

筆記にかかる認知負荷の考察に基づいた 講義映像再生システムの提案

井元 麻衣子^{1,a)} 渡邊 大喜¹ 宮川 和¹ 中山 丈二¹ 山田 智広¹ 片岡 良治¹

概要：インターネットメディアの発展によって、ネットワークを介したマルチメディア教材配信サービスの展開が目覚ましい。マルチメディア教材利用においても、授業の理解を促進するには講義を聴きながらの筆記行為が重要であるが、学習者には講義内容の理解と筆記行為を同時にこなす高度な認知能力が要求される。そこで、本研究では、筆記行為と認知負荷の関係に基づいて、学習者の筆記にかかる認知負荷低減を目的とした講義映像再生システムを提案する。

1. はじめに

インターネットメディアの発展によって、ネットワークを介して、視覚コンテンツ(テキスト, 図, 動画)と聴覚コンテンツ(ナレーション)からなるマルチメディア教材を配信するサービスが広がりを見せており、世界規模で教育における学習環境に大きな変革をもたらしている。その中でも、高等教育機関等がインターネット上に講座を無償公開するオープンコースウェアや大規模公開オンライン講座(MOOC: Massive Open Online Course)の展開が目覚ましい。2001年にMITがオープンコースウェアを始めたのをきっかけにKhan Academy^{*1}やUdacity^{*2}などが展開されている。MOOCにおいては2012年に米国でCoursera^{*3}, edX^{*4}などのプラットフォームが公開されたのを皮切りに、学習者が世界の有名大学の講義を大学に通わずに受講できる環境が急速に整いつつある。

教室での授業や文章読解に対する理解を促進するには、講義を聴きながらノートを取ることが重要である[1]ことが知られているが、コンピュータを利用したマルチメディア教材視聴においては、タブレットPC上に表示させたマルチメディア教材への手書きによる文字の書込み行為や下線引きなどの筆記行為が有効である[2]ことが実験によって示されている。タブレットPCでは視覚コンテンツと筆記領域とを近接して提示できるため、学習者の視線移動頻

度が減少する。また、分裂注意原則が起こりにくいと考えられる[3]。分裂注意原則[4]とは、学習者が物理的や空間的または時間的に分離した複数の情報を統合するときに認知負荷が発生することである。

しかしながら、学習者のノートを取る速さは教授者の発話速度よりも1/10程度と遅い[5]ことから、マルチメディア教材利用においても、学習者には講義内容の理解と筆記行為を同時にこなす高度な認知能力が要求される。具体的には、教授者の発話を筆写したり文字に起こすためには、視聴中に絶えず教授者の発話を更新・記録しておかなければならない[6]とともに、教授者の発話中に内容を書ききれぬとは限らない。教材を視聴しながらの筆記行為は学習者にとって困難であり、筆記中に聞き逃した箇所を巻き戻したり、筆記中に動画を一時停止させざるを得ず、1講義あたりの視聴時間は長くなりやすい。マルチメディア教材の最適な視聴時間は15~20分である[7]とされており、視聴時間の増大は集中力の維持困難や学習効果の低下につながる恐れがある。

そこで、本研究では、マルチメディア教材視聴における筆記行為と認知負荷の関係を考察し、筆記にかかる認知負荷を軽減する講義映像再生システムの実現を目指す。本稿では、先行研究に基づいてデザイン指針を議論し、デザイン指針に基づいたシステムを提案する。提案システムの概要を図1に示す。提案システムは以下の特徴をもつ。

- 学習者の筆記情報から筆記種別を推定する。
- 講義映像を1つ以上のシーンに分割し、シーン間の関連性とシーンごとの筆記にかかる認知負荷を推定する。
- 筆記種別と筆記にかかる認知負荷の推定結果に基づいて、講義映像の再生速度を動的に調整する。

