

# 回顧的思考発話の中で誤情報を提示する mRTA 法の効果

森 順平<sup>1</sup> 西本 一志<sup>1</sup>

**概要：**定性調査やユーザビリティテストなどで用いられている回顧的思考発話法（Retrospective Think-Aloud 法：RTA 法）には、十分な発話量が得られなかったり記憶の正確な想起が難しかったりするなどの問題がある。我々は、これらの問題を解決するために、RTA 法の実施時に提示するタスク遂行時の記録情報に誤情報を混入する方法である mRTA 法（misinformed Retrospective Think-Aloud 法）を提案している。mRTA 法の効果を検証するための実験を実施した。その結果 mRTA 法は、特に熟達度の高い作業者に対しては記憶の正確な想起と発話量の増加に有効である一方、熟練度の低い作業者においては記憶の正確な想起をむしろ好ましくない影響を与える可能性があることを確認した。

## 1. はじめに

思考発話法（Think-Aloud：TA）は、あるタスクを遂行している間に頭に思い浮かんだことをすべて発話して報告することによる行動分析の手法である。回顧的思考発話法（Retrospective Think-Aloud: RTA）は思考発話法の 1 種であり、タスクを遂行した後に、タスクを遂行している間の様子を記録した映像などの情報を提示しながら、タスクを遂行時に考えていた内容を思い出してもらい、発話させる手法である [1], [2]。実際に RTA は、Web サービスやアプリケーションの開発の中で行うユーザビリティテスト [1] や、スポーツやゲーム、経営等における戦略分析など様々な場面で用いられている。

しかし、RTA 法ではタスク遂行後に発話するため、タスク遂行時の記憶を正確かつ十分に思い出させることが難しいことや、記憶の合理化が行われてしまうこと、その結果として十分な発話量が得られず正確な思考内容を詳しく探ることができないことといった問題点が指摘されている [3]。これらの問題点はいまだに十分に解決されていない。そこで我々は、この問題の解決策のひとつとして、誤情報（misinformation）を活用した「**mRTA (misinformed Retrospective Think-Aloud)**」法を提案している [4]。

誤情報の活用事例として、知見ら [5] がプログラミングの授業に誤情報を活用した研究がある。この研究では、授業の中に誤情報を混ぜたことで、その授業を受けた生徒らが授業に対して受動的な態度から能動的な態度に変化したと報告されている。通常の授業では説明されてもその内容

をただ受け取るのみであったが、誤情報を混ぜることを教示し、正情報と誤情報の判別を生徒らに自主的に行わせたことで、能動的な態度に変化したとされている。また小森ら [6] は、誤情報を与えて指摘意欲を誘発することで、何についてどの程度精通しているかという知識レベルをあぶりだすことができたと報告している。このような、誤情報を提示されることで参加態度が受動的から能動的に変化する効果や指摘意欲がかき立てられ記憶を想起させる効果を RTA に活用することで、正確かつ詳細な記憶の想起を促し発話量を増加させることができると考えられる。

そこで mRTA 法では、RTA を行う際、序盤で提示するタスク遂行中の情報に若干の誤情報を混ぜ込む手段をとる。これにより、発話者の報告姿勢を能動的なものに変化させ、後続する情報提示の際に、誤情報を提示した場合のみならず、誤りを含まない正しい情報を提示した場合においても発話内容の正確性や発話量が向上することが期待される。以下本稿では、mRTA 法の有効性を実験によって検証するとともに、mRTA 法の適用可能な対象について議論する。

## 2. 関連研究

TA に関する研究はこれまでに多く行われている。加藤 [7] は学習過程におけるユーザの考えを明らかにすることを目的として、話し手に学習中に浮かんだ疑問を質問させることで、どのような疑問や問題に直面するか明らかにすることができることを報告した。Boren ら [8] は、発話は常にコミュニケーションの性質を持っているとして、話し手と聞き手が自然なコミュニケーションをとることでより多くの有用なデータを取ることができると報告した。Ball

