アバターの種類や数が視聴者の記憶想起に及ぼす影響
The Effects of Avatar Type and Number on Viewers' Memory Recall

瀬島 章仁[†]，安田 哲也[‡]，小林 春美[§*]
Akihito Sejima, Tetsuya Yasuda, Harumi Kobayashi
[†]東京電機大学大学院，[‡]東京大学，[§]東京電機大学
Graduate School of Tokyo Denki University, The University of Tokyo, Tokyo Denki University
h-koba@mail.dendai.ac.jp

概要
近年 AI 利用など人間以外のアナウンサーによるニュースの読み上げが広がっている. 本研究では, ニュース映像に登場するアバターの数や種類（人間, 動物）が視聴者のニュースで使用された語に関する記憶に与える影響を調べた. アバターの種類や数にかかわらず, 使用された語に関する記憶は比較的よく保たれていた. 一方, 1 体のアバターで伝える場合に比べ, 4 体のアバターが次々に登場する場合は, 視覚的注意が分散し映像に関係する語も誤って選ばれやすくなる傾向がみられた.

キーワード : 非言語的要素, 視線、口の動き、動物アバター, 記憶想起

1. 目的

本研究の目的は, YouTube でよく見られる情報発信動画を模し, アバターを利用して, その動画の印象や提示された情報に関する視聴者の記憶を調べることにより, 適切なアバターの特徴を調べることである. アバターを使用した動画は YouTube などで多くみられ, 全国各地で人気コンテンツとなっている. ニュース動画は現代社会において情報伝達の主要な手段となり, その中で AI がニュースの読み上げに利用される例が増加している. しかし, AI によるニュースの読み上げのような純粋な音声情報だけでは, 視聴者が視聴するためのモチベーションの維持・向上が問題となることが考えられる. アバターを導入して視覚的な情報を提供することにより, 視聴者が情報を効果的に受容しやすくなる可能性があるが, どのように導入すべきかについてはあまり検討されていない.

検討すべき問題として第一に, 情報提示をするアバターの数の問題がある. Mizuho et al. (2023)では, バーチャル・オムニバス講義と呼ばれる新しい遠隔講義形態が提案された. この講義は, 多数でオムニバス講義を行うというわけではなく, 1 人の講師だが自身のアバターを変えることで複数の講師で疑似的にオムニバス講義を行うというものである. また, この講義では人間アバターだけでなく動物アバターやロボットのようなアバターなども使用された. この研究では、異なる環境での学習が記憶力を向上させることが示されていることを基に, 複数の講師から仮想的に学習することでも同様の効果があるか調べるものであった. 90 分間の講義を 4 つのセッションに分け, セッションごとに異なるアバターを使って教える「4 アバター状態」と, すべてのセッションで同一のアバターを使用する「1 アバター状態」を比較する実験を行った. 結果, 講義後の理解度テストの点数は, 1 アバター状態よりも 4 アバター状態の方が有意に高いことが示された. これは, 複数のアバターで講義を行うことで, 学生が講義に注意しやすくなる効果が得られる可能性を示唆した.

第二に、視線や口の動きといった、情報の伝達において非言語的な情報に関する問題がある. なかでも人の視線は相手との会話促進などに影響を与えることが知られている (大坊, 1998). 他者の瞬間的な視線移動は, その方向に対しての注意を促すことがわかっている (Driver et al., 1999). また, 顔が提示された場合に直視の方が逸視よりもその顔を再生しやすいことがわかっている(Mason et al., 2004). 学習効果に関しては、Pi et al. (2020)がビデオ教示場面を調べており, 教材に視線を向けた方がアイコンタクト（直視）よりも学習効果が増加したという知見を示している. そこでアバターの助けが参加者の注意を促すという仮定を基に, 視線と口の動きがニュース内容の記憶に及ぼす効果を調べることにした.

本研究においては人間に加え動物アバターも使用することにした. その理由は, 動物は一般に視聴者の興味をひきやすいため, 人間アバターによる効果と比較したかったからである. 動物アバターでは容姿だけでなく, 口の動きに特徴がある. 人間アバターでは, 口の動き方と話している言葉の音声がおおむね一致しているので違和感なく聞きやすいため, 理解しやすいと考えられる. 一方, 動物アバターの場合は口の形状が人間と異なっており, 口の動きが「パクパク」と口の開閉のみで単調になりがちであるため, 人間アバターに比べて理解しにくい可能性がある.

