



# モニタ上の 3D アバタによる環境的文脈操作を用いた 記憶支援手法の効果

瑞穂嵩人<sup>1)</sup>, 鳴海拓志<sup>1)</sup>, 葛岡英明<sup>1)</sup>

1) 東京大学大学院情報理工学系研究科

(〒 113-8656 東京都文京区本郷 7-3-1, {takato, narumi, kuzuoka}@cyber.t.u-tokyo.ac.jp)

**概要:** ビデオ会議などではフィルタ機能などによって利用者の見た目を変容させることができる。本研究では、誰といるかなどの環境的文脈が記憶に影響することに注目し、モニタ上の発言者の見た目を多様なアバタに切り替えることで記憶を支援しようと考えた。lab.js を用いたオンライン実験で、100 個の単語を 25 語ずつのサブリストに分けて、全てのサブリストを同じアバタが読み上げる条件 (1-Avatar) と各サブリストを異なるアバタが読み上げる条件 (4-Avatar) の想起成績を比較した。96 名の参加者間比較の結果、覚えらる単語の総数には条件間に差がなかったが、ある単語がどのサブリストに含まれていたかを判定する課題において 4-Avatar 条件の方が有意に成績が高かった。

**キーワード:** マルチ文脈効果, 環境的文脈依存記憶, アバタ, 遠隔講義, 遠隔会議

## 1. はじめに

オンライン講義やビデオ会議において、利用者は自身の表情や身振りを WEB カメラを通じて相手に共有することができる。近年では画像処理技術の発達に伴って、自身の姿を実際とは異なる姿に変容させることも容易になった。例えば、Nakazato et al. [1] は WEB カメラ映像上のお互いの表情を笑顔に変形させると通常時よりも多くのアイデアを出すことが出来ると報告している。また、雨宮ら [2] は「どの先生の授業を受けてみたいか?」という質問によって評価が高いアバタと低いアバタを選定し、受講希望度が高い講師アバタで講義をすると受講希望度の低い講師アバタを用いた場合に比べて匿名チャットにおける発言数が増加することを明らかにした。以上のような操作は対面時には不可能な操作である。このように、遠隔コミュニケーションシステム特有の効果を実現し、活用していくことが期待される。

本研究は記憶の代表的な性質の一つである環境的文脈依存性に注目する。環境的文脈依存性とは、記憶には場所や匂い、同伴者などの環境情報 (環境的文脈) が取り込まれており、後で想い出すときにそれらの環境的文脈が手がかりとして使用される性質である [3]。本研究では特に Smith (1982) [4] のマルチ文脈効果に注目する。Smith は、100 個の英単語の記憶課題において、全て同じ部屋で覚える条件 (1-Room)、50 語ずつ部屋を変えて覚える条件 (2-Room)、25 語ずつ部屋を変えて覚える条件 (4-Room) を参加者間比較した。いずれとも異なる部屋で自由再生テスト (覚えていた単語を出来るだけ多く回答する) を実施した結果、正答数は 1-Room 条件が 19.9 語、2-Room 条件が 21.9 語、4-Room 条件が 26.2 語であり、複数の部屋で覚える方が正答数が高かった。これは単語を想い出すときに手がかりとして使用できる環境的文脈が多いほど想起が容易になるた

めであると考えられている (符号化変動仮説)。

本研究では、部屋を操作する代わりに、実験者の見た目を様々なアバタに変容することを通じて「誰が発言しているか」という環境的文脈を操作する手法を提案し、同様のマルチ文脈効果が得られるかを検証する。モニタ上のアバタが環境的文脈として記憶に作用するかは解明されていないが、(1) モニタ上の視覚情報 (背景ビデオや写真など) は環境的文脈として記憶に作用する [5]、(2) 実験者が誰か (外見や性別) や何人で一緒に実験を受けるかといった要素も環境的文脈として記憶に作用することが示唆されている [6] ことから、提案手法が有効である可能性が高いと考えた。

