

ELEVATE：学習者自身の自己聴取音声で聴く講義システム

福田 航希^{1,a)} 阪井 瞭介¹ 松下 嶺佑¹ 國見 友亮² 高道 慎之介^{1,2,b)}

概要：本論文では学習者が自身の合成自己聴取音声を用いて講義を聴くシステムである ELEVATE を提案する。自己聴取音声により自分の発話に対する行為主体感や心理的所有感を高め、結果として講義への没入感や学習効率を向上させることが期待される。実験の評価により、本システムによる主観的な講義体験の変化や学習者の記憶に対する効果があることを示す。

1. はじめに

COVID-19 をきっかけにオンライン授業を受ける機会が増加し、現在も定着しつつある [1]。中でもオンデマンド授業はフレキシビリティの高さが評価されている一方、学習者の主体的な関与が不足しがちであり、モチベーションが維持しづらいことや、ドロップアウトしやすいことが指摘されている [2]。学習者が受け身で講義を聞く状況では、学習内容の定着や学習意欲が低下する可能性が高く、効率的な学習には、学習者自身が学習プロセスに能動的に関与する仕組みが求められている。

関連する研究として、音読や自己関連付け効果によって学習効果が向上することが示されている [3, 4]。また、自己聴取音（発話者が自分自身で聞く声）は、自分の発話に対する行為主体感や心理的所有感を高める可能性がある。これらの心理的要素が学習に与える影響を活用することで、オンライン授業における学習効果を向上させる新しいアプローチが期待できる。

本論文では、自己聴取音声を活用した新しい講義システム、ELEVATE (Enhanced Learning Experience with self-listening Voice-Adoptive Teaching Environment) を提案する。本システムは、学習者が学習者自身の合成自己聴取音声で講義を聴くことを可能とし、学習者各々にとってより効果的な学習体験を提供することを目的とする。本研究の貢献は、次の3点に集約される。(1) 自己聴取音が学習者に講義内容の定着や関与を促進する可能性を示した。(2) テキスト音声合成技術と自己聴取音再現技術を組み合わせ、新たな講義体験を提供するシステムを提案した。(3) 合成自己聴取音声は心理的所有感や行為主体感に与える影

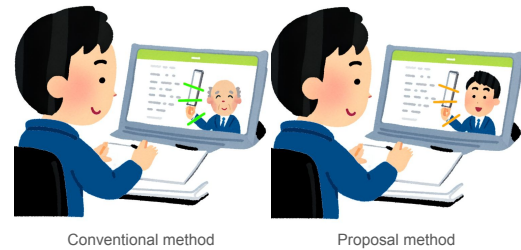


図1 提案システムの概念図。従来（左）は誰が受講しようとも同じ講義者による説明を受ける。提案システム（右）では、学習者自身の声で講義を受ける。

響と、学習者の感情やモチベーションに対する影響を明らかにした。

本研究のリサーチクエスションは、以下の通りである。

RQ1:合成自己聴取音に対して心理的所有感が生じるか。

RQ2-1:本システムによって、講義に対する行為主体感が生じるか。

RQ2-2:講義に対する行為主体感に伴って、講義内容に対する心理的所有感が生じるか。

RQ3-1:本システムによる講義体験が学習者の感情にポジティブ、あるいはネガティブな変化をもたらすか。

RQ3-2:学習者の感情の変化に伴って、講義内容に関する好感度に変化があるか。

RQ4:本システムが学習者の記憶力を向上させるか。

2. 関連研究

2.1 心理的所有感と行為主体感

心理的所有感 (psychological ownership) とは、自己の身体などが自分のものだという主観的体験である [5]。心理的所有感の対象は有形・無形に関わらず幅広い [6]。例えば、我々は自分の声に対して所有感を持つ [7]。心理的所有感は、(1) コントロールの経験、(2) 自己投資、および、(3) 知識や親しみであるの3条件によって形成され、そのうち1つの要因のみからでも心理的所有感は生じう

¹ 慶應義塾大学

² 東京大学

^{a)} fukuda_koki@keio.jp

^{b)} shinnosuke_takamichi@keio.jp

る [8, 9, 10, 11].