<sup>1</sup> 北陸先端科学技術大学院大学  
Graduate School of Advanced Science and Technology,  
Japan Advanced Institute of Science and Technology

熟達度	実験群	統制群
初心者	2	2
中級者	7	7
上級者	3	3

ら [9] は、タスク遂行中に視線情報を取得し、RTA を行う際にスクリーン上に視線情報の映像を再生し、その映像を見ながら発話させたところ、従来の TA の手法よりも豊富な情報を引き出すことができたと報告している。Walendy ら [10] も視線情報と TA を組み合わせた手法を提案しており、視線情報と発話内容から包括的に解釈することでより理解することができると報告している。

### 3. 本実験

この実験は、北陸先端科学技術大学院大学・知識科学倫理審査委員会の承認（承認番号 KSEC-D2024103103）を受けて実施した。

#### 3.1 目的

本実験では、タスク課題として麻雀の次の一手問題を行い、その後 mRTA を実施、もしくは既存手法である RTA を実施し、比較分析することで、mRTA の有効性を検証する。

#### 3.2 実験手順

##### 3.2.1 被験者の群分け

本実験には麻雀経験のある大学院生 24 名が被験者として参加した。被験者はそれぞれ 12 人ずつ、mRTA を実施する実験群と通常の RTA を実施する統制群に振り分けた。被験者には、実験に参加する前に被験者自身の麻雀プレイにおける熟達度を以下の種別を基に申告してもらい、その情報を基に実験群と統制群に熟達度のばらつきがないように振り分けた。実験群と統制群の被験者の内訳を表 1 に示す。

- 初心者：プレイする際には役の成立と点数計算の補助が必要な方。
- 中級者：プレイする際には点数計算のみ補助が必要な方。
- 上級者：プレイする際には補助は不要な方。

##### 3.2.2 次の一手問題

「次の一手問題」では被験者に図 1 のような麻雀のある局面を示す盤面画像を 1 枚見せ、その局面で次にどのような行動をとるかを考え、最終的な意思決定のみを解答用紙に記述してもらった。盤面画像は 1 枚ずつ合計 12 枚提示し、また、被験者ごとに思考する時間を統制するために、1 問あたり 60 秒の時間制限を設定した。なお、次の一手問題を実施する前に、被験者にはその対局におけるルール設

