

聴覚情報のみによる 理解しやすいマイクロタスクデザイン

鍾 穎^{1,a)} 松原 正樹^{1,b)} 小林 真^{2,c)} 森嶋 厚行^{1,d)}

概要: マイクロタスク型クラウドソーシングは、これまで解決できなかった問題を解決する方法を提供し、人々の働き方を広げる可能性をもつ。また聴覚情報のみによるマイクロタスクは、クラウドワーカが画面を使用できない状況でも実行できる。本論文では、聴覚情報のみによるマイクロタスクでの理解容易性について調査を行い、理解しやすいタスクデザイン方法を提案する。理解しやすいタスクには、軸の一貫性が重要な要素であることを示し、一貫性のある聴覚情報のみによるマイクロタスクを開発するフレームワークを提案する。クラウドワーカや視覚障害者を対象とした実験により、軸の一貫性があるマイクロタスクの方が軸の一貫性のないマイクロタスクよりも優れた結果を得られることを示した。

キーワード: クラウドソーシング, マイクロタスクデザイン, 軸の一貫性

Understandable microtask design with only auditory information

Abstract: Microtasks expand ways for people to work, which we could not imagine in the past. When people have pockets of time, they can perform microtasks. This paper pursues this approach further by exploring the design of microtasks that interact with workers with audio and physical means only, with only auditory information. Such a task can be performed in situations where workers cannot use display devices. This paper shows that consistency in navigation is an important factor and proposes a principled framework that develops consistent non-visual microtasks. The experimental result with real-world workers and visually impaired people shows that the resultant task design allows them to produce better results than tasks without consistency.

Keywords: Crowdsourcing, Microtask design, Consistency

1. はじめに

クラウドソーシングはこれまで解決できなかった問題を解決する方法を提供し、人々の働き方を広げる可能性をもつ。クラウドソーシングの重要

な概念の1つとしてマイクロタスクがある。マイクロタスクでは依頼者(リクエスタ)と働き手(クラウドワーカ)がコミュニケーションを必要とせず短期間で実行できる。その結果、これまででは作業しなかったような状況で作業できるようになった。たとえば、5分の空き時間ができた場合にマイクロタスクを実行できる。マイクロタスクでは、一般にタスクの作業指示と結果入力フォームの両方が画面上に表示され、クラウドワーカは結果入力フォームにタスク結果を入力する。

本論文では、クラウドワーカが働ける状況を広

¹ 筑波大学
University of Tsukuba
² 筑波技術大学
Tsukuba University of Technology
a) ying.zhong.2018d@mlab.info
b) masaki@slis.tsukuba.ac.jp
c) koba@cs.k.tsukuba-tech.ac.jp
d) mori@slis.tsukuba.ac.jp

げるために異なるモダリティによるマイクロタスクの実現方法について探究する。具体的には聴覚情報を用いたマイクロタスクデザインを提案する。具体的には、タスクの指示文とタスクに関連するデータはすべて *audio matrix* で与えられると仮定する。*Audio matrix* とは、 N 次元行列のそれぞれの要素の位置に音声クリップが配置されたものである。*Audio matrix* タスク (AM task)*¹ が与えられた時、クラウドワーカは *audio matrix* 上の位置を矢印キーにより操作し該当位置の音声クリップを聞くことができる、そしてタスク結果を入力することができる。

このようなインターフェースを用いれば、ベッドやビーチサイドで寝そべっている時などディスプレイを見ない状況でもタスクを実行できる。他には視覚障害者がクラウドワーカの対象になりうる。これまで多くの視覚障害者の PC 作業の様子を観察した経験上、スクリーンリーダの助けを借りてタブ付きのスプレッドシートを扱えることがわかっている。このようなスプレッドシートは 3 次元の *audio matrix* として理論的にモデル化できる。 N 次元の *audio matrix* は、そのようなツールの自然な数学的抽象化である。