¹ 日本電信電話株式会社 NTT サービスエボリューション研究所
NTT Service Evolution Laboratories, NTT Corporation
1-1 Hikarinooka Yokosuka-Shi, Kanagawa, 239-0847, Japan

a) imoto.maiko@lab.ntt.co.jp

*1 <https://www.khanacademy.org>

*2 <https://www.udacity.com>

*3 <https://www.coursera.org>

*4 <https://www.edx.org>

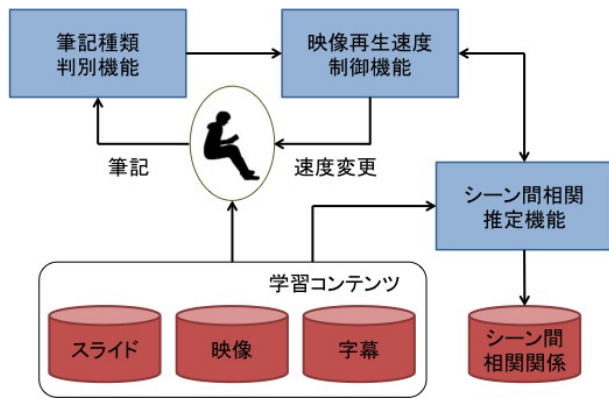


図1 提案システムの概要

本稿の構成は以下のとおりである。第2章で関連研究について述べる。第3章では提案システムのデザイン指針について説明する。そして、第4章でデザイン指針に基づいた講義映像再生システムを提案する。最後に、第5章で本稿をまとめ今後の課題について論ずる。

2. 関連研究

これまでも、学習者のマルチメディア教材視聴における筆記を支援する研究が提案されている。

谷内は学習者のノートを記述するペースに合わせて動画講義の再生ステータスの自動調整をおこなう動画講義プレイヤーを提案している[8]。この手法では、ノート記述量がすでに視聴したシーンよりも多ければ再生速度を遅くし、次のシーンに移る際にキー入力が続いている場合は、入力が終わるまで自動的に動画を一時停止させる。したがって、学習者の筆記量が多ければ多いほど講義映像を視聴するのに時間を要する。

八重樫らは動画中に学習者による印付けと他視聴者の印付けを表示可能なストリーミングプレイヤーを提案している[9]。この手法では、投稿されたテキストに下線引きをおこない、それを共有することで、他視聴者がどの部分に着目しているかを把握しやすくし、要点をとらえやすくする。また、他視聴者の筆記を共有することで、学習者自身の筆記を振り返るときに筆記の過不足を認識することができる。しかしながら、学習者ごとに既有知識の差があり、筆記が必要な箇所は学習者によって異なると考えられる。

栗原らは音声認識と文字認識を組み合わせ、学習者がペンで文字列を書き始めると、それに続く可能性のある文字列リストを提示し、それらを選択することで残りの筆記行為を省略できる予測手書き入力システムを開発している[10]。この手法では、手書き文字認識の認識誤りを学習者に意識させないことを目的としており、本研究とは主張が異なる。

3. デザイン指針

本章では、筆記における筆記種別の違いと筆記内容に対

する学習者の既有知識量に着目し、これらと筆記にかかる認知負荷の大きさの関係を先行研究の知見に基づいて考察し、提案システムのデザイン指針を議論する。

3.1 筆記種別と認知負荷の関係に基づくデザイン指針

本節では、筆記と認知負荷の関係を考察し、それに基づいて提案システムのデザイン指針を立てる。

学習効果を向上させるためには、学習時に聴覚情報として得られる講師の発言と筆記時に視覚情報となる教材や学習者自身の筆記は時間的に近いことが重要である[11]。Paivioらが提唱した「デュアル・チャンネル・モデル」によれば、人間が情報を処理する際に聴覚情報と視覚情報が同期することにより、これらが相互に作用し、情報伝達効率がより強化されると考えられている。前述したように、筆記速度は発話速度の約10分の1である[5]から、筆記に応じて講義映像の再生速度を遅くすることで認知負荷を低減させることができると考えられる。

学習時における筆記は、学習者への認知負荷の観点から以下の3つに分類できると考えられる。

マーク 教材上の文章に下線などのマークを書き込む。

筆写 教材の空白部分やノートなどの教材以外の媒体に、講師の発言内容を書き写す。

注釈 教材の空白部分やノートなどの教材以外の媒体に、要点や気づいたことを自分の言葉で書き留めたり、図表を描画する。

学習者への認知負荷の大きさは筆記種別ごとに異なり、「注釈」>「筆写」>「マーク」であると考えられる。「注釈」には講師の発言を聞きながら自身がすでに持っている知識と関連づけた上で文章を書き起こす思考能力が要求され、聴覚情報への注意が散漫となりやすく、講師の発言をそのまま書き写す「筆写」より認知負荷は大きいと考えられる。