心理的所有感に関連する概念に、行為の心理的所有感を示す行為主体感 (agency) がある [5]. 行為主体感とは、自己が行為の作用主体だという感覚のことである [5]. 例えば、人間は発話した声をフィードバックされることで、発話に対する行為主体感を得る [12]. 一方で、高い成績や到達度テストの得点、学業継続率といった学業的成功と関連している指標に、学生エンゲージメント (student engagement) がある [13, 14]. 学生エンゲージメントに合意のある定義は無いものの [15], 行為主体感は学生エンゲージメントの 1 要素である [16]. また、単語の記憶タスクにおいても、行為主体感は成績を高めることが示されている [17].

本研究では、学習者が学習者自身の自己聴取音声聴くことによって、あたかも自分自身が講義をしているかのような感覚、すなわち講義に対する行為主体感が生じ、効率的な学習が行えるのではないかと、という仮説を立て、これを検証する。また、講義に対する行為主体感が生じたとき、前述した心理的所有感が生じる 3 条件の (1) と (3) を満たす^{*1}。したがって、講義内容に対しても心理的所有感が生じるのではないかとという仮説を立て、これを検証する。

2.2 テキスト音声合成

テキスト音声合成 (text-to-speech; TTS) とは、任意のテキストから対応する音声波形を合成する技術である [18]. 昨今のテキスト音声合成の生成モデルは、音声時間長について明示型と非明示型に大別される。明示型は、入力テキストの各音素についてその音声時間長を内部で明示的に推定する [19]. 一方の非明示型は時間長を推定せずブラックボックス的に扱う [20]. 前者の明示型は、合成音声の話速の制御が容易であり、本研究のように動画に合わせて話速を容易に制御することも可能である。

2.3 音読と記憶

人間は視覚提示された情報を記憶するとき、音韻ループを用いて情報を一時的に保つ。音韻ループは音韻貯蔵庫と構音リハーサルによって構成される。視覚情報は聴覚コードに変換され、脳内で何度も繰り返し復唱する。このとき、繰り返される脳内での発話を内言と呼ぶ。そして、音読は内言を発話として外に出す行為である [21]. 音読は読解に習熟した成人にとっても音読の構音運動が理解を支える役割を持ち、認知資源が少ない場合には黙読よりも有用である [4]. 本システムの狙いの一つは、この音読の作用を外部的化することである。

2.4 自身の録音音声・合成音声認知に与える影響

人は自己に関連する情報を重要視する傾向があり、自己関連刺激は他者関連刺激よりも認知的処理を促進する。こ

れを自己関連付け効果と呼び、自己音声もこの効果を持つ刺激に含まれる [3]. 神経科学の研究では、自己関連刺激が高い感情反応を示す皮膚電流反応を引き起こし [22], これによってある刺激に選択的に注意を集中させる注意バイアスを引き起こすことがわかっている [23].

自己音声の効果を利用してユーザの行動変容を目指したのが自己音声アラームである [24]. ユーザの音声を事前に録音し、リマインダやアラーム音として再生することで、行動を起こすモチベーションや注意を喚起する効果が得られることを報告している [24].

テキスト音声合成が発展したことにより、任意のテキストについてユーザの声を合成できるようになり、上記の類のアプローチをより柔軟に実現できるようになった。これを利用して、合成自己音声を教育的ゲームのアバター音声とする研究があり、合成自己音声による指示で没入感の向上が認められている [25].