本論文では、マイクロタスクの特定のクラスである「オブジェクト比較タスク」に焦点を当てる。オブジェクト比較タスクは多くの実際のアプリケーションに現れ、またクラウドソーシングプラットフォームのオンラインタスクで広く見られる。例えば、エンティティ同一性判定タスク [1] はオブジェクト比較タスクでよく使われ、データのクリーニングと統合に不可欠である。

本論文の貢献は以下の通りである。第一に、ナビゲーションの意味の一貫性としての「軸の一貫性」が、視覚的な表現がないマイクロタスクでは重要な要素であることを実験的に示した。視覚的な情報がなければ、クラウドワーカはアクションを実行するときに何が起こるかを予測する必要がある。これは図 1 の *audio matrix* タスクによって示されている。そこではクラウドワーカは概念的に Excel のようなスプレッドシートの想起する。たとえば、図 1 下段真ん中のタスクでは、クラウドワーカがキーボードの右矢印キーなど、同じキーを押すたびに同じ意味する（異なるオブジェ

クトの同じ属性の値を比較する）場合、タスクが理解しやすくなる。しかし、図 1 上段真ん中のタスクでは、その一貫性を保証しない。下向きの矢印キーは、異なるオブジェクトの同じ属性の値を比較するだけでなく、タスクの指示も読み取るために押す必要がある。

第二に、軸の一貫性のある *audio matrix* タスクが視覚的表現を用いたタスクと同等な結果をもたらすことを示した。

2. 関連研究

2.1 音声ガイダンス

音声ガイダンスは、対話型音声応答 (IVR) で広く使用されている。対話型音声応答は、タッチフォン (タッチフォン) のキーを押して回答者が答えを提供する高品質の記録されたインタラクティブスクリプトに置き換えられる電話インタビュー技術である [2]。IVR に関して多くの研究がなされている。Yin らは視覚的な表示とキーワード検索に焦点を当て、電話の音声メニューのナビゲーションは視覚的なチャンネルの増強によって大幅に改善できることを発見した [3]。また、Aggarwal らは視覚障害者の使用に焦点を当てた [4]。

本論文では、クラウドソーシングプラットフォームでの音声ガイダンスの応用を検討し、視覚的な表現を持たない理解できるマイクロタスクを設計したいと考えている。

2.2 認知の一貫性

1940 年代から 1960 年代にかけて栄えた認知整合性理論は、複数の変数を含むタスクの処理を探索しようとした研究分野の 1 つだった。これらの理論は、主にフリッツ・ハイダーのバランス理論 [5], [6] に基づいていた。

認知一貫性理論は、構造ダイナミクスの 4 つの原理によってアニメーション化された。第一に、認知状態は、要素的ではなく、総合的に決定される。第二に、構造特性は動的である。第三に、精神的なプロセスの動的な特徴は、彼らが異なる構造的特性の状態に定住する傾向があるようにである。第四に、現在の実験プロジェクトに最も関連し、構造レベルで発生するこれらの動的な変化は、変化、すなわち「再構成」を伴う [7]。

