

## 皮膚電気活動を用いた数独問題の難易度評価

棟方渚<sup>†1</sup> 志水雅俊<sup>†2</sup> 松原仁<sup>†3</sup>

ペンシルパズルの一種である、「数独 (ナンバープレイス)」は近年世界的なブームとなっており、問題の自動生成や解法の解説など計算機科学的な研究だけではなく、人間の思考過程を解明する認知科学的な解析が行われるなど、様々な方面から研究が行われている。人間の認知過程の解明を目指す研究では、その方法の一つとしてチェスや将棋を初めとした思考ゲームにおける問題解決の思考に着目してきた。思考の記録には、ユーザにその過程を発話させ、その内容を分析する発話プロトコル解析が一般的に用いられてきた。しかしこの手法は、発話を行うことで思考が妨げられるといった問題や、発話されない無意識的な思考を扱うことができないといった問題点が挙げられる。そこで著者らは、「数独」におけるユーザの思考の分析手法として、意識的な行動である発話ではなく、無意識的に絶えず変化する生体情報に着目した。特に、人間の思考過程の移り変わりや計算に対する集中の有無を測定できる皮膚電気活動を用いて、数独を解くユーザのスキルレベルを導き出した。また、その解析から導き出したユーザのスキルレベルが、実際にどの程度対応できていたのかどうかを調べる評価実験を行った。

### Evaluation of SUDOKU Difficulty Levels Using the User's Skin Conductance Response

NAGISA MUNEKATA<sup>†1</sup> MASATOSHI SHIMIZU<sup>†2</sup>  
HITOSHI MATSUBARA<sup>†2</sup>

A type of Pencil puzzle, "Sudoku(or Number Place)", has become world-wide popular in recent year. Sudoku puzzle generation and solving were researched from various fields, is not only computer science research but cognitive science research that reveal human thinking processes. Cognitive science researcher focused on that is thinking processes of problem solving in games such as chess and shogi as one way. To analyze the thinking process, researcher used protocol analysis. However, this method has some problem that thinking is blocked by utterance and unconscious thinking is not used. In order to solve this problem, we focused on biological information that always changes unconsciously as analysis method of thinking but utterance. We examine users' skill level of sudoku by Skin conductance response that can measure change in thinking process and concentration. We examine users' skill level of sudoku by Skin conductance response that can measure change in thinking process and concentration.

#### 1. はじめに

ペンシルパズルの一種である、「数独 (ナンバープレイス)」は近年世界的なブームとなり、問題の自動生成や解法の解説など計算機科学的な研究だけではなく、人間の思考過程を解明する認知科学的な解析が行われるなど、様々な方面から研究が行われている[1][2]。また、数独は初心者から上級者まで楽しめる幅広い難易度をもつため、雑誌や新聞に掲載やゲームや携帯アプリとして、老若男女様々なユーザ層で親しまれている。

また、数独は制約問題という観点から計算機科学の分野において様々な研究が行われている。数独は3x3のブロックに区切られた9x9の正方形の枠内に1~9までの数字を入れるといった単純なルールで、一般的な問題では予め25箇所ほどのマスが数字で埋められている。McGuireらは数独が問題として成り立つためには、少なくとも17箇所のヒントが必要であることを解明した[3]。また、難易度と問題構造との関連については、未だ解明されておらず、数独問題の構造[2]との関連やヒント数との関連、人間が解く場合

の難易度を想定し、数独パズルの解法ロジックを分類してそれらの汎用性に順位をつけて解析するといった試みがある[4]。数独の難易度は幅広く用意されており、初心者から上級者まで、楽しめるよう設定されている。その一方で、ユーザのスキルやレベルに合った問題を提供することは非常に難しく、未だに実現できていない。数独はシンプルな問題であるが、空きマスとヒントとなる数字の情報だけでは、外観的にその難易度を推定することは困難である。現状では、ユーザは問題とともに記された(多くの場合は根拠のない)難易度を参考に問題を選択している。