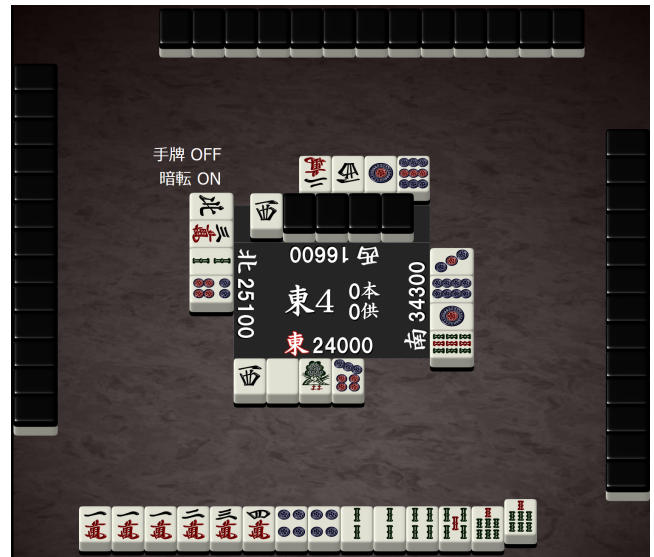


図 1 盤面画像の例

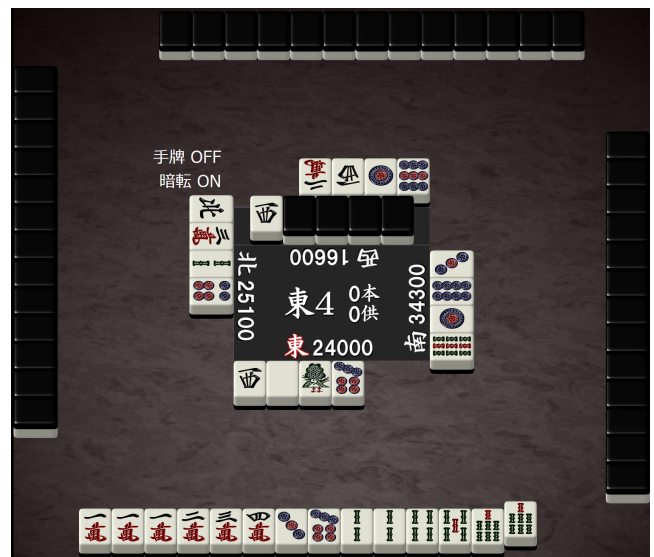


図 2 誤画像の例。図 1 における 2 個の四筒の頭が三筒と七筒に差し替えられている。

定（四人半荘、赤牌あり）や盤面画像のどの箇所にもどのような情報が映されているか、および回答の制限時間について説明した。また、盤面上のどこに注視しているかを確認することを目的として視線データも実験中に収集した。

##### 3.2.3 RTA・mRTA

12 枚の盤面画像に対する次の一手問題がすべて終了した後、実験群には mRTA、統制群には RTA を実施した。次の一手問題と同様に盤面画像を 1 枚提示し、提示された画像から次の一手問題でどのようなことを考えていたか思い出してもらい、半構造化インタビューを行った。併せて、次の一手問題と同様に、視線データを収集した。なお、次の一手問題が終了した時点で回答用紙はすべて回収し、被験者が自分の回答を見られないようにしておいた。RTA の場合は次の一手問題で提示した画像と同じ画像 12 枚を提示した。mRTA では 12 枚のうち 3 枚を誤画像に変更して

提示し、誤画像が含まれていることは教示せずに行った。図2は、図1の誤画像として実際に実験で提示した画像である。誤画像は提示する画像12枚のうち、前半の1, 3, 6枚目に配置した。こうして前半に誤画像を提示することで、被験者の報告姿勢を能動的なものに変化させ、後半で正画像を提示した場合でも報告内容の正確さや発話量が向上するかどうかを検証する。RTAとmRTAを行う際には、それぞれに誤情報の要素以外で発話内容に差が生じることを軽減するために、実験者が被験者に質問する際のルールを以下の通り設定した。

- (1) 各盤面について次の一手問題を解いた際にどんな回答をしたか質問し、被験者が行った意思決定を覚えているか確認する。
- (2) (1)を行った後、その行動を取った理由について質問し、意思決定をする上で考えていた内容を聴取する。  
(1)にて被験者が発話したものと回答用紙に記した意思決定が異なっている場合には、理由を尋ねた後に回答に記した意思決定を被験者に教示し、その内容について覚えていることを聴取する。
- (3) (2)を行った後、(1)と(2)の発話内の考え以外で、回答するときを考えていた行動があったか、あった場合には何を考えていたかを聴取する。

12問すべてのインタビューが終了した後、発話する際にどのようにして次の一手問題の時の記憶を引き出していたかを問うた。

### 3.3 結果

#### 3.3.1 記憶想起の正確さ

記憶の想起の仕方は主に盤面画像を見て、画像として想起するパターンと、画像の情報を見てから戦略を考え、その内容から次の一手問題で考えていた内容を思い出すパターンの2パターンであった。また、被験者の一部には、今(RTA・mRTAの実施中)の自分が考えていることは、次の一手問題を考慮中の自分が考えていたことと必ず同じであるという(根拠のない)信念を持ち、記憶想起を行わずに今の考えを発話している者(これを以下では「非想起者」と呼ぶ)が3人いた(実験群の中級者に1名、統制群の中級者に1名、初級者に1名)。