認知一貫性理論は、認知の一般的な理論に向けた願望の中で開発され、比較的小きな構造のため

*1 *Audio matrix* タスク
<https://bit.ly/2XrJnQf>

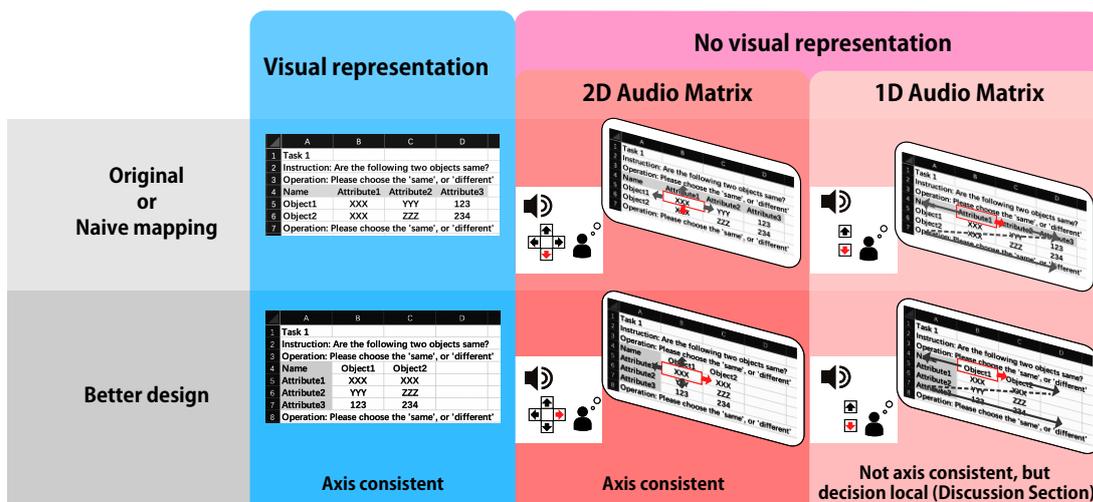


図1 6つのタスク表現の例。(左上)オリジナルビジュアルタスク。(中央上)オリジナルのナイーブマッピングによる2D audio matrix タスク。(右上)オリジナルのナイーブマッピングによる1D audio matrix タスク A。(左下)軸の一貫性があるビジュアルタスク。(中央下)軸の一貫性がある2D audio matrix タスク。(右下)軸の一貫性がある2D audio matrix から生成した1D audio matrix タスク B(このタスク自体には軸の一貫性が無いことに注意)。この2D オーディオマトリックスタスクでは、ユーザはそれぞれ上下矢印キーと左右矢印キーを使用して垂直方向と水平方向の方法で位置をナビゲートすることができる。

Name	City	Age
Dave Smith	New York	18

t_1

Name	City	Age
David Smith	New York	18

t_2

(a) structured

Description
Kingston 133x high-speed 4GB compact flash card ts4gcf133, 21.5 MB per sec data transfer rate, dual-channel support, multi-platform compatibility.

t_1

Description
Kingston ts4gcf133 4GB compactflash memory card (133x).

t_2

(b) textual

Name	Brand	Price
Adobe Acrobat 8		299.99

t_1

Name	Brand	Price
Acrobat 8	Adobe	299.99

t_2

(c) dirty

図2 オブジェクト比較の例 [1]

に設計される傾向があった。認知の一貫性の原理は社会心理学で広く議論されており、認知、感情的なつながり、行動に対する自己反映を含む幅広い社会心理学的構成要素に影響を与えるように見える [8]。人間は一貫性があると考えられる傾向がある。

2.3 Human-Powered Joins

オブジェクト比較(図2)は、human-powered join (データベース結合演算の Human-in-the-loop 版)の重要な要素の一つである [9]。この問題は、情報統合、自然言語理解、World Wide Web 上の情報処理、およびセマンティックネットワークにおいて重要な役割を果たす。その結果、人工知能、データベース、データマイニング、Web コミュニティで大きな注目を集めている [10]。Human-powered

join 演算を必要とするタスクでは、指示に基づいて2つのエンティティが同一であるか、または1つのエンティティを選択し答える。

クラウドソーシングは、これらの問題を解決する方法として注目を集めている。Corleone は、すべての主要なステップでクラウドワーカを使用する HOC ソリューションである [11]。ハイブリッドな人機械系のアプローチが提案されている。すなわち最初に機械がすべてのデータを粗めにパースし、人は最も可能性の高いペアのみを検証する [12]。クラウドソーシングのフレームワークは、ソーシャルメディアにおける具体的に組み込むものとして提案された [13]。

3. タスクデザイン

軸の一貫性の重要性を検証するため、オブジェ

クト比較タスクについて複数のタスクを設計した。クラウドワーカーは2つまたは3つの製品の情報（価格とレーティング）を提示され、価格と評価に基づいて、その中から最も費用対効果の高い製品を選んでもらう。製品の中から1つだけ明らかに他よりもコスト効率が高くなるように設定した（価格は低いほどレーティングは高いほど、より費用対効果が高くなることを意味する）。