本研究の予測では, すべてのセッションで同一のアバターを使用する「1 アバター状態」よりもセッションごとに異なるアバターを使って教える「4 アバター状態」の方が記憶の程度（記憶度）は高いと予測した. これは動物・人間いずれのアバターでも, 講義中にアバターが変化することでセッションごとに講義が新鮮なものとしてされ, 情報の整理がしやすくなる可能性を考えたからである. また, 人間の口の動きはより馴染みがあるため, 動画の記憶度が動物のアバターよりも高いと予測した.

本研究では, 動画の記憶度を測定する指標として単

語選択テストを用いた. 具体的には, 実験刺激であるニュース動画の視聴後, 30 個の単語が提示されその中から「実際に動画内で発せられた単語 (使用語)」を 10 個選択するよう参加者に求めた. このテストは, 参加者がどれだけ正確に講義内容を記憶していたかを測定するためのものであった. 提示された単語の内訳は, 実際にニュースで使用された単語 (使用語) が 10 個, 使用された語に関係ある語 (使用関連語) が 10 個, 動画には映像として登場するが, 音声とは無関係な語 (映像関連語) が 10 個とした. このように異なるタイプの語を含めた理由は, 参加者が単に動画のテーマに基づいた推測によって選択肢を選んでいるのか, あるいは正確に発語を記憶して使用されたと思った単語を選んでいるのかを区分するためである. 例えば, 使用関連語を選ぶ傾向が他の語に比べて高ければ, 動画内容について全体的に理解しているが正確な言語情報の記憶は曖昧であることを示唆する. 一方で, 使用語の選択率が高ければ, 言語情報についてより正確な記憶がされていると判断できる. 映像関連語を選ぶ傾向が高ければ, 動画に注意を向ける一方, 言語情報についての記憶は曖昧であることを示唆する.

2. 方法

2.1. 参加者

理工系大学に通う 18 歳から 24 歳の学生 15 人 (男性 10 人, 女性 5 人 : 平均年齢 20.2 歳) であった.

2.2. 条件

全ての刺激動画は顔方向の動きと口の動きが有るものとした. アバターが人間または人間ではない動物が解説するニュース動画となっている. アバターの数条件とアバターの種類条件をそれぞれ 2 水準で設けた. アバターの数条件は, 1 つのニュース動画を 4 つのセッションに分け, セッションごとに異なるアバターを使って教える「4 アバター状態」と, すべてのセッションで同一のアバターを使用する「1 アバター状態」を比較するというものであった. アバターの種類条件は, アバターが人間であるか動物であるかというものであった. よって条件は, アバターの数条件 (1 アバター状態, 4 アバター状態) とアバターの種類条件 (人間, 動物) の 2 条件とした.

2.3. 刺激

アバターは, メッセージアプリの機能にあるミー文字 (開発会社名 : Apple 社, iOS12 から導入された機能) を使用して作成した. 全ての刺激は, 画面の左側にアバターが配置され, 右側にニュース動画が配置される形式となっている. ニュース動画は, いずれも天気に関する街頭インタビューを内容とし, 視覚的な一貫性を保つために 4 種類作成した.

1 つ目の刺激は, 男女含めた 4 人のアバターがセッション毎に解説するニュース動画である. 2 つ目の刺激は, 1 人のアバターが全てのセッションを解説するニュース動画である. 3 つ目の刺激は, 4 匹のアバターがセッション毎に解説するニュース動画である. 4 つ目の刺激は, 1 匹のアバターが全てのセッションを解説するニュース動画である (図 1) . また, 全ての刺激はアバターの口の動き条件があり, ニュース動画の解説音声とアバターの口の動きが一致している状態で提示した. これは, 視覚的な手がかりとして口の動きと聴覚情報との統合が, 情報の理解度に寄与することができると考えられるからである.

図 1 動物のアバターが解説するニュース動画

2.4. 手続き

参加者は実験同意書が明記された Google form の QR コードをスマホで読み取り回答を行った. そして, 参加者に対して以下のような概要の説明を行った.

・これから 4 つの刺激動画を見て頂き, それぞれの動画を視聴した後にテストを行う.

・このテストは 30 個の単語が並べてある.

・この 30 個の単語の内訳は, 実際に実験動画に発せられた単語 (使用語) が 10 個, 実際に実験動画に発せられた単語の関連語 (使用関連語) が 10 個, 映像として登場するが, 音声とは無関係な語 (映像関連語) が 10 個の計 30 個である. そこから実際に発せられた単語と思うものを選択する.