表2に、mRTAにおける誤画像の誤り内容判断の正答率を示す。表2内の「完全正答率」は改変された部分と改変される前の情報の両方を答えることができた回数の割合を表し、「部分正答率」は改変された部分を答えることができた(完全正答を含む)回数の割合を示している。非想起者を除外の行は、3人の非想起者の結果を除外した場合の割合を示している。

表3に、被験者による次の一手問題の回答とmRTA/RTAで想起した結果の一致率(想起率)を示す。表3内の「完全正答率」は最初の意思決定確認にて決定手を完全に思い

表2 実験群(mRTA)における正誤判断の正答率

		上級者	中級者	初級者
全員	完全正答率	0.89	0.19	0.00
	部分正答率	1.00	0.71	0.50
非想起者を除外	完全正答率	0.89	0.22	0.00
	部分正答率	1.00	0.83	0.50

出した問題の割合であり、「部分正答率」は決定手の一部でも答えることができた問題(完全正答を含む)の割合を示す。非想起者を除外の行は、3人の非想起者の結果を除外した場合の割合を示している。

#### 3.3.2 発話量

被験者の中には日本語を母語としない留学生7名も含まれていたが、発話量の比較では日本語を母語とする被験者のデータのみを分析対象とする。さらに、非想起者の3名と、思考発話に習熟していて極端に発話量が多かった1名の被験者を除外した13名について発話内容を分析した。なお、これらの被験者のデータを除外した結果、熟達度が初心者レベルの被験者が各群1名ずつになってしまったため、初心者については発話量の分析は行わなかった。

分析対象の発話として、最初の質問として問うた意思決定の内容が次の一手問題での回答と異なっていた場合のものは除外した。これは、異なっている場合には回答の内容を教示しているため、発話量が構造上多くなってしまっているためである。発話内容を平仮名変換し、フィラーを除去した。また誤画像の際の発話については、盤面の違いに関する発話は除去した。以上の前処理を行った上で、その文字数を熟達度ごとに実験群と統制群で比較した。表4に、各群における提示画像ごとの平均発話文字数を示す。実験群に対して誤画像を提示した画像1, 3, 6については実験群の結果を太字にしている。

実験群と統制群の発話量の結果を、マン・ホイットニーのU検定で比較した。表5に全画像を対象とした比較結果を、表6には実験群に3枚の誤画像を提示し終えたあとの画像7~12に関する比較結果を、それぞれ示す。なお、誤画像(画像1, 3, 6)についてはNが3しかないため、比較検定は行っていない。

- 全画像を対象とした検定の結果(表5)では、  
上級者グループについては実験群(Median=441.8 (212.0 - 695.0))と統制群(Median=274.0 (134.5 - 375.5))の間に1%水準で有意差が見られた( $z = 2.5981$ ,  $p = 0.0094 < 0.01$ )。  
中級者グループについては、実験群(Median=285.8 (197.5 - 363.3))と統制群(Median=266.2 (168.5 - 317.3))の間に有意差は見られなかったが、有意傾向が認められた( $z = 1.7321$ ,  $p = 0.0833 < 0.1$ )。
- 両群に対して後半の画像7~12を提示した場合を対象とした検定の結果(表6)では、  
上級者グループについては実験群(Median=412.0

表 3 想起率

		実験群 (mRTA)			統制群 (RTA)		
		上級者	中級者	初級者	上級者	中級者	初級者
全員	完全正答率	0.97	0.79	0.54	0.89	0.90	0.75
	部分正答率	0.97	0.88	0.71	0.92	1.00	0.83
非想起者を除外	完全正答率	0.97	0.88	0.54	0.89	0.92	0.92
	部分正答率	0.97	0.96	0.71	0.92	1.00	1.00