データセットは、Kaggle^{*2}のショップ製品の公開データである。これは、この研究で使用される製品の価格と評価に関する情報が含まれている。

図3に示すように4つのaudio matrixタスクを含む以下の6つのタスクデザインを実験で比較対象とした。

- (1) オリジナルのビジュアルタスク（図1(左上)）
図3にこのタスクのスキーマを示す。
- (2) 軸の一貫性があるビジュアルタスク（図1(左下)）
図4にこのタスクのスキーマを示す。
- (3) オリジナルのナイーブマッピングによる2D audio matrixタスク（図1(中央上)）
このタスクのスキーマは、図3と同じである。ただし、このタスクには視覚的な情報がなく、クラウドワーカーは音声を聞いて質問に答える。クラウドワーカーは上、下、左および右の矢印を押しながら、オーディオマトリックスの位置をナビゲートすることによって質問に答える。
- (4) 軸の一貫性がある2D audio matrixタスク（図1(中央下)）
このタスクのスキーマは、図4と同じである。ただし、このタスクには視覚的な情報がなく、クラウドワーカーは音声を聞いて質問に答える。クラウドワーカーは上、下、左および右の矢印キーを押しながら、オーディオマトリックスの位置をナビゲートすることによって質問に答える。
- (5) 1D audio matrixタスクA（図1(右上)）
このタスクのスキーマは、図3と同じである。クラウドワーカーは上、下の矢印キーを押しながら、ジグザグの方向に沿ってオーディオマトリックスの位置をナビゲートすることによって質問に答える。たとえば、図3の場合、再生順序は「Instruction」, 「Opera-

Task12		
Instruction: Which product is more cost-effective based on the prices and ratings of the following products?		
Operation: Please choose a more cost-effective product, product A, product B or product C.		
Name	Price	Rating
Product A	60	5
Product B	60	3
Product C	60	4
Operation: Please choose a more cost-effective product, product A, product B or product C.		

図3 ナイーブマッピング 2D audio matrix タスク

Task1		
Instruction: Which product is more cost-effective based on the prices and ratings of the following products?		
Operation: Please choose a more cost-effective product, product A or product B.		
Name	Product A	Product B
Price	34	24
Rating	4	5
Operation: Please choose a more cost-effective product, product A or product B.		

図4 軸の一貫性がある 2D audio matrix タスク

tion」, 「Name」, 「Price」, 「Rating」, 「Product A」, 「60」, 「5」, 「Product B」, ... となる。

- (6) 1D audio matrixタスクB（図1(右下)）
このタスクのスキーマは、図4と同じである。たとえば、図4の場合、再生順序は「Instruction」, 「Operation」, 「Name」, 「Product A」, 「Product B」, 「Price」, 「34」, 「24」, 「Rating」... となる。

4. 実験1: クラウドワーカー実験

4.1 実験参加者

Amazon Mechanical Turk^{*3}を通じて180人（6種類×30人）のクラウドワーカーが実験に参加した。報酬はビジュアルタスクを完了したときに一人あたり\$0.01であり、オーディオタスクを完了したときに一人あたり\$0.05であった。

4.2 手続き

クラウドワーカーは12個の同じタスクデザインオブジェクト比較タスクを行うよう要求された。12個のうち6つのマイクロタスクには2つの製品があり、他の6つのマイクロタスクには3つの製品があった。

4.3 結果

4.3.1 正解率

図5は、実験1の各タスクデザインの正解率を示している。費用対効果の高い製品を見つけるには、クラウドワーカーは2つの属性を同時に比較する必要があった。聴覚情報のみを用いたタスクでは、軸の一貫性があるaudio matrixタスクの正解