したがって、本研究では、以下のようなデザイン指針に基づいて提案システムを開発する。

指針1 筆記種別を判定し、判定結果に基づいて講義映像の再生速度を遅く調整する。筆記中の再生速度は、筆記種別が「マーク」「筆写」「注釈」の順に遅くする。

3.2 筆記部分と既有知識の関係に基づくデザイン指針

本節では、講義内容に対する学習者の既有知識と筆記にかかる認知負荷の関係を考察し、それに基づいて提案システムのデザイン指針を立てる。

学習者への認知負荷の大きさは、講義内容に対する学習者の既有知識量によっても異なると考えられる。学習者にとって未習の項目 t の解説が始まった直後は、項目 t に対する知識が乏しいために項目 t の認知に対する負荷がかかるため、筆記にかかる認知負荷が大きいと考えられる。しかし、教授者による解説が進むにつれ、学習者の項目 t に対する理解は深まっていき、それにともなって学習者の筆

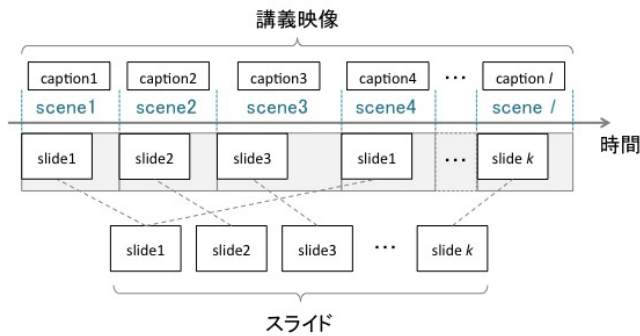


図 2 講義映像のシーン分割

記にかかる認知負荷は小さくなっていくと考えられる。

また、筆記がおこなわれるシーンでは学習者にとって未習の項目が解説されていることが多いと考えられる。したがって、学習者 s がシーン n において学習者が筆記するとき、学習者 s にとってシーン n の項目は未習であり、シーン n と講義内容の関連性があるシーン集合 C にも学習者 s にとって未習の項目が含まれる可能性は高く、学習者 s はシーン集合 C においても筆記する可能性が高いと考えることができる。

これらの考察に基づくと、あるシーンにおいて学習者が筆記するとき、そのシーンと関連性があるシーンでの筆記にかかる認知負荷は徐々に減少すると考えることができる。

したがって、本研究では、以下のようなデザイン指針に基づいて提案システムを開発する。

指針 2 学習者による筆記タイミングと講義映像のシーン間の関連性に基づいて講義映像の再生速度を調整する。

4. 提案システム

本章では、第 3 章で述べたデザイン指針に基づいた講義映像再生システムを提案する。

4.1 提案システムの概要

本提案システムは以下の 3 つのステップから成る。本システムは、講義映像に対して Step1 を予め実行しておき、学習者の入力操作を検知したら Step2, 3 を実行する。

Step1: 講義映像シーン間の講義内容の関連性推定 (前処理)