表 4 mRTA/RTA における平均発話文字数

	実験群 (mRTA)		統制群 (RTA)	
	上級者	中級者	上級者	中級者
画像 1	401.5	280.0	173.5	270.7
画像 2	654.5	323.8	375.5	255.0
画像 3	423.5	340.8	175.5	262.7
画像 4	521.0	266.8	241.5	261.0
画像 5	470.0	366.3	309.0	312.3
画像 6	230.5	270.5	302.0	168.5
画像 7	460.0	335.0	331.0	245.0
画像 8	212.0	220.5	266.0	242.3
画像 9	695.0	277.8	272.5	317.3
画像 10	469.5	291.5	134.5	269.7
画像 11	364.0	360.5	314.5	303.0
画像 12	247.5	197.8	275.5	274.3

表 5 U 検定の結果・全問題

		N	Median	Z	p-value
上級者	実験群 (mRTA)	12	441.8	2.5981	0.0094**
	統制群 (RTA)	12	274.0		
中級者	実験群 (mRTA)	12	285.8	1.7321	0.0833†
	統制群 (RTA)	12	266.2		

表 6 U 検定の結果・画像 7~12

		N	Median	Z	p-value
上級者	実験群 (mRTA)	6	412.0	1.2810	0.2002
	統制群 (RTA)	6	284.6		
中級者	実験群 (mRTA)	6	274.0	0.3203	0.7488
	統制群 (RTA)	6	272.0		

(212.0 - 695.0) と統制群 (*Median*=284.6 (197.8 - 360.5) の間では有意差が見られなかった ( $z = 1.2810$ ,  $p = 0.2002$ ).

中級者グループについては、実験群 (*Median*=274.0 (134.5 - 331.0) と統制群 (*Median*=272.0 (242.3 - 317.3) の間に有意差は見られなかった ( $z = 0.3203$ ,  $p = 0.7488$ ).

### 3.3.3 視線データ

次の一手問題に回答する際の初級者の視線データの 1 例を図 3 に、中級者の視線データの 1 例を図 4 に、上級者の視線データの 1 例を図 5 に、それぞれ示す。これらは、すべて同じ次の一手問題の画像である。初級者グループと中級者グループは手牌の部分にのみ注視度が高い領域が集中しており、手牌を中心に見ていることが伺えた (図 3, 4)。一方上級者グループでは、手牌だけでなく、局数や点数状