しかし、自己音声を聞く体験はネガティブな体験となる可能性がある。例えば、自己音声の録音の評価は他者の音声と比べネガティブになる [26]. 一つの考えられる理由として、voice confrontation [27] と呼ばれる現象がある [24]. これは、予想される音声と録音された音声と異なることから生じる認知的不協和のことである [27]. 次項で説明する合成自己聴取音声は、この現象を解決し得ると期待される。

2.5 自己聴取音の人工再現

マイクで録音した自身の音声と、発話しながら聞く自身の音声は、その聞こえが大きく異なる。これは、両者で音声の伝播経路が異なるためである。前者は気導音と呼ばれ、口唇から放射された音声マイクロフォンに収録される。一方で後者は自己聴取音と呼ばれる。自己聴取音は、口唇から空気中を通して鼓膜に到達する気導音と、骨や体内組織を通して到達する体内伝導音からなる骨導音からなる [28]. 我々は、気導音ではなく自己聴取音を聴取して、音声に関する認知を獲得している。

マイクロフォンで収録した気導音から、体内伝導音を人工的に再現し、自己聴取音を生成する信号処理技術がある [29]. 伝導経路の違いは、周波数振幅特性 (周波数ゲイン) と音波到達時間に影響する。空気伝導に比べ体内伝導はローパスフィルタの役割を持つため [30], 適切なローパスフィルタを設計することで、空気伝導と体内伝導の間の周波数振幅特性差を疑似的に再現できる。なお、空気伝導音と体内伝導音の混合割合は、聴取者間、音素毎での個人差が認められている [28, 29].

3. 提案システム

提案する講義視聴システムの概略を図 2 に示す。このシステムでは、学習者自身の朗読音声を収録し、その音声を基に合成自己聴取音を生成することで、学習者自身の自己聴取音で講義動画を視聴する。

^{*1} 授業に対する行為主体感が生じたとき、本システムによる体験は、講義内容の知識を整理しながら、プレゼンテーションをする体験となるため。

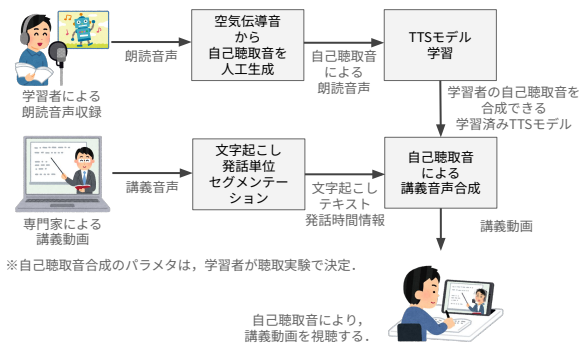


図 2 提案システムのフローチャート

まず、学習者は任意の文を朗読してその音声（気導音）を収録する。次に、この音声に対して自己聴取音加工を行う。加工のパラメータには個人差があるため、いくつかのパラメータ候補の中から学習者自身が、最も自身の自己聴取音の聞こえに近いものを選ぶ。このパラメータを用いてすべての収録音声を加工して、学習者の疑似的な自己聴取音を生成する。その後、朗読テキストと生成自己聴取音のペアデータを用いて、テキスト音声合成モデルを学習する。学習されたこのモデルは、任意のテキストから自己聴取音を合成することができる。

次に、専門家の講義する講義動画を用意する。この講義動画の音声に対し自動音声認識を施し、書き起こし文と文単位の発話開始終了時刻を得る。その後、書き起こし文をテキスト音声合成モデルに与えることで、その文を読み上げた自己聴取音を合成する。合成の際には、合成音声の時間長が元の時間長に合うように、合成音声の話速を制御する。最後に、元の講義動画の映像と合成音声を結合することで、学習者の自己聴取音で講義する講義動画を作成する。

学習者は、作成された講義動画を視聴して学習する。自己聴取音を用いた講義音声を聞くことで、記憶の定着が向上することが期待される。このテキスト音声合成モデルは任意の文の自己聴取音を合成できたため、本システムは様々な内容の講義動画に適用可能である。

