながら指差すなど）がほぼ同時に行われると考えた。つまり、それぞれの入力

*Construction of Geometric Problem Solving Assist System Based on Integrative Understanding of Speech and Finger Pointing

[†]Kanta Kiyohara, Graduate School of Technology, Industrial and Social Sciences, Tokushima University

[‡]Ryota Nishimura, Graduate School of Technology, Industrial and Social Sciences, Tokushima University

[§]Norihide Kitaoka, Department of Computer Science and Engineering, Toyohashi University of Technology

¹<https://www.geogebra.org/>

²<http://julius.osdn.jp/>

³<https://cloud.google.com/speech-to-text/?hl=ja>

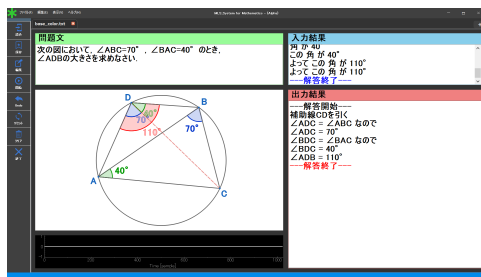


図2 システム GUI

時間的に近い対象同士を統合理解させることで、指示語のような音声のみでは機械が理解できない表現を認識することが可能になった。さらに、理解した内容に応じて、ユーザの解答を問題内に描画し、数式以外の入力も数式に変換して出力することができる。

最後に、システムの GUI を図 2 に示す。緑色の枠で囲まれたテキストボックスには問題文、その下に問題図形が表示され、ユーザは図形内の角や線などを指し示しながら音声入力（“ここに補助線を引いて”、“この角とこの角が等しいので”など）することができる。さらにその下の枠では、マイク入力の音声波形が表示される。青色の枠では、音声認識結果が表示され、赤色の枠では、統合理解した結果（“DC に補助線を引く”、“ $\angle ABC = \angle ADC$ ”など）が表示される。また、メニューバー、ツールバーを備えており、システムの設定や問題編集などの操作が行える。

4 評価実験

本実験により、以前のシステムと提案システムとの比較評価を行う。また、提案システムに対して不慣れた被験者（男性：5 名，年齢：21～23 歳）に協力してもらった。

4.1 実験手順

実験の手順としては、以下の通りである。

1. 解答する幾何学問題の解法が分かるかを確認。
2. 提案システムの操作方法を教示。
3. 手順 1 と同様の問題をシステムを用いて解答。
4. アンケートを実施し、システムの評価を行う。

4.2 実験結果

手順 4 で実施したアンケート（“1：いいえ”～“5：はい”とした 5 段階評価）の結果を図 3 に示す。実験結果から、以前のシステムよりも GUI に関しては高評価が得られたが、全体的な評価は以前よりも下回る結果となった。原因としては、被験者が不慣れたシステムの操作のコツを掴むまで時間がかかったことや、音声認識の誤認識（“B”↔“D”，“角と”↔“角度”など）によるシステムが想定していない入力をしてしまったことなどが挙げられる。しかし、入力文法を制限していないにも関わらず、評価は大きく下がらな

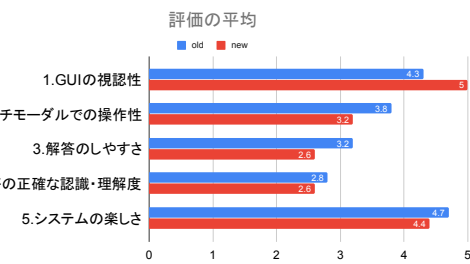


図3 以前のシステムとの比較結果

かったことから、以前のシステムよりも、詳細なマルチモーダル使用の分析が可能であるといえる。さらに、誤認識の種類を特定し、認識結果の誤認識箇所を補正するような仕組みを導入できれば更なるユーザビリティの向上が期待できる。

5 おわりに

本稿では、音声と指差しの2つのモダリティを入力として扱い、それぞれのモダリティを統合理解することで、幾何問題を解答することができるシステムを構築した。評価実験により、以前の研究結果と比較して、GUI に関しては高評価であったが、特に音声認識のユーザビリティに関しては低評価となってしまった。しかし、入力文法を制限しないことで、ユーザが自由な解答を行えるため、音声と指差しの関係性についての調査を行うことに適したシステムといえる。

今後の課題として、誤認識を補正するような仕組みを実装し入力理解能力を高めることで、より自然な入力に対応できるようにすることである。また、音声と指差しの関係性についての調査を行うための実験を行うことも考えている。

参考文献

- [1] J. E. Driskell et al., “The Effect of Gesture on Speech Production and Comprehension”, Human Factors, Vol.45, No.3, Fall 2003.
- [2] S. Oviatt et al., “When do we interact multimodally?: Cognitive load and multimodal communication patterns”, In Proceedings of the 6th International Conference on Multimodal Interfaces ACM, N.Y., 129-136, 2004.
- [3] A. Sriramulu et al., “Dynamic Adaptive Gesturing Predicts Domain Expertise in Mathematics”, 2019 International Conference on Multimodal Interaction, 105-113, October 2019.
- [4] 清原侃太, 西村良太, 北岡教英, “幾何問題における音声と指差しを用いた学習支援システムの構築とその評価”, 日本音響学会研究発表会講演論文集, Vol.2-P-17, 989-992, 2019 年 3 月.