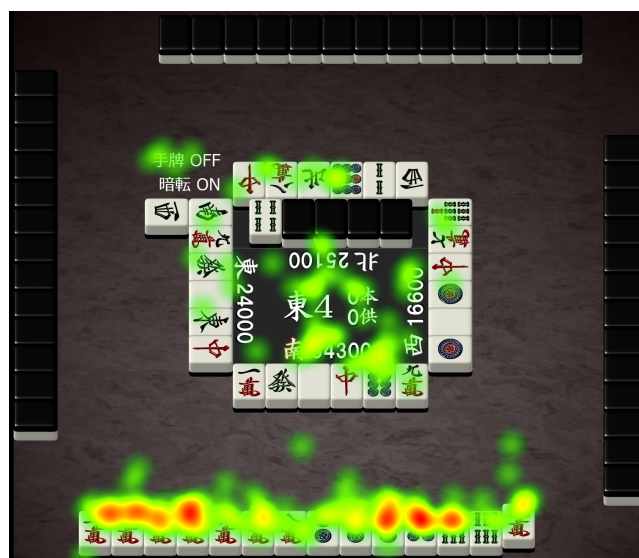


図 3 初級者

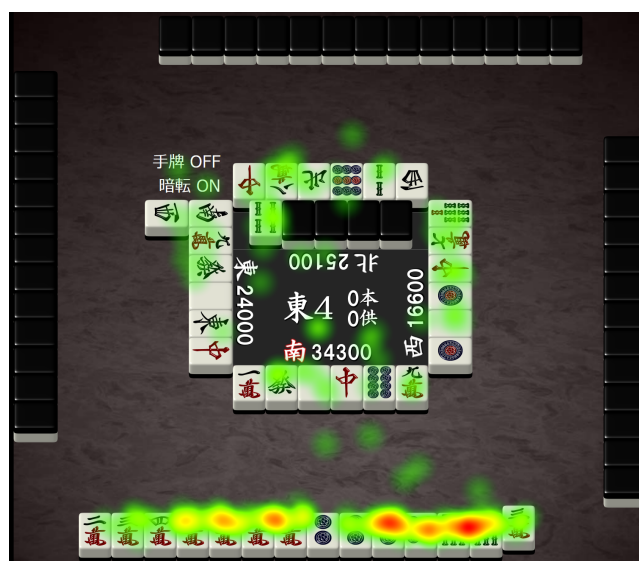


図 4 中級者

況、河の牌にも注視度が高い領域が分布しており、盤面全体を見ていることが伺えた (図 5)。

## 4. 考察

### 4.1 記憶想起の正確さについて

表 2 から、実験群において上級者グループの完全正答率は中級者や初級者と比べてかなり高く、元の盤面の情報を正しく引き出せていたことが伺える。このことから、上級



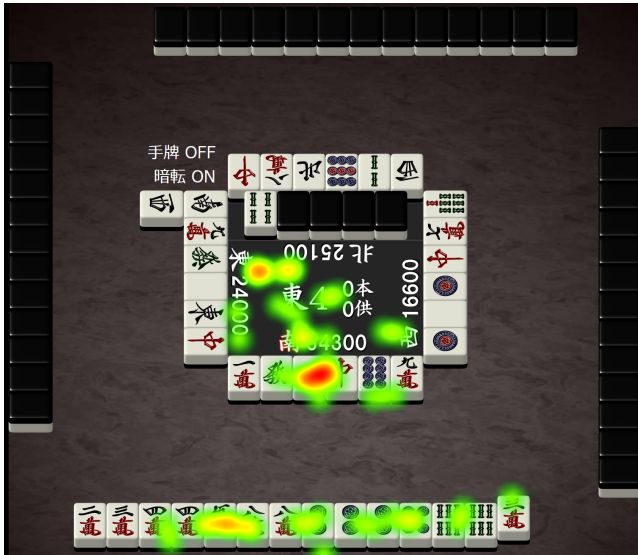


図 5 上級者

者グループでは mRTA を使用することでより正確な記憶の想起を促すことができる可能性が示された。また表 2 から、実験群の中級者グループでは完全正答率は低いものの、部分正答率は特に非想起者を除外した場合には高い確率で正誤判断を行えていることがわかる。このことから、中級者ではある程度の記憶の想起は行えているが、盤面の情報を詳細に記憶できてはいないことが考えられる。

図 3, 4, 5 に示した次の一手問題の正画像を図 6 に示す。mRTA を行った実験群では、この正画像の盤面情報を一部改変して誤画像 (図 7) として使用している。上級者グループは全員、この図 7 の誤画像を見て、手牌の改変箇所を見抜くことができていた。一方、実験群の初級者グループや中級者グループには誤画像を見抜けなかった被験者がいた。それらの被験者の視線情報を確認したところ、改変箇所が停留点になっていたにもかかわらず、改変されていることに気づいていなかった。図 3, 4 に示した視線データの元となった被験者は、いずれも誤画像であることを見抜けていなかった。このように、中級者や初級者は盤面情報の詳細を覚えきれておらず、改変に気づくことができないケースが多々あるため、誤情報を提示してもそれが正確な記憶の想起のきっかけとなり難いことが示唆される。これに対し上級者は、幅広く視線が向きながらも盤面情報を正確に覚えているため、誤情報が正確な記憶の想起とその内容に関する発言へのきっかけとなり得ることが示唆される。

De Groot ら [11] や伊藤ら [12] は将棋やチェスのエキスパートは、初級者が数個ほどの駒しか覚えられていない間により多くの盤面情報を覚え、その内容を正確に再現することができることを報告している。麻雀においてもこの性質が見られたと考えられる。上級者グループは、元の盤面の情報を詳細に覚えていたため完全正答率が高くなった。これに対し、中級者グループは元の盤面との違いがあることは