4. 評価実験

提案システムの効果を検証するために、実験参加者（学習者）を対照群と介入群に分けて実験的評価を行った。対照群は、通常の動画講義の受講のように、専門家の音声付きの講義動画を視聴した。介入群は、提案システムを用いて、学習者自身の自己聴取音声付きの講義動画を視聴した。講義動画の視聴後には、講義内容に関するテストと実験に関するアンケートを実施し、本システムによって自分の声で講義を受けることが、認知や学習効果、心理状態に及ぼす影響を検証した。

4.1 システム実装

介入群は、講義の音声を合成自己聴取音に変換した状態で受講した。自己聴取音の人工生成方法として、先行研究 [29] に基づき、4 次バターワースフィルタ（カットオフ

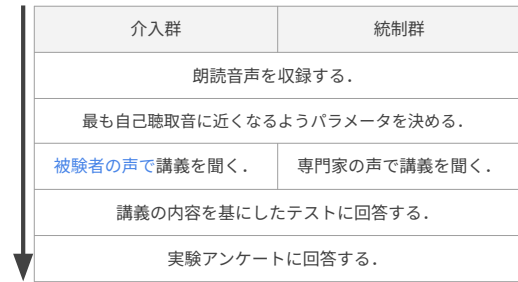


図 3 実験手順の概略

周波数 800 Hz) を用い気導音から疑似体内伝導音を合成し、時間領域で気導音と疑似体内伝導音を加算した。

授業動画の文字起こしと発話単位セグメンテーションには、Whisper-small *2 [31] を使用した。言語指定は日本語とした。認識誤りにより提案システムの効果が変化することを避けるため、文字起こしおよびセグメンテーションの誤りを手作業で修正した。

授業音声の学習および合成には、VITS2 [32] をベースにした、日本語特化モデル Style-Bert-VITS2 JP-Extra*3 を使用した。学習の初期モデルには事前学習モデル*4 を利用し、サンプリング周波数は 48 kHz とした。

4.2 参加者

日本語母語話者であり、かつ視聴覚に障害が無いことを条件として参加者を募集した。15 名の参加者（男性 11 名、女性 4 名、平均 24.6 歳）が実験に参加した。5 名が介入群、10 名が統制群として実験に参加した。

4.3 手順

実験手順を図 3 に示す。介入群と統制群に依らずすべての参加者には、用意したテキストを朗読しその音声を収録させた。講義動画の内容を参加者が予習することを避けるために、朗読テキストは講義内容と無関係の、心理学史や哲学に関する日本語文とした。収録音声の時間長は、全被験者で平均して約 23 分だった。収録機器は、参加者が所有するスマートフォンとした。サンプリング周波数などの収録設定が参加者間で統一するため、収録用アプリとその収録設定を事前に指定して参加者に収録させた。また、収録の際には、雑音が收音されないよう周囲環境に配慮すること、テキストの内容を読み間違えたときはその文から再収録することを参加者に伝達した。これらの設定と伝達により、音声合成の構築に十分な音質の音声を収録できることを、事前に確認している。

収録音声（気導音）のうち 1 発話を無作為に選択し、気導音と疑似体内伝導音の混合割合を 0%（気導音のみ）から 100%（疑似体内伝導音のみ）まで 10% 刻みで変えた 11 通りの音声候補を、参加者に提示した。参加者には、この

*2 <https://huggingface.co/openai/whisper-small>

*3 <https://github.com/litagin02/Style-Bert-VITS2>

*4 <https://huggingface.co/ku-nlp/deberta-v2-large-japanese-char-wwm>

音声のうち最も「自分が聞く自分自身の声」に似ていると感じるものを選択させ、その混合割合を当該参加者の設定値とした。参加者には音声のみを提示し、混合割合は提示しなかった。音声の聞き直し可能回数は無制限とした。介入群の参加者については、当該参加者の混合割合値を用いて、すべての収録音声疑似体内伝導音に加工した。