1 つの講義映像を 1 つ以上のシーンに分割し、各シーン間の講義内容の関連性を推定する。

Step2: 筆記種別の推定 学習者の入力操作を検知し、入力操作が「マーク」、「筆写」、「注釈」のうちどれであるかを推定する。

Step3: 映像再生速度の算出 Step1, Step2 の推定結果に基づいて、講義映像の再生速度を調節する。

次節以降で各ステップの処理の詳細を述べる。

4.2 Step1: 講義映像シーン間の講義内容の関連性推定

MOOC で提供される学習コンテンツの多くは、スライ

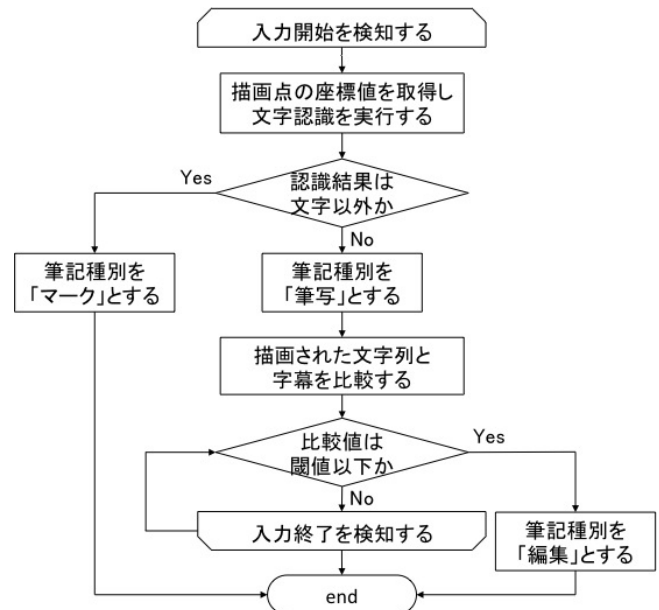


図 3 Step2 の処理手順

ド形式の教材、講義映像、字幕の 3 つを含む。本ステップでは、まず、図 2 に示すように、講義映像を映像中のスライドの切替タイミングでシーンという単位で分割する。学習コンテンツによっては 1 つのスライドで講師が複数の話題を言及することもあるため、複数のシーンで説明に同一のスライドが使われていることがある。したがって、シーンの分割数はスライドの枚数と必ずしも一致しない。

次に、本ステップでは字幕情報を用いて各シーン間の講義内容の関連性を推定する。例えば、シーン s で階層型クラスタリング手法が解説され、シーン s' で階層型クラスタリングと非階層型クラスタリングの違いが解説されているとき、この 2 つのシーンはいずれもクラスタリング手法の解説という共通点があり、講義内容に関連があると考えられる。

シーン間の講義内容の関連性は、具体的には、まず、形態素解析エンジンを用いて字幕情報を単語単位に分割し、シーン s における単語 w の重要度 W_{ws} を $TF \cdot IDF$ によって算出する。そして、シーン s_1 とシーン s_2 の特徴ベクトル \vec{v}_{s1} 、 \vec{v}_{s2} を用いて、シーン s_1 とシーン s_2 の相関関係 $r(s_1, s_2)$ を以下の式によって算出する。

$$r(s_1, s_2) = \cos\theta_{(s_1, s_2)} = \frac{\vec{v}_{s1} \cdot \vec{v}_{s2}}{|\vec{v}_{s1}| |\vec{v}_{s2}|} \quad (-1 \leq r(s_1, s_2) \leq 1)$$

4.3 Step2: 筆記種別の推定

本ステップでは、マウス操作やタッチ操作を検知して筆記に対する描画処理をおこない、描画点の座標情報を取得して文字認識を実行し、その結果に基づいて学習者の入力操作の筆記種別を推定する。本ステップの処理手順を図 3 に示す。具体的には、文字認識の結果、入力された描画点の集合が文字でないとき認識された場合は、筆記種別を「マーク」と推定し、処理を終了する。入力された描画点

であると認識された場合は、筆記種別を「筆写」か「注釈」かを判別しなければならない。しかしながら、判別にはある程度の文字数が必要であるため、判別までには時間がかかる。そこで、本手法では、受講中の学習者による筆記の60%は講師の発言の筆写である [12] という知見に基づいて、筆記種別を「筆写」であると仮決めする。そして、学習者による文字入力が行われている間、一定時間間隔で学習者によって入力された文字列と字幕データに対してレーベンシュタイン距離を算出し、2つの文字列を比較し、距離が閾値以下になれば、筆記種別を「注釈」に更新する。

4.4 Step3: 映像再生速度の算出

本ステップは、学習者の入力操作中における講義映像の再生速度を、Step1で推定した講義映像シーン間の相関関係とStep2で推定した筆記種別に基づいて算出し、講義映像の再生を制御する。本ステップはStep2で筆記種別を推定したタイミングで実行される。

本ステップでは、第3章で示したデザイン指針に基づいて、以下の条件を満たすような再生速度 v を算出する。

条件 1 再生速度 v は $v < v_0$ を満たす。

条件 2 再生速度 v は筆記種別が「マーク」「筆写」「注釈」の順に小さくなる。

条件 3 t_2/t_1 が大きいほど、再生速度 v は小さくなる。

条件 4 t_2/t_3 が大きいほど、再生速度 v は小さくなる。

C_1 : 入力操作中に再生されているシーン n との相関関係

$r(n, c)$ が閾値 $M(0 \leq M \leq 1)$ 以上のシーン c の集合

C_2 : 集合 C_1 に含まれ、かつ、シーン n 以降に再生されるシーンの集合 ($C_1 \supseteq C_2$)