認識できたが、元の盤面の情報を詳細に覚えてはいなかったため、完全正答率が低くなっていたと思われる。さらに表 3 から、上級者グループでは統制群に比べて実験群の方がやや高い精度で想起を行えている傾向が認められる。一方、中級者と初級者に関しては、実験群の方が統制群よりも想起率が低い傾向が認められる。以上の結果から、誤情報を混ぜ込む mRTA 法は、中級者や初級者に対しては情報の正確な想起に関してあまり有効ではなく、むしろ好ましくない影響を与える可能性があることが示唆される。

#### 4.2 発話量について

表 5 から、上級者グループでは実験群の方が統制群よりも有意に発話量が多くなっていることがわかる。この結果から、上級者においては作業序盤の画像 1, 3, 6 で誤画像を提示することにより全体の発話量が増加することが示された。また、中級者グループでも実験群の方が統制群よりも発話量が多くなる傾向が認められた。ゆえに、発話量の増加に関しては mRTA 法は中級者に対しても有効である可能性が示唆された。しかしながら先述のとおり、中級者については想起情報の正確さが低下する傾向があるので、全体としての有効性については一概に良いとは言えない結果となった。

表 6 から、実験群に対して 3 つの誤画像を提示し終わった後の、画像 7~12 の正画像を提示した際における発話量については、上級者・中級者ともに実験群と統制群の間には有意差が認められない結果となった。ただし、上級者については Median の値が実験群では 412.0、統制群では 284.6 と大きく異なっていることから、上級者については誤画像の提示がもたらした発話量増加への影響が後半でも続いている可能性が示唆されている。有意差が認められなかったのは N が 6 と少ないためである可能性が考えられるので、今後データ数を増やして検証を進めたい。

#### 5. おわりに

本研究では、RTA に誤情報を加えることで、記憶の想起や発話量の増加を促す mRTA 法を提案した。本実験では麻雀プレーヤーが次の一手問題について考える思考内容を題材にして、タスクに誤情報を加えた mRTA を実施し、既存手法の RTA と比較することで、記憶の想起や発話量の増加を促すことができるか検証した。その結果、mRTA を実施することで、熟達度の高い被験者については RTA と比べてより正確な記憶の想起と発話量の増加を促すことができることが確認された。そのため、mRTA は熟達度の高い作業者を対象とした調査では有用であると考えられる。一方、熟達度が低い作業者に対しては、想起の正確さに好ましくない影響を与え、発話量も増加しない可能性があることが示唆された。ゆえに、mRTA 法を適用する際は作業者の熟達度に十分配慮する必要があると考えられる。