講義内容には、参加者の多くが未学習であり、かつ専門性の高い音声情報処理を選択した^{*5}。講義内容の内訳は4セクションとした。各セクションの動画時間は約10分で、動画の合計時間長は約40分だった。対照群の参加者には当該講義内容の専門家の録音音声で、介入群の参加者には参加者自身の合成自己聴取音声で、動画を視聴させた。動画には講義スライドのみを写し、発話者の顔や身体を写さないものとした。動画の再生は1度のみとし、動画の早戻しや早送りは認めなかった。専門家は日本語母語話者の男性1名(30歳台)とした。

講義動画視聴の直後、参加者は、講義に関するテストに回答した。なお、テスト実施について、講義動画視聴前に参加者に通達している。講義動画視聴とテスト回答はGoogle Meet上で実施し、実験中に参加者が不適の行動(例えば、動画を視聴していない)をとらないように確認した。テスト後にアンケートを実施した。テスト内容とアンケート内容は後述する。

4.4 評価尺度

4.4.1 記憶力

本システムが記憶に及ぼす影響を調査するため(RQ4)、講義動画視聴直後に講義内容に関する自由記述問題40問から成るテストを行う。問題は、講義の4セクション間で問題数が大きく偏らないよう配した。また、問題は、その解答が動画中のスライドに陽に表示されるものと、音声のみで言及されるものから成る。模範解答は、単語あるいは短いフレーズのみとした。

4.4.2 心理的所有感

合成自己聴取音声に対しても心理的所有感が生じているか(RQ1)、また講義内容に対する心理的所有感が生じているか(RQ2-2)を検証するため、日本語版 Psychological Ownership Scale (POS-J) [33]を用いて、講義音声と講義内容それぞれについてアンケートを行う。

4.4.3 行為主体感

講義に対して行為主体感が発生しているか(RQ2-1)を調査するため、先行研究[34]を参考に、「まるで講義を自分がしているかのような感覚」などの5設問に対し、7段階リッカート尺度で評価をする。

4.4.4 感情

講義体験が感情に及ぼす影響を調査する(RQ3-1)ため、

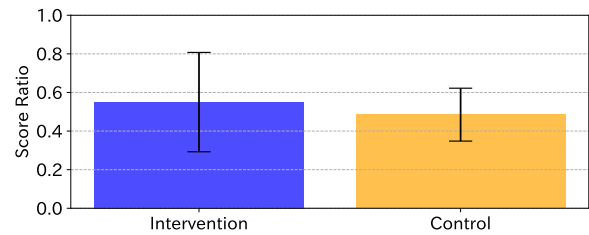


図4 テストの得点割合(平均±標準偏差)。2群間に有意差は見られなかったが、介入群の得点割合が高くなる傾向が見られる。講義を受けているときの感情を示すものとして、「やる気が湧いた」「神経質な」等のワードを提示し、6段階リッカート尺度によるアンケートを行う。人間の基本的な感情はPANASスケール[35]によって評価でき、5つのポジティブ感情と6つのネガティブ感情を測定する[36, 24]。また、先行研究[36, 24]に倣い、これらに加え、自己聴取音が起こしうる感情として、不快だ、恥ずかしい、イライラした、うしろめたいの4感情を加えた。

4.4.5 好感度

講義内容に対する好感度(RQ3-2)を調べることを目的として、「講義の内容をもっと学習してみたいと思いましたか?」などの3設問を、1から7の7段階リッカート尺度で評価する。

5. 実験結果

介入群と統制群の2群間における有意差を検証するため、各群の参加者数が20名以下であることから、マン・ホイットニーのU検定を両側検定として用いた。全ての統計分析はscipy.statsモジュール^{*6}を用いて行った。以下、平均値をM、標準偏差をSDと記す。また、事前知識について、「講義の内容をどのくらい知っていましたか?」という設問の1から7の7段階リッカート尺度によるアンケートを行ったところ、介入群について $M = 2.6$ 、統制群について $M = 3.6$ であり、2群間に有意差は見られなかった($p = 0.344 > 0.05$)。