説明後, 同意のあった参加者に対し, 実験参加用の QR コードを提示し, 参加者はその QR コードを読み取り実験に参加した. なお, 参加者には必要に応じてイヤホンやヘッドフォンなどをつけて動画を再生するよう求めた.

参加者は自身で Google Forms 上の実験動画を流した後に, その動画の内容をどの程度覚えて理解したのかを知るためにテストを計 4 回行った. このテストは参加者が実験刺激を見た後にクールタイム (1 分間) を設け, その後に回答を行うものである. またこのテストは, 1 回につき 30 個の単語を提示した. 単語の内訳は, 実際に実験動画で発せられた単語 (使用語) が 10 個,

実際に実験動画に発せられた単語の関連語（使用関連語）が 10 個，映像として登場するが，音声とは無関係な語（映像関連語）が 10 個の計 30 個である．そして，参加者に対し実際に実験動画で発せられた単語を 30 個の単語の中から 10 個選択してもらった．実験結果の信頼性を高めるために，参加者が見る実験刺激の順番はランダマイズした．

2.5. 分析方法

データの分析には統計ソフトウェア R およびパッケージ lme4，ggeffects を使用した．得られたデータは二値（選択または非選択）であり，さらに同一の被験者が複数条件で回答する繰り返し測定デザインであったため，個人差をランダム効果として考慮しつつ，固定効果の影響を検討できる一般化線形混合モデル（GLMM: Generalized Linear Mixed Model）を用いた．GLMM は，二項分布に従う応答変数に対応しながら，被験者間のばらつきをモデルに取り込むことができる．

固定効果は アバターの種類（Avatar Type:人間アバターまたは動物アバター），登場するアバターの数（Number of Avatar:1 または４），単語カテゴリー（Word Category: 実際に実験動画で発せられた単語(使用語: Actual Words)，実際に実験動画に発せられた単語の関連語(使用関連語: Related Words)，映像として登場するが，音声とは無関係な語(映像関連語: Video related words）であった．

分析に使用したモデルは，交互作用項が含まれた一般化線形混合モデルを構築した後，MuMIn パッケージを利用し，前向きのモデル選定を行った．その結果，Number of Avatar，Word Category とその交互作用が含まれたモデルが選定されたため，分析にはこのモデルを利用した．なお，Number of Avatar 要因のコーディングには，コントラストコーディング(e.g.，-0.5，0.5)を用いた．Word Category 要因は Forward Difference Coding を用いた．3 水準において，このコーディングを利用すると順に比較することになる．よって，Word Category においては，Actual Words と Related Words(Word Category1)，Related Words と Video related words（Word Category2)との比較をすることになる．

3. 結果

GLMM の結果，切片（Est. = -1.1366，z = -9.141，p < .01)，Word Category1（Est. = 2.2313，z = 8.241，p < .01)，Word Category2（Est. = 1.1653，z = 3.855，p < .01）が有意であった．加えて，Number of Avatar と Word Category2 の交互作用が有意であった(Est. = -2.1271，z = -3.50，p < .01)．しかしながら，Number of Avatar の主効果は有意ではなかった（Est. = 0.2230，z = 0.970，p = .33)．

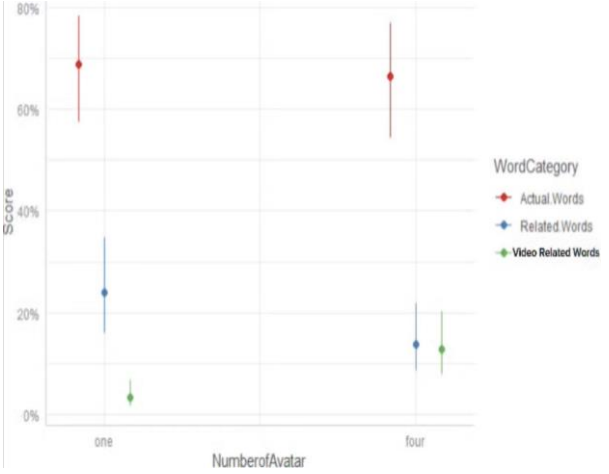
アバター数と単語カテゴリーの交互作用は有意であったため，dummy coding を利用した単純主効果検定を行った．

1 アバター状態は，Word Category1（Est. = -1.9453，z = -5.239，p < .01)，Word Category2（Est. = -4.1742，z = -9.279，p < .01）が有意であった．よって，アバターが 1 体のみでニュースを伝えた場合は，使用語は使用関連語より高い割合で選択され，使用関連語は映像関連語より高い割合で選択されたことがわかった．