本研究では、100 個の日本語名詞の記憶課題において、単語を読み上げる実験者の見た目をアバタに変容し、全て同じアバタで読み上げる条件 (1-Avatar) と 25 個ずつ異なるアバタで読み上げる条件 (4-Avatar) を参加者間比較する (図 1)。符号化変動仮説に基づき 4-Avatar 条件の方が 1-Avatar 条件よりも記憶成績が高いと考えられる。評価指標として、どれだけ多くの単語を思い出せるかを自由再生テスト、ある単語がいつ読み上げられたかを思い出せるかをソースモニタリングテストで計測する。前者の成果はオンライン講義などの教育分野において学習効率の向上に活用できる可能性があり、後者の成果は遠隔会議や遠隔診療などの場面で記憶の混濁を防ぐ応用が考えられる。

## 2. 実験方法

本実験はオンラインで実施した。参加者は自宅などから実験用 WEB サイトにアクセスし、画面上の指示に従った。

### 2.1 実験参加者

健全な男女 96 名 (平均 24.5 ± 4.4 歳, 男性 56 名, 女性 40 名) が参加した。参加者は全員日本語母語話者であった。

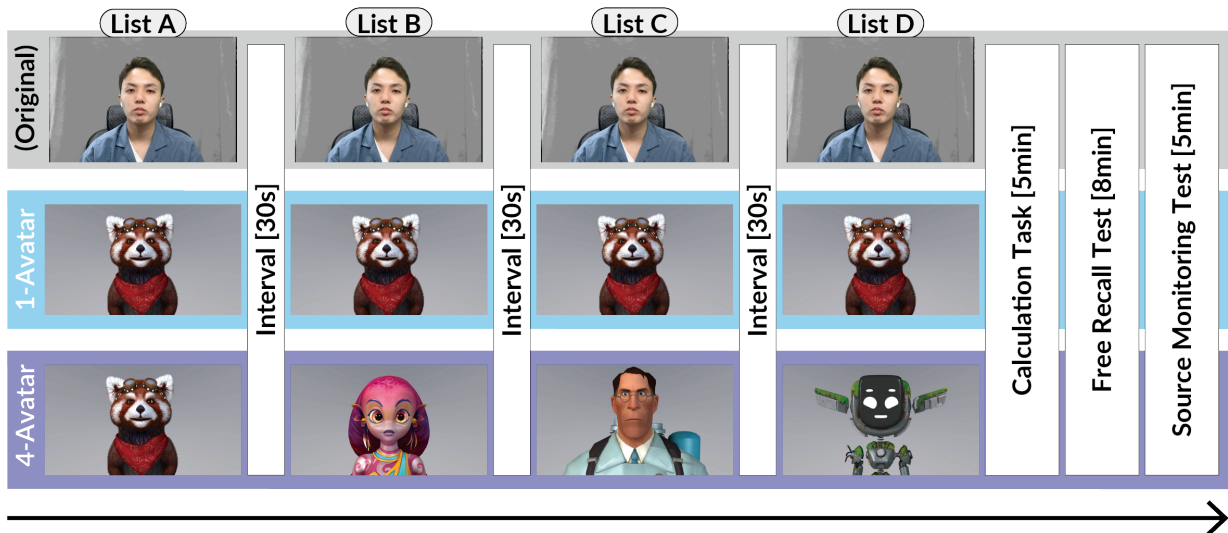


図 1: 参加者は実験者が読み上げる単語リストを 4 本覚え (List A, B, C, D), インターバル後に想起テストを受けた。全て同じアバタで読み上げられる条件 (1-Avatar) とそれぞれ異なるアバタで読み上げられる条件 (4-Avatar) を比較した。