5.1 記憶力

テストの正答率を図4に示す。テストの正答率は、介入群で $M = 0.550$ ($SD = 0.257$)、統制群で $M = 0.485$ ($SD = 0.137$)であった。有意差検定の結果、2群間に有意差は見られなかった。 $(p = 0.623 > 0.05)$ ただし、介入群に得点割合が高くなる傾向が見られ、本システムによって記憶力を高める可能性が示唆される。

5.2 心理的所有感

講義の音声と講義内容に対する心理的所有感の質問項目及びアンケート結果を表1に示す。ただし、「それ」が指す内容は質問項目に後続する丸括弧括弧内のものである。「まるで、それは私という人間の一部分であるかのように感じる。(講義音声)」及び「まるで、私は自分の一部をそれ

^{*5} 後述するが、事前知識について1から7の7段階リッカート尺度によるアンケートを行ったところ、2群間に有意差は見られなかった。

^{*6} <https://scipy.org/>

表 1 心理的所有感におけるアンケート結果. 合成自己聴取音に心理的所有感が生じたことが読み取れるが、講義内容に心理的所有感は生じなかった. * は $p < 0.05$ で有意差ありを示す.

Question item	Intervention		Control		p-val
	M	SD	M	SD	
まるで、それは私という人間の一部であるかのように感じる。(講義音声) *	3.60	1.949	1.40	0.699	0.032
まるで、私は自分の一部をそれから得ているかのように感じる。(講義音声) *	3.00	1.581	1.50	1.269	0.045
まるで、それは自分のアイデンティティの中心であるかのように感じる。(講義音声)	1.60	0.894	1.20	0.632	0.254
それは、実際の自分になりたい自分とのギャップを埋める助けをしてくれるような気がする。(講義音声)	2.40	2.074	1.50	1.581	0.083
それは、私が望むようなアイデンティティを獲得する助けをしてくれるような気がする。(講義音声)	2.20	1.643	1.40	1.265	0.083
それと私には、たくさんの共通点があるような気がする。(講義音声)	4.00	1.871	2.20	1.398	0.102
まるで、私はそれを所有しているかのように感じる。(講義音声)	2.40	1.673	1.30	0.675	0.125
まるで、それは私の所有物であるかのように感じる。(講義音声)	2.40	1.673	1.20	0.632	0.059
まるで、私はそれに対して高い所有権を有しているかのように感じる。(講義音声)	2.80	2.049	1.50	1.269	0.142
まるで、それは私という人間の一部であるかのように感じる。(講義内容)	2.00	1.000	2.00	1.491	0.789
まるで、私は自分の一部をそれから得ているかのように感じる。(講義内容)	1.80	0.447	2.20	1.814	0.688
まるで、それは自分のアイデンティティの中心であるかのように感じる。(講義内容)	1.60	0.548	1.90	1.370	1.000
それは、実際の自分になりたい自分とのギャップを埋める助けをしてくれるような気がする。(講義内容)	3.40	2.302	1.90	1.912	0.214
それは、私が望むようなアイデンティティを獲得する助けをしてくれるような気がする。(講義内容)	3.40	2.302	2.70	2.263	0.640
それと私には、たくさんの共通点があるような気がする。(講義内容)	3.00	1.581	2.30	1.567	0.408
まるで、私はそれを所有しているかのように感じる。(講義内容)	1.00	0.000	1.40	0.699	0.219
まるで、それは私の所有物であるかのように感じる。(講義内容)	1.00	0.000	1.40	0.699	0.219
まるで、私はそれに対して高い所有権を有しているかのように感じる。(講義内容)	1.00	0.000	1.70	1.252	0.133

表 2 行為主体感に関連するアンケート結果. 本システムが、学習者に講義内容を自ら発話しているかのような感覚を抱かせ、より授業に没入させることが示唆される. * は $p < 0.05$ で有意差ありを示す.