そこで本研究では、数独の難易度推定を中心課題とし、数独を解くユーザの振る舞いや数独の解答結果をベースにユーザのスキルレベルと問題との関係性を評価した。ユーザの振る舞いは、解答時間などの他に、興奮や集中など思考の移り変わりの影響をうける皮膚電気活動を取得・記録した。これまで、思考ゲームを用いたユーザの思考過程の記録には、ユーザにその過程を発話させ、その内容を分析する発話プロトコル解析が一般的に用いられてきた。しかしこの手法は、発話を行うことで思考が妨げられるといった問題や、発話されない無意識的な思考を扱うことができないといった問題点が挙げられる。本研究では、「数独」にお

†1 北海道大学大学院情報科学研究科  
Hokkaido University

†2 公立はこだて未来大学  
Future University-HAKODATE

けるユーザの思考の分析手法として、意識的な行動として記録される発話ではなく、無意識的に絶えず変化し、偽りのない素直な反応である生体情報に着目した。特に、人間の思考過程の移り変わりや計算に対する集中の有無に強く影響を受ける皮膚電気活動を用いて、数独を解くユーザの振る舞いを評価する。また、その生体情報から得られた振る舞いが、どの程度実際のスキルと整合していたのかどうかを調べる簡単な評価実験を行った。

## 2. 皮膚電気活動

本研究では、数独を解答しているユーザの自然な状態を測定するために、ユーザにとって負担の少ない測定部位である掌に装着する装置として、皮膚電気活動に注目した(図1)。皮膚電気活動は、精神性発汗を電気的にとらえたものである。人の手掌や足底は、緊張や動揺などの心的興奮によって発汗を生じる。これらの精神性発汗は、自分では感じられないほどの微量のものから、手掌が湿ってしまうほどの大量の発汗まで様々である。それらの皮膚電気活動のうち一過性の反応を測定した SCR(Skin Conductance Response)は、交感神経支配下の汗腺活動を電気的に測定して、情動状態、認知活動、情報処理過程を評価する方法である。例えば、SCRは精神的ストレスや、感情の指標として、医学や心理学などで注目されている[5]。村井らはSCRを用いた授業評価を行い、主に授業における集中の度合いをSCRから評価した[6]。その結果、反応が極度に大きいときは、新奇性のある外部刺激に対する反射である定位反応を示し、新奇性が失われると反応は弱まることが示された。一方、反応が極度に小さいときは注意集中状態にあると考えられ、反応が出ている箇所だけではなく、SCRは反応が出ていない箇所にも注意を向ける必要があると指摘している。本研究においても、注意集中状態の判定として、SCRの無反応箇所の分析を行う。

## 3. 実験1

本研究では、「数独」におけるユーザの思考の分析手法として、生体情報による評価に着目した。特に、人間の思考過程の移り変わりや計算に対する集中の有無に強く影響を受ける皮膚電気活動を用いて、数独を解くユーザの振る舞いを評価し、その振る舞いが実際に実験で導いたユーザのスキルと整合していたかどうかを調べる評価実験を行った。

### 3.1 実験で用いた数独問題の難易度

実験は、以下の2つの数独問題を用いた。一般的に数独の問題の難易度は、空白のマス数や空白のマスに入る数字の候補数を中心に考えられている。空白のマス数が多いほど埋めなければならない数字が増えるため、単純にマス数が増える時間にも影響を与えられられている。また、空白のマスに入る数字の候補数が多いほど、考えられる数」

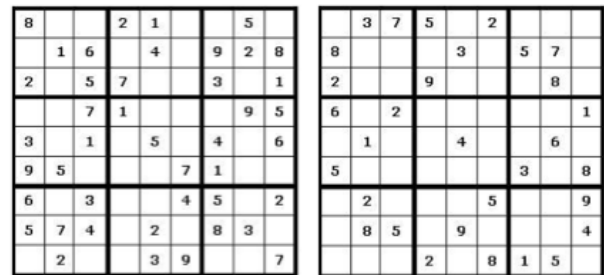


図3 実験で用いた数独問題 (左図:1問目, 右図:2問目)