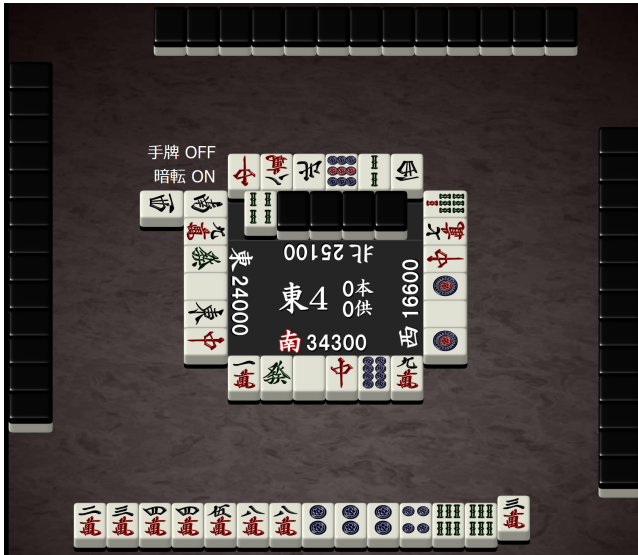


図 6 図 3, 4, 5 に示した視線画像の正面像

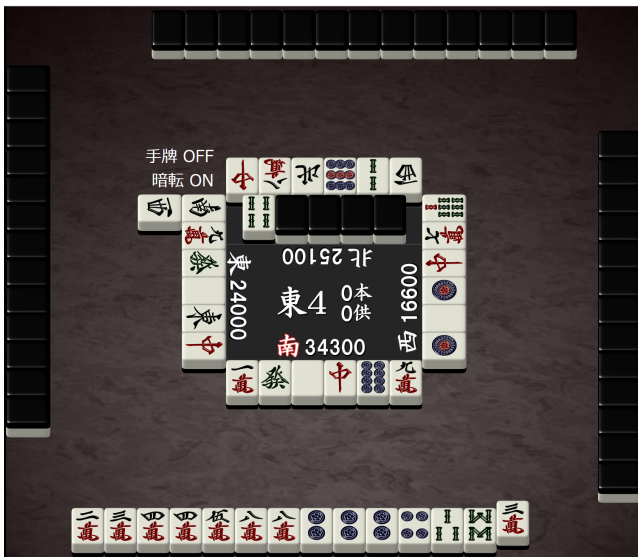


図 7 図 6 の誤画像。2 個の六索が三索と八索に差し替えられている。

現時点では分析の内容は発話内容の量や再認に関する分析のみであり、発話内容の質に関してはまだ検討の余地がある。引き続き分析を行い、mRTA の有用性について検討していきたい。

**謝辞** 実験にご協力いただいた協力者の皆さんに厚くお礼申し上げます。本研究は JSPS 科研費 JP24K02976 の助成を受けたものです。

## 参考文献

- [1] Nielsen, J.: *Usability engineering*, Morgan Kaufmann (1994).
- [2] Henderson, R. D., Smith, M. C., Podd, J. and Varela-Alvarez, H.: A comparison of the four prominent user-based methods for evaluating the usability of computer software, *Ergonomics*, Vol. 38, No. 10, pp. 2030–2044 (1995).
- [3] Russo, J. E., Johnson, E. J. and Stephens, D. L.: The validity of verbal protocols, *Memory & cognition*, Vol. 17,

- pp. 759–769 (1989).
- [4] 森順平, 高宗楓, 西本一志: 誤情報の活用による Retrospective Think Aloud 法改良の試み, 日本認知科学会第 41 回大会発表論文集, No. P-3-5, pp. 639–642 (2024).
- [5] 知見邦彦, 樋山淳雄, 宮寺庸造: 失敗知識を利用したプログラミング学習環境の構築, 電子情報通信学会論文誌 D, Vol. 88, No. 1, pp. 66–75 (2005).
- [6] 小森麻友香, 高島健太郎, 西本一志: 潜在的技能保有者を顕在化するための娯乐的 Know-who 支援手法, 情報処理学会研究報告, Vol. 2019-HCI-182, No. 9, pp. 1–8 (2019).
- [7] Kato, T.: What “question-asking protocols” can say about the user interface, *International Journal of Man-Machine Studies*, Vol. 25, No. 6, pp. 659–673 (1986).
- [8] Boren, T. and Ramey, J.: Thinking aloud: Reconciling theory and practice, *IEEE transactions on professional communication*, Vol. 43, No. 3, pp. 261–278 (2000).
- [9] Ball, L. J., Eger, N., Stevens, R. and Dodd, J.: Applying the PEEP method in usability testing, *Interfaces*, Vol. 67, No. Summer, pp. 15–19 (2006).
- [10] Walendy, R., Weber, M., Li, J., Becker, S., Wiesen, C., Elson, M., Kim, Y., Fawaz, K., Rummel, N. and Paar, C.: I see an IC: A Mixed-Methods Approach to Study Human Problem-Solving Processes in Hardware Reverse Engineering, in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–20 (2024).
- [11] De Groot, A. D. and De Groot, A. D.: *Thought and choice in chess*, Vol. 4, Walter de Gruyter (1978).
- [12] 伊藤毅志, 松原仁他: 将棋の認知科学的研究 (1)-記憶実験からの考察, 情報処理学会論文誌, Vol. 43, No. 10, pp. 2998–3011 (2002).