Question item	Intervention		Control		p-val
	M	SD	M	SD	
まるで講義を自分がしているかのように感じた	3.40	2.074	1.60	0.966	0.068
まるで頭の中から声がするようだった*	3.00	2.345	1.10	0.316	0.041
まるで自分自身の声を聞いたようだった*	4.20	2.387	1.20	0.422	0.010
まるで聞いた言葉を自分が発しているかのように感じた*	4.20	2.168	1.30	0.483	0.016
まるで調整された自分の声を聞いているかのように感じた	5.20	2.490	2.20	2.300	0.071

から得ているかのように感じる。(講義音声)」の 2 項目について統計的に有意な差が見られた. 講義音声に関する質問項目では、介入群のスコアは統制群よりも高い傾向があり、合成自己聴取音にも心理的所有感が生じることが示唆される. 一方で、講義内容に関しては統計的に有意な差は見られなかった.

5.3 行為主体感

行為主体感に関する質問項目およびアンケート結果を表 2 に示す. 「まるで頭の中から声がするようだった」「まるで自分自身の声を聞いたようだった」および「まるで聞いた言葉を自分が発しているかのように感じた」の 3 項目について、有意差検定を行った結果、2 群間に統計的に有意な差が見られた. また、全ての項目において介入群のスコアは統制群よりも高い傾向が示された.

5.4 感情

感情に関する質問項目とアンケート結果を表 3 に示す. 「やる気がわいた」、「不快だ」の 2 項目について統計的に有意な差が見られた.

5.5 好感度

授業内容に対する好感度に関する質問項目とアンケート結果を表 4 に示す. どの質問項目に関しても有意差は見ら

表 3 感情に関する質問項目のアンケート結果. 「やる気がわいた」、「不快だ」の 2 項目について統計的に有意な差が見られ、本システムが学習者の不快感を増大し、やる気を削いでしまうことが示唆される. * は $p < 0.05$ で有意差ありを示す.

Question item	Intervention		Control		p-val
	M	SD	M	SD	
やる気がわいた*	2.20	0.837	3.40	0.966	0.049
神経質な	1.80	1.304	1.60	0.699	1.000
イライラした	2.00	1.414	1.40	0.966	0.431
熱狂した	1.20	0.447	2.00	1.247	0.238
苦悩した	2.80	2.049	2.20	1.549	0.558
びりびりした	1.20	0.447	1.20	0.422	1.000
興奮した	1.00	0.000	2.00	1.247	0.079
うしろめたい	1.20	0.447	1.50	0.972	0.693
はずかしい	2.60	1.517	1.70	1.494	0.272
機敏な	1.20	0.447	1.60	1.350	0.930
怯えた	1.00	0.000	1.20	0.632	0.572
決心した	1.00	0.000	1.10	0.316	0.572
恐れた	1.60	1.342	1.20	0.422	0.930
うろたえた	1.80	1.304	1.30	0.675	0.431
不快だ*	2.40	1.140	1.40	1.265	0.029

表 4 講義に対する好感度に関する質問項目のアンケート結果. すべての項目で、統計的に有意な差 ($p < 0.05$) は見られなかった.

Question item	Intervention		Control		p-val
	M	SD	M	SD	
講義の内容をもっと学習してみたいと思いましたか?	4.40	2.408	5.00	1.491	0.756
講義の内容は面白かったですか?	5.60	1.673	5.30	1.889	0.847
講義をまた受けてみたいと思いましたか?	4.20	2.168	4.60	1.838	0.656

れなかった.

6. 議論

6.1 リサーチクエスションへの回答

RQ1: 合成自己聴取音に対して心理的所有感が生じるか
5.1 節から、介入群は統制群よりも高いスコアを示し、専門家の音声に比べ、自分の声を聞いているという感覚が生じていることが確認された. そのため、合成自己聴取音に対して心理的所有感が生じることが示唆される.

RQ2-1: 講義に対する行為主体感が生じるか

実験結果より、自己聴取音声で講義を聞くだけでも、