*2 Kaggle: <https://www.kaggle.com>

*3 Amazon Mechanical Turk: <https://www.mturk.com>

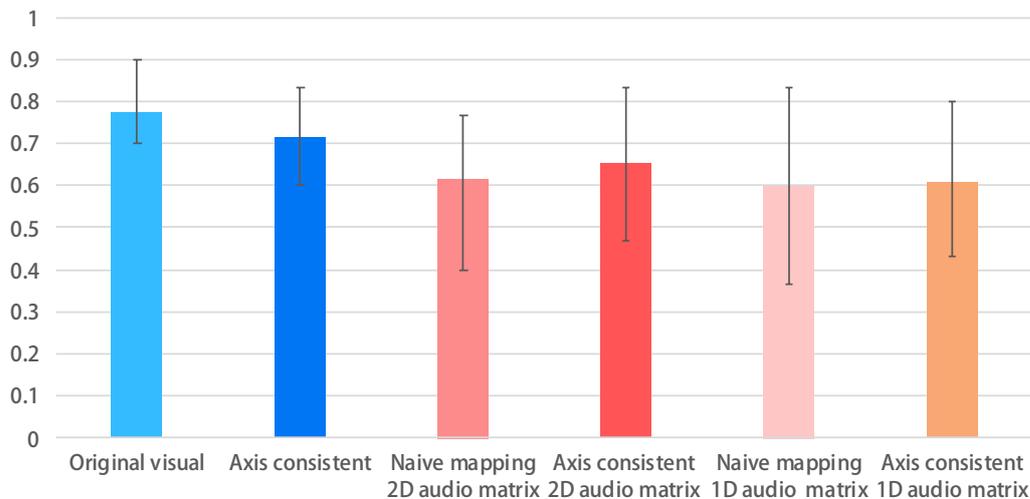


図5 実験1における6つのタスクデザインの正解率

表1 実験1に各オーディオタスクデザインの平均回答時間

	オーディオ (2D)	オーディオ (1D)
ナイーブマッピング (タスク A)	3m10s	5m54s
軸の一貫性 (タスク B)	3m18s	3m34s

率が最も高くなった。

異なるタスクデザインの正解率を比較するために、対応なしのtテストを実施した。その結果、元の構造と軸の一貫性がある構造との間のタスク設計による大きな影響はなかった（ビジュアルタスク ($t_{22} = 3.793, p = 0.643$), 2D audio matrix タスク ($t_{22} = 0.574, p = 0.457$), 1D audio matrix タスク ($t_{22} = 0.029, p = 0.867$))。

4.3.2 回答時間

表1は、実験1のオーディオタスクの合計回答時間の平均を示している。1D audio matrix タスクの平均回答時間は、2D audio matrix タスクの平均回答時間よりも長くなった。ただし、1D audio matrix タスク B の平均回答時間は他の 2D audio matrix タスクと同等でしたが、1D audio matrix タスク A の平均回答時間ははるかに長くなった。

5. 実験2:視覚障害者ケーススタディ

5.1 手続き

視覚障害者1人を対象に実験を行いインタビューを行った。実験1と同様に24つのマイクロタスクを提示した。各タスクデザインには6つのマイクロタスクがあった。3つのマイクロタスクには2つの製品があり、他の3つのマイクロタスクには3つの製品があった。

5.2 結果

5.2.1 正解率

表2に示すように、4つ全ての audio matrix タスクでの正解率は高かった。被験者は4つのタスクをすべて理解した。このうち、ナイーブマッピング 2D audio matrix タスクと軸の一貫性がある 2D audio matrix タスクが最も高く、1であった。1D audio matrix タスク A が0.83であった。1D audio matrix タスク B は最も低く、0.67であった。

5.2.2 回答時間

表3に示すように、4つのオーディオタスクデザイン中、1D audio matrix タスク B 最も長く、83秒であった。1D audio matrix タスク A は25秒であった。ナイーブマッピング 2D audio matrix は16秒であった。軸の一貫性がある 2D audio matrix タスク最も短く、10秒であった。