4 アバター状態は，Word Category1（Est. = 2.5172，z = 6.454，p < .01)は有意であったが，Word Category2（Est. = 0.1017，z = 0.253，p = .80）は有意ではなかった．よって，４アバター状態で，それぞれのアバターがニュースを伝えた場合は，使用語は使用関連語より高い割合で選択されたが，使用関連語と映像関連語は変わらない割合で選択されたことがわかった．

使用語における Number of Avatar は有意であった（Est. = 1.4504，z = 3.158，p < .01)が，映像関連語においては Number of Avatar は有意ではなかった(Est. = -0.1048，z = -0.286，p = 0.775)．

図２：１アバター状態と４アバター状態での
各単語カテゴリーの選択確率
(赤線：Actual Words，青線：Related Words，
緑線：Video Related Words)



4. 考察

本研究では，1 アバター状態での映像関連語は選択されにくいことがわかった．この要因として 1 アバター状態では視覚的・聴覚的注意が 1 点に集中するため，映像として登場するが，音声とは無関係な語（映像関連語）は相対的に無視されやすくなったと考えられる．一方で，アバターが複数登場する条件では，注意が分散し，視覚情報が強調されるため，使用関連語と映像関連語が選択される傾向に差が見られなくなることも示された．これらの結果は，情報提示の方法（アバター数）が注意の向け先を変化させ，語の選択行動に影響を与えることを示唆していると考える．

続いて，Word Category1（使用語と使用関連語との比較）と Word Category2（使用関連語と映像関連語との比較）がともに有意であったことから，語の種類ごとに選択率に明確な差があることがわかった．使用語はニュース中で実際に聞こえる語であるため，音声による言語的処理が促進され，認知的によりアクセスしやすかったと考えられる．また，使用関連語も多少の意味的連想を通じて選択されやすいが，使用語ほどの直接的な結びつきはないため，使用語との間に有意な差が生じたと考えられる．加えて，映像関連語は意味的にも聴覚的にも弱い結びつきであったため，使用関連語と比べても選択されにくかったと考えられる．

さらに，Number of Avatar と Word Category の交互作用が有意であったことは，提示されるアバターの数によって語のカテゴリーごとの選択行動が異なることを示している．具体的には，1 アバター状態は，使用語・使用関連語・映像関連語の間に明確な序列（使用語 > 使用関連語 > 映像関連語）が見られたが，4 アバター状態は，使用語と使用関連語の間には有意差があったものの，使用関連語と映像関連語の差は見られなかった．これは，アバター数の増加によって視覚的な負荷や情報量が増加し，意味的な結びつきの処理が困難になったためと考えられる．すなわち，複数のアバターが順次提示されることで視覚情報が多くなった結果，聴覚情報と視覚情報の対応づけが曖昧になり，映像関連語が誤って選択される傾向が高まった可能性がある．

5. 結論

本研究では，アバター数および単語カテゴリーが語の選択行動に及ぼす影響を検討した．その結果，1 アバター状態は，発話と意味的に関連する語（使用語や使用関連語）が優先的に選択され，映像として提示されものを表す語は選択されにくい傾向が明らかとなった．一方，4 アバター状態は，使用された語に関する記憶は優先的に選択されていたものの、1 アバター状態と比べ、視覚的注意が分散することで映像関連語の選択率が相対的に上昇し，意味的関連性による選択傾向が弱まることが示唆された．このことは，実験後の参加者の自発的コメントからも伺えた．

これらの結果は，アバターの提示数が注意の配分や情報処理の深度に影響を与え，それが語の意味的関連性に基づく選択行動を変化させることを示している．特に，1 アバター状態による明確な情報提示は，受け手にとって意味的整合性を重視させ，不要な情報を排除する認知的フィルタリングを促進する可能性がある．今後の研究では，さらなる実験設計や実験参加者数を増やし，より詳細な理解を深めることが求められる．

文献

Driver IV, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive orienting. *Visual cognition*, *6*(5), 509-540.

Mizuho, T., Amemiya, T., Narumi, T., & Kuzuoka, H. (2023, March). Virtual omnibus lecture: investigating the effects of varying lecturer avatars as environmental context on audience memory. *Proceedings of the Augmented Humans International Conference 2023*, pp. 55-65.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694.

Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101*(3), 621-629.