Figure 3 Sudoku puzzle 1 and 2 using the first experiment  
 字が増えるため、解答により時間がかかると考えられている[2]。実験で用いた問題の難易度をChen[7]らの提案する「数独エントロピー」を用いて評価し、「世界基準ナンプレ300」[8][9]で用いられている公称難易度(8段階: Ultra Easy, Very Easy, Easy, Medium, Medium Plus, Hard, Very Hard, Spicy)と対応させたものを以下に示す。

表1 数独問題の難易度

Table 1 Sudoku puzzles difficulty Levels

	1問目	2問目
数独エントロピー	0.2389	0.6613
空きマス	38	50
全ての候補数	70	152
公称難易度	Ultra Easy	Easy

表1から、1問目と2問目を比較すると、「数独エントロピー」、「空きマス」、「全ての候補数」のどの項目も、2問目の方が高い値となっている。また「公称難易度」も「Ultra Easy」と「Easy」となっており、1問目よりも2問目の方が難しい問題であることがわかる。しかし、これらの指標は、どちらも問題を解き始める前の盤面の状態から数独の難易度を判定したものであり、実際に解いているユーザの状態やその過程を含めての難易度判定ではない。また、空白のマスが多ければ多いほど、単純に難易度が上がる仕組みではないと言われており[10]、公称難易度と、ユーザの体験が一致しているとは限らないと考えられる。

### 3.2 実験1概要

被験者を5名とし、数独を解答している際のユーザの振る舞いやSCRを測定した。具体的な実験手順は以下の通りである。まず、事前に数独に不慣れな人には、解答方法など数独に関する説明書を提示した。次に被験者への教示として、計2問(図1)の数独の問題を可能な限り早く解くことと、正答率や解答速度に応じて報酬を出す旨を伝えた。実験中、問題への解答が終了したときや疑問が浮かんだとき、間違いを修正できなくなり最初からやり直したいときには実験者にその旨を伝えさせた。実験は、周囲の環境の影響を考慮し、防音室で行った。

### 3.3 実験1結果

1問目と2問目の数独問題について、「解答時間」、被験

者の「SCR 反応量」, 「SCR 無反応量」の平均を表 2 にまとめた。「解答時間」は難易度を推定する指標として重要である。また、「SCR 反応量」は、リラックス状態では全く出現しないが、焦りなどを感じると強い反応を示すため、判断の難しい箇所に直面した際に、被験者の焦りなどの心的興奮が測定できると考えた。また、「SCR 無反応量」は、注意集中状態の判定として用い、SCR の反応があまり確認されない場合、注意集中が持続していると判断する。実験 1 の結果を表 2 に示す。

表 2 実験 1 結果 (被験者の平均)

Table 2 Results of first experiment ( average )

	1 問目(Ultra Easy)	2 問目(Easy)
解答時間	282.8 (s)	668.2 (s)
SCR 反応量	143.72	119.88
SCR 無反応量	73	79.4

実験 1 結果(表 2)より、2 問目は全体の解答時間が長く、SCR 反応量は 1 問目の方が高い値を示していることがわかる。また、SCR 無反応量が 2 問目の方が高い値を示している。このことから、全体的に 2 問目の解答時の方が、1 問目に比べて集中できており、焦りなどの心的興奮も減少していたと考えられる。これについては、数名の被験者から「実験が始まった時(1 問目)は緊張していた」との内省が得られており、実験環境を意識してしまったことが強く影響していたと考えられる。次にそれぞれの項目について、それぞれの被験者(A-E)ごとに解析した結果を以下に示す。

表 3 解答時間の評価 (被験者ごと)

Table 3 Results of game time ( with each participant )

	A	B	C	D	E
1	392 (s)	232 (s)	263 (s)	247 (s)	280 (s)
2	1098 (s)	558 (s)	609 (s)	681 (s)	395 (s)

表 4 SCR 反応量の評価 (被験者ごと)

Table 4 Results of SCR Values ( with each participant )

	A	B	C	D	E
1	33.6454	134.966	263.19	116.919	169.964
2	55.0951	70.7153	268.101	73.5979	131.885

表 5 SCR 無反応量の評価 (被験者ごと)