5.2.3 インタビュー

実験参加者へのインタビューも行った。

質問1 タスクを理解できましたか？

私は 1D audio matrix タスク B をよく理解することはできなかったが、私は他のタスクを理解することができた。

質問2 1D audio matrix タスク B の問題は何だと思うか？

(同じオブジェクトの) 価格とレーティ

表 2 実験 2 に各オーディオタスクデザインの平均正解率

	オーディオ (2D)	オーディオ (1D)
ナイーブマッピング (タスク A)	0.83	1
軸の一貫性 (タスク B)	1	0.67

表 3 実験 2 に各オーディオタスクデザインの平均回答時間

	オーディオ (2D)	オーディオ (1D)
ナイーブマッピング (タスク A)	16s	25s
軸の一貫性 (タスク B)	10s	83s

ングの項目が隣り合っていない。比較するのは難しい。

質問 3 1D タスクのうちどちらのタスクが好きか?

1D audio matrix タスク A.

質問 4 2D タスクのうちどちらのタスクが好きか?

どちらも同じくらい。

質問 5 この 4 種類のタスクでどのタスクが好きか?

テーブルのような (上下左右操作できる 2D) タスク。

6. まとめ

我々は、聴覚情報のみによるマイクロタスクに関する実験結果を報告した。その結果、軸の一貫性がある audio matrix タスクデザインは、視覚的な表現を持たない Human-powered join 演算を使用してマイクロタスクを理解するのに有効であることを示した。この実験を通じて、我々は、軸の一貫性が聴覚情報のみによるマイクロタスクデザインの重要な要因であることを確認した。

今後は、タスクのデザインを変更し、より多くの被験者を実験に予定である。

謝辞 本研究の一部は JST CREST (JP-MJCR16E3) による。

参考文献

- [1] Mudgal, S., Li, H., Rekatsinas, T., Doan, A., Park, Y., Krishnan, G., Deep, R., Arcaute, E. and Reghavendra, V.: *Deep learning for entity matching: A design space exploration*, SIGMOD (2018).
- [2] Corkrey, R. and Parkinson, L.: *Interactive voice re-sponse: Review of studies 19892000.*, Behavior Research Methods, Instruments, Computers (2002).
- [3] Yin, M. and Zhai, S.: *The benefits of augmenting telephone voice menu navigation with visual browsing and search.*, CHI (2006).
- [4] Aggarwal, P., Luthra, G., Koradia, Z., Seth, A., Gadodia, P., Abraham, G. and Jain, N.: *A case study on the use of ivr systems by visually impaired*

people. the 3rd ACM Symposium on Computing for Development (2013).

- [5] Heider, F.: *Attitudes and cognitive organization.*, Journal of Psychology (1946).
- [6] Heider, F.: *The psychology of interpersonal relations.*, The psychology of interpersonal relations (1958).
- [7] Simon, D., Snow, C. J. and Read, S. J.: *The redux of cognitive consistency theories: Evidence judgments by constraint satisfaction.*, Journal of Personality and Social Psychology (2004).
- [8] Heberlein, T. A. and Black, J. S.: *Cognitive consistency and environmental action.*, Environment and Behavior (1981).
- [9] Marcus, A., Wu, E., Karger, D., Madden, S. and Miller, R.: *Human-powered sorts and joins.*, PVLDB (2011).
- [10] Shen, W., Li, X. and Doan, A.: *Constraint-based entity matching.*, AAAI (2005).
- [11] Gokhale, C., Das, S., Doanand, A., Naughton, J. F., Ram-palli, N., Shavlik, J. W. and Zhu, X.: *Corleone: hands-off crowdsourcing for entity matching.*, SIGMOD (2014).
- [12] Wang, J., Kraska, T., Franklin, M. J. and Feng, J.: *Crowder: Crowdsourcing entity resolution.*, VLDB (2012).
- [13] Jiang, L., Wang, Y., Hoffart, J. and Weikum, G.: *Crowdsourced entity markup.* CrowdSem (2013).