## 2.2 実験条件

実験はアバタ要因 × 教示要因 × 系列内位置要因の三要因混合計画で行われた。アバタ要因は参加者間要因で、全ての単語を同じアバタが読み上げる条件 (1-Avatar) と、25 単語ずつ異なるアバタが読み上げる条件 (4-Avatar) の二水準であった (図 1)。教示要因は参加者間要因で、単語とアバタを関連付けて覚えるよう教示を与えられる条件 (Instructed) と、そのような教示を与えられない条件 (Uninstructed) の二水準であった。系列内位置要因は参加者内要因で、最初の 25 単語 (List A), 次の 25 単語 (List B), その次の 25 単語 (List C), 最後の 25 単語 (List D) の四水準であった。参加者は、年齢および性別のバランスを考慮した上で、アバタ × 教示の 4 グループにランダムに割り振られた。

## 2.3 記銘材料

参加者は熟知価 4.00 – 4.99 の三音節の日本語名詞 100 語 [7] の記憶課題に取り組んだ。熟知価とはどの程度馴染みのある単語かを示す値で、熟知価が高いほど覚えやすいことが知られている。100 語を 25 語ずつのサブリストに分けて、それぞれ List A, B, C, D とした。このとき各サブリスト内で単語同士に意味的な関連がないように分配した。図 1 に示すように、参加者は List A, B, C, D の順で各サブリストを覚えた。各サブリスト内での単語順序は、実装の都合、あらかじめランダム化した一通りであった。

## 2.4 実験システム

オンライン実験システムの構築には lab.js (<https://lab.js.org/>) を使用した。lab.js は、HTML, CSS, JavaScript などに関する知識がなくても GUI で実験システムを作成できる lab.js Builder が提供されていることや、中上級者はコーディングを追加して機能をカスタマイズすることが出来る点で注目されている [8]。本実験では、これらの実装上の利点とミリ秒のオーダーでの時間管理 (想起テストの制限時間など) が可能な点に注目し、lab.js を使用した。

参加者は lab.js で構築された WEB ページ上で動画を視

聴し、アバタが読み上げる単語を覚えた。各動画は以下の手順で作成した。まず、Macbook の Photo Booth アプリで第一著者が単語リストを読み上げる様子を撮影した (図 1, 上段)。マイクは SHURE MV7 を用いた。各動画は「これから単語を読み上げるので出来るだけ多く覚えてください。順序は問いません。」という台詞から始まり、続いて 25 個の単語が 3 秒/単語の速度で連続で読み上げられるという構成であった。各サブリストに対応する合計 4 本の動画が作成された。次に、Animaze (<https://www.animaze.us/>) を用いて各動画の第一著者の見た目をアバタに変換した。Animaze は、WEB カメラ映像からユーザの頭の動きや表情を検出し、モニタ上のアバタ (キャラクタ) に反映するソフトウェアである。アバタは Animaze が提供する中から意味や色合いが互いに異なるように選び、レッサーパンダ、エイリアン、ヒト、ロボットの 4 種類であった (図 1)。アバタの背景には交絡を防ぐため灰色を用いた。4 本の元動画をそれぞれ 4 種類のアバタに変換し、合計で 16 本の動画が作成された。

各動画は lab.js で作成された WEB ページ上に埋め込まれ、ブラウザをフルスクリーンにすれば画面一杯に表示されるような大きさで提示された。実装の都合、動画の再生位置や再生速度などを操作するインタフェースを非表示にすることが出来ず、参加者はこれら进行操作出来る状況にあったが、事前にこれらを行わないように指示し、事後に該当する WEB ページの滞在時間が動画の長さよりも短い場合には解析から除去することで対応した。

## 2.5 手順

まず配布された説明書を読み、Google Form で実験参加に同意する旨を回答した。続いて実験用 WEB サイトにアクセスし、以降は画面上の指示に従って実験を進めた。

はじめに、ブラウザをフルスクリーンにすることや音量の調整などの実験実施に関する注意事項等を確認した。また、実験の構成について、動画を 4 本視聴し合計 100 単語を覚えることと、インターバル後に 2 種類の想起テストを

実施することが伝えられた。ただしテストの詳細については知らされなかった。その後、Instructed 条件群には「単語とアバタの両方を関連付けながら覚えていってください。これは後で行うテストに役立つ可能性があります。」と教示が与えられた。他方、Uninstructed 条件群にはそのような教示は与えられなかった。参加者は List A, B, C, D の順番で動画を視聴し単語を覚えた。n 番目の動画の WEB ページには「セッション n」と表示されていた。サブリスト間は 30 秒で、このとき画面にはプログレスバーが表示された。