Table 5 Results of no reaction of SCR Values

	A	B	C	D	E
1	0.88	0.7	0.63	0.86	0.58
2	0.89	0.87	0.67	0.87	0.67

これらの結果に基づいて、それぞれのユーザの振る舞いを相対的に評価するために、実験結果から得られたそれぞれの項目について、順位を設定しその順位を元に、各ユーザのポイントを計算し、ユーザの評価を行う。解答時間が短

時間であるほど、ユーザの数独スキルは高いと考え、その順位に応じて上から 5 から 1 のポイントを割り当てた。また、SCR 反応量については、その反応量が少ないほど、解答中に焦りや興奮を感じていないと考えられるため、スキルレベルが高いと判断し、その順位に応じ、5 から 1 ポイントを付与した。SCR 無反応量は、その量が少ないほど、問題の解答に注意集中できていると考えられるため、その順位に応じて 5 から 1 のポイントを付与した。各項目におけるユーザ順位と、ポイントを以下に示す。

表 6 スキルレベルの設定 (被験者ごと)

Table 6 Setting of User's Skill

	順位	ポイント
解答時間	E>B>C>D>A	5>4>3>2>1
SCR 反応量	B<D<E<C<A	5<4<3<2<1
SCR 無反応量	B>E>C>D>A	5>4>3>2>1

表 6 は、「解答時間」、被験者の「SCR 反応量」、「SCR 無反応量」について、それぞれの被験者のデータを順位付けしたものと、その順位に応じて付与するポイントを示した。ポイントは、解答時間が早いほど高く、SCR の反応が少ないほど高く設定した。表 6 の要領で各ユーザのポイントを計算した相対的なスキルレベルの結果を表 7 に示す。

表 7 ユーザのスキルレベル (被験者ごと)

Table 7 User's Skill Level ( with each participant )

	A	B	C	D	E
計	7	12	7	8	11

表 7 から、相対的なユーザスキルは、B>E>D>C=A の順に高い結果となった。

## 4. 実験 2

実験 1 の被験者に数独の問題を解いてもらい、実験 1 で導き出したユーザのスキルとを比較する。

### 4.1 実験 2 概要

実験は、以下の数独問題を用いた。

			9	4	1
	7	5	6		
	3	9		6	7
2			1	9	7
	4			1	
3	5	7			2
	9	3		7	1
		4	8	2	
4	2	9			

図 3 実験で用いた数独問題

Figure 3 Sudoku puzzle using the second experiment

図 3 の数独問題について Chen[7]らの難易度を分析したものを表 8 に示す。実験 1 で使用した図 2 の数独問題に比べて、数独エントロピー、空きマス、候補数の全ての項目で高い数値を示した。

表 8 数独問題の難易度

Table 8 Sudoku puzzles difficulty Levels

	3 問目
数独エントロピー	0.7033
空きマス	51
全ての候補数	161
公称難易度	Easy

実験 2 は、実験 1 に参加した 5 名の被験者に対し、図 3 の数独の問題を可能な限り早く解くことと、正答率や解答速度に応じて報酬を出す旨を伝えた。実験中、問題への解答が終了したときや疑問が浮かんだとき、間違いを修正できなくなり最初からやり直したいときには実験者にその旨を伝えさせた。実験は、周囲の環境の影響を考慮し、防音室で行った。

#### 4.2 実験 2 結果

表 9 解答時間の評価 (被験者ごと)

Table 9 Results of the game time ( with each participant )

	A	B	C	D	E
3	1140 (s)	346 (s)	608 (s)	1023 (s)	652 (s)

表 10 SCR 反応量の評価 (被験者ごと)

Table 10 Results of SCR Values ( with each participant )

	A	B	C	D	E
3	374.3763	101.5771	411.8198	195.7653	108.4799

表 11 SCR 無反応量の評価 (被験者ごと)

Table 11 Results of no reaction of SCR Values

	A	B	C	D	E
3	0.69	0.77	0.58	0.5	0.68

これらの実験結果から、表 12 のルールに合わせてスキルレベルを導く。

表 12 スキルレベルの比較 (被験者ごと)