v_0 : 学習コンテンツ作成時に設定された再生速度

t_1 : 集合 C_1 に含まれるすべてのシーンを速度 v_0 で再生した場合の合計再生時間

t_2 : 集合 C_2 に含まれるすべてのシーンを速度 v_0 で再生した場合の合計再生時間

t_3 : シーン n 以降を速度 v_0 で再生した場合の合計再生時間

本ステップでは、速度 v を以下の式によって算出する。

$$v = \begin{cases} v_0 \times \beta & (K \leq \frac{1}{t_2/t_1 + t_2/t_3}) \\ v_0 \times \beta \times \frac{1}{t_2/t_1 + t_2/t_3} & (1 < \frac{1}{t_2/t_1 + t_2/t_3} < K) \\ v_0 & (\text{otherwise}) \end{cases}$$

実数 β は、講義映像の再生速度を0.7倍まで低速にしても学習者が違和感を感じにくい [8] とされることから、Step2で推定した筆記種別が「マーク」のとき $\beta = 0.9$ 、「筆記」のとき $\beta = 0.8$ 、「注釈」のとき $\beta = 0.7$ とする。閾値 K は、速度 v が $v > v_0$ とならないような実数である。

5. まとめと今後の課題

本稿では、講義映像を視聴中の筆記にかかる認知負荷低減を目的として、学習者の筆記活動に基づいて講義映像の

再生速度を動的に制御する講義映像再生システムを提案した。今後は、被験者による実験を通して以下の項目について検証し、デザイン指針の妥当性や効果を検証する予定である。

- (1) 筆記種別によって認知負荷の大きさが異なる。
- (2) 筆記種別に応じた講義映像の再生速度調節が認知負荷低減に効果的である。
- (3) 講義内容に関連がある複数のシーンと筆記には相関があり、かつ、筆記量は単調に減少する。

筆記にかかる認知負荷は学習コンテンツの特性や学習者個人の特性にも依存すると考えられるため、実験を通して、今後も筆記と筆記にかかる認知負荷の関係について深く考察する必要があると考えている。

参考文献

- [1] Williams, R.L. and Eggert, A.C.: Note-taking in college classes: Student patterns and instructional strategies, *The Journal of General Education*, Vol.51, No.3, pp.173-199 (2002).
- [2] 安藤雅洋, 植野真臣: eラーニングにおけるタブレットPCを用いた書き込みの分析、*日本教育工学論文誌*, Vol.35, No.2, pp.109-123 (2011).
- [3] Moreno, R. and Mayer, R.E.: Cognitive principles of multimedia learning: The role of modality and contiguity, *Journal of Educational Psychology*, Vol.91, pp.358-368 (1999).
- [4] Mayer, R.E.: *The Cambridge handbook of Multimedia Learning*, Cambridge University Press.
- [5] Foulin, J.N.: Pauses et debits: les indicateurs temporels de la production ecrite [Pauses and fluency: chronometrics indices of writing production], *L'Annee Psychologique*, pp.483-504 (1995).
- [6] Baddeley, A.D.: Exploring the central executive, *Quarterly Journal of Experimental Psychology*, 49A, pp.5-28 (1996).
- [7] Frederick, P.J.: The lively lecture: 8 variations, *College Teaching*, Vol.34, No.2, pp.43-50 (1986).
- [8] 谷内正裕: 学習者のノート記述情報から再生状態を動的に調整する動画講義プレーヤー、*メディア教育研究*, Vol.5, No.2, pp.85-94 (2008).
- [9] 八重樫文, 北村智, 久松慎一, 酒井俊典, 望月俊男, 山内祐平: iPlayer: eラーニング用インタラクティブ・ストリーミング・プレーヤーの開発と評価、*日本教育工学論文誌*, Vol.29, No.3, pp.207-216 (2006).
- [10] 栗原一貴, 後藤真孝, 緒方淳, 五十嵐健夫: 音声ペン: 音声認識結果を手書き文字入力で利用できる新たなペン入力インタフェース、*コンピュータソフトウェア*, Vol.23, No.4, pp.60-68 (2006).
- [11] Paivio, A.: *Mental representations: A dual coding approach*, Oxford University Press, Oxford (1986).
- [12] Bretzing, B.H. and Kulhavy, R.W.: Note-taking and Passage Style, *Journal of Educational Psychology*, Vol.73, No.2, pp.242-250 (1981).