4 本目の動画を視聴後、参加者は 5 分間の計算課題に取り組んだ。これはインターバル中のリハーサル（覚えた単語を頭の中で反復すること）を抑制するための課題であった。例えば「 $1+2 \times 3 =$ 」のように 3 つの桁の数の和算や積算からなる計算式が一つ表示され、参加者はキーボードを用いて回答した。回答を確定するか、5 秒経過すると次の計算式に切り替わった。

計算課題の後、8 分間の自由再生テストを行った。参加者は「覚えている単語を全て回答してください。順序は問いません。」と指示され、縦に並んだ 100 個の回答欄に、上から順に一単語ずつキーボードで回答を入力した。ページの上にはテストの残り時間がプログレスバーで表示されていた。次にソースモニタリングテストを行った。100 語のうち一つが画面中央に表示され、「この単語は何番目のセッションで覚えましたか？」という問いに 1 から 4 のいずれかの数字をキーボードで回答した。これを 100 単語全てについてランダムな順序で行った。時間制限はなかった。最後に実験全体に関するアンケート（当日の体調、実験を実施する上でわからなかった点、感想など）に回答した。

## 2.6 評価指標および仮説

本実験の評価指標は、(1) 自由再生テストの正答数、(2) ソースモニタリングテストの正答数であった。符号化変動仮説に基づき、想起時の手がかりが多い 4-Avatar 条件の方が 1-Avatar 条件よりも想起成績が向上すると考えられる。

**H1** 自由再生テストの正答数は 4-Avatar > 1-Avatar

**H2** ソースモニタリングの正答数は 4-Avatar > 1-Avatar

また、森井ら [9] らは、モニタ上の単語とその背景ビデオを関連付けて覚えさせる教示を与えると、そのような関連付け教示を与えない場合に比べて背景ビデオが想起に与える影響が増大することを示した。同様に、本実験において単語とアバタの関連付け教示を与えることは、参加者が想起時にアバタを手がかりとして使用する傾向を強め、4-Avatar 条件優位の傾向を強化すると考えられる。

**H3** 自由再生テストの正答数において、4-Avatar 条件優位の傾向は Instructed 条件の方が Uninstructed 条件よりも大きい

**H4** ソースモニタリングの正答数において、4-Avatar 条件優位の傾向は Instructed 条件の方が Uninstructed 条件よりも大きい

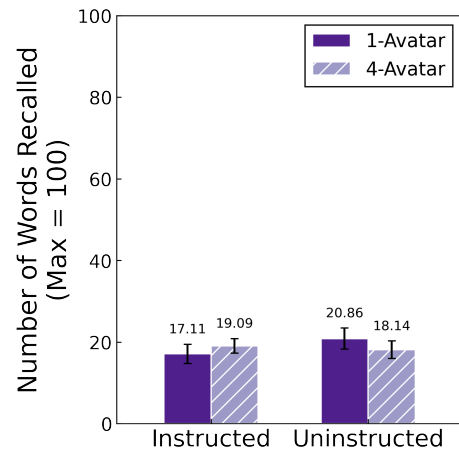


図 2: 自由再生テストの正答数 ( $M \pm SE$ )。

## 3. 結果

実験システムの不具合があった参加者が 6 名、各動画ページの滞在時間が動画の長さよりも短かった参加者が 6 名、集中力をなるべく統制するため早朝および深夜を避けて（9 時から 21 時まで）に実施するよう通知していたところ 5 時に実施した参加者が 1 名いた。上記の 13 名を除外し、残りの 83 名のデータについて解析を実施した。83 名の内訳は、Instructed 1-Avatar 条件群が 19 名、Uninstructed 1-Avatar 条件群が 21 名、Instructed 4-Avatar 条件群が 22 名、Uninstructed 4-Avatar 条件群が 21 名であった。