Table 12 Setting of User's Skill

	順位	ポイント
解答時間	B>C>E>D>A	5>4>3>2>1
SCR 反応量	B<E<D<A<C	5<4<3<2<1
SCR 無反応量	B>A>E>A>B	5>4>3>2>1

これらの結果から、ユーザの相対的なスキルレベルは、B>E>D>C>A となった。具体的なポイントを、実験 1 と比較したものを以下 (表 13) に示す。

表 13 スキルレベルの比較 (被験者ごと)

Table 13 User's Skill ( with each participant )

	A	B	C	D	E
実験 1 (ユーザ評価)	7	12	7	8	11
実験 2 (ユーザ評価)	4	12	9	10	10

この結果から、それぞれの被験者の相対的なレベルは、難易度の違う問題において、ある程度一致していることが示

された。具体的には、B と E の被験者は、難易度の違う数独問題においても、同じような結果が得られており、その他の被験者においても、相対的なスキルレベルの順位はほぼ変化していない。このことから、実験 1 の最初の 2 つの問題で得られたユーザの相対的なスキルレベルが、その他のレベルの問題を解いた場合に得られたスキルレベルとで、ほぼ一致していたことが示された。しかし、同じ難易度を感じられるような問題であっても、あるユーザにとっては苦手なものや得意なもの、間違えやすいものなどは全く違うため、ユーザスキルについては、より詳細に分析する必要がある。

#### 5. おわりに

本研究では、数独の難易度推定を中心課題とし、数独を解くユーザの振る舞いや数独の解答結果をベースにユーザのスキルレベルと問題との関係性を評価した。ユーザの振る舞いは、数字を記入するタイミングなどの行動の他に、興奮や集中など思考の影響をうける皮膚電気活動を取得・記録した。今回得られたユーザのスキルレベルを活用し、本研究の中心課題であるユーザベースの数独問題の難易度推定を行いたいと考えている。このような研究は、様々な思考ゲームにおいて、ユーザのレベルに合うコンピュータのレベル調整や、難易度のリコメンデーションなど、様々な応用可能性をもつ。

#### 参考文献

- 1) Rory Sobolewski, Richard B. Reilly, Rory Sobolewski, Simon Finnigan, Paul Dockree. : Monitoring of Cognitive Processes in Older Persons, Proceedings of the 4th IEEE International Conference on Neural Engineering 4th IEEE International Conference on Neural Engineering, pp132-135 (2009).
- 2) 前田一貴, 奥乃博:数独の問題作成支援システムの設計と開発, 情報処理学会第 70 回全国大会, 4ZH-4, Mar. 2008.
- 3) Gary McGuire, Bastian Tugemann, Gilles Civario : There is no 16-Clue Sudoku: Solving the Sudoku Minimum Number of Clues Problem (2012).  
<http://arxiv.org/abs/1201.0749v1>
- 4) 土出智也, 真貝寿明: 数独パズルの難易度判定- 解法ロジックを用いた数値化の提案-, 大阪工業大学紀要 (理工編) 56 巻, p1-18. (2011).
- 5) 山崎勝男, 藤澤清, 柿木昇治, 宮田洋: 新生理心理学 3 巻, 北大路書房. (1998).
- 6) 村井護晏: 皮膚抵抗反応による授業評価の可能性について, 日本教科教育学会誌 vol.14, pp145-151. (1990).
- 7) Zhe Chen: Heuristic Reasoning on Graph and Game Complexity of Sudoku, (2009).  
<http://arxiv.org/abs/0903.1659v1>
- 8) 世界基準ナンプレ 300 Vol.1, 笠倉出版社. (2010)
- 9) 世界基準ナンプレ 300 Vol.2, 笠倉出版社. (2010)
- 10) 数独データランド, 数独通信 Vol.16, ニコリ, 2009 年 11 月 10 日. (2009).
- 11) Foley, J. D. et al.: Computer Graphics - Principles and Practice, System Programming Series, Addison-Wesley, Reading, Massachusetts, 2nd edition (1990).