### 3.1 自由再生テストの正答数

図 2 に自由再生テストの正答数を示す。整列ランク変換後 [10]、アバタ  $\times$  教示  $\times$  系列内位置の三元配置分散分析を実施した結果、系列内位置の主効果のみ有意で ( $F(3, 237) = 9.86, p < .001, \eta_p^2 = 0.11$ )、その他の主効果および交互作用はいずれも有意でなかった ( $ps > .1$ )。系列内位置の主効果について、Holm の方法で ART-C [11] による多重比較を実施した結果、List B の正答数が List A ( $t(237) = 5.19, p < .001$ )、List C ( $t(237) = 3.76, p = .001$ )、List D ( $t(237) = 3.74, p = .001$ ) よりも有意に低かった。

### 3.2 ソースモニタリングテストの正答数

図 3 にソースモニタリングテストの正答数を示す。整列ランク変換後 [10]、アバタ  $\times$  教示  $\times$  系列内位置の三元配置分散分析を実施した結果、アバタの主効果 ( $F(1, 79) = 4.58, p = .04, \eta_p^2 = 0.05$ )、系列内位置の主効果 ( $F(3, 237) = 2.76, p = .04, \eta_p^2 = 0.03$ )、アバタ  $\times$  教示の交互作用 ( $F(1, 79) = 4.27, p = .04, \eta_p^2 = 0.05$ ) が有意で、その他の主効果および交互作用はいずれも有意でなかった ( $ps > .1$ )。アバタ  $\times$  教示の交互作用について、Holm の方法で ART-C [11] による多重比較を実施した結果、Instructed 1-Avatar 条件と Instructed 4-Avatar 条件 ( $t(79) = 3.15, p = .01$ )、Instructed 4-Avatar 条件と Uninstructed 1-Avatar 条件 ( $t(79) = 2.80, p = .03$ ) に有意差が認められた。Instructed 4-Avatar 条件と Uninstructed 4-Avatar 条件の差は有意傾向で ( $t(79) = 2.52, p = .05$ )、それ以外は有意でなかった ( $ps > .1$ )。また、系列内位置の主効果について、Holm の

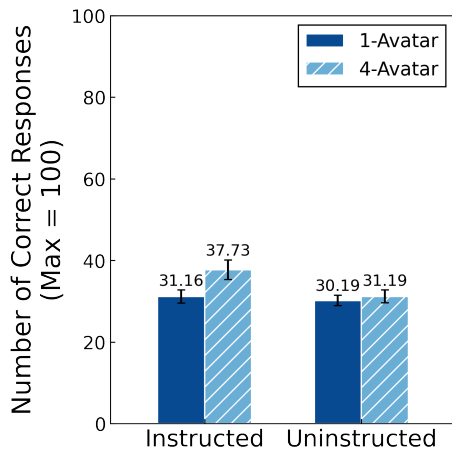


図3: ソースモニタリングテストの正答数 ( $M \pm SE$ ).

方法で ART-C [11] による多重比較を実施した結果、いずれの組み合わせでも有意差は認められなかった ( $ps > .1$ ).

#### 4. 考察

自由再生テストの正答数について、1-Avatar 条件と 4-Avatar 条件に有意差は認められず、仮説 H1 は支持されなかった。また、アバター × 教示の交互作用も認められず、仮説 H3 は支持されなかった。以上の理由について、(1) 参加者がモニタを注視していなかった可能性と (2) 単語とアバターの関連付けが困難であった可能性があったと考えている。まず (1) について、単語が聴覚提示であったため画面を見なくても記憶課題を遂行できた。こうした事態を防ぐために、事前に実験中は画面を見ているように指示していたが、参加者からは「無意識のうちに画面外を見るなどしてアバターはあまり見ていなかった」とコメントがあった。この場合、アバターの情報が記憶に取り込まれず、結果としてテスト時に手がかり情報として使用することが出来ない。ところで聴覚提示を選択した理由は Smith (1982) [4] に倣ったからであるが、彼が用いた部屋という環境的文脈はどこを向いていても提示されている点で今回の提案手法と違いがあったと考えられる。(1) については、単語をアバターと並べて視覚提示し共に記憶に取り込まれるようにした場合について追加で検証したいと考えている。次に (2) について、複数の参加者から「アバターと何の関連もない単語を結びつけるのは難しかった」とコメントがあった。関連付け教示の効果を明らかにした森井ら [9] は日常風景のビデオという環境的文脈を用いたため、例えば学校やスーパーといった場所の情報が含まれている分、関連付けが容易であったという違いがあったと考えられる。関連付け教示が環境的文脈の効果を高める条件についてはさらなる調査が必要である。

ソースモニタリングテストの正答数について、Instructed 条件でのみ 4-Avatar 条件が 1-Avatar 条件よりも有意に正答数が高く、仮説 H2 は一部支持され仮説 H4 は支持された。Uninstructed 条件でアバター条件間に有意差が認められなかったのは、自由再生テストの結果と同様に、アバターをあまり見ていなかったためだと考えられる。本結果は、例えば遠隔会議に対しては、発言者のアバターを会議ごとに切り

替えることで、ある話題がどの会議で話し合われたのかを区別することが容易になるといった応用が考えられる。ただしその際に聞き手はアバターを注視し内容とアバターを関連付けておく必要がある可能性がある。

本実験は lab.js を用いたオンライン実験であった。システムのトラブルなどによって解析から除外した参加者が多数出てしまった点については改善の余地がある。また統制が不十分な点についても留意して結果を解釈する必要がある。他方、1ヶ月程度の期間で96名という人数の実験を実施できた点は、オンライン実験による研究活動の加速が期待できる。また、今後は本研究の知見を活かし、遠隔講義や会議といった実際の現場において3Dアバターによる環境的文脈操作を用いた記憶支援の実証も考えている。

**謝辞** 本研究の一部は JST ムーンショット型研究開発事業 (JPMJMS2013)、科研費挑戦的研究(萌芽) (21K19784)、NEDO の委託業務 (JPNP21501015-0) の支援を受けた。

#### 参考文献

- [1] N. Nakazato et al.: Smart Face: Enhancing creativity during video conferences using real-time facial deformation, *Proc. of CSCW '14*, 75–83, 2014.
- [2] 両宮ら: 遠隔講義における講師アバターの見かけによって変化する受講希望度が授業への積極的参加行動に与える効果—オンライン授業への導入事例—, *日本バーチャルリアリティ学会論文誌*, 26(1), 86–95, 2021.
- [3] S. M. Smith and E. Vela: Environmental context-dependent memory: A review and meta analysis, *Psychonomic Bulletin & Review*, 8(2), 203–220, 2001.
- [4] S. M. Smith: Enhancement of recall using multiple environmental contexts during learning, *Memory & Cognition*, 10(5), 405–412, 1982.
- [5] S. M. Smith and I. Manzano: Video context-dependent recall, *Behavior Research Methods*, 42(1), 292–301, 2010.
- [6] T. Isarida and T. K. Isarida: Effects of simple and complex-place contexts in the multiple context paradigm, *The Quarterly Journal of Experimental Psychology*, 63(12), 2399–2412, 2010.
- [7] 小柳ら: 日本語三音節名詞の熟知価, *心理学研究*, 30(5), 357–365, 1960.
- [8] 大杉, 小林: GUI ベースの web 実験作成ツール (lab.js) の紹介と実践, *認知心理学研究*, 19(1), 1–15, 2021.
- [9] 森井ら: ビデオ文脈依存再生における学習方法とテスト方法の効果, *基礎心理学研究*, 34(2), 229–238, 2016.
- [10] J. O. Wobbrock et al.: The aligned rank transform for nonparametric factorial analyses using only anova procedures, *Proc. of CHI '11*, 143–146, 2011.
- [11] L. A. Elkin et al.: An aligned rank transform procedure for multifactor contrast tests, *Proc. of UIST '21*, 754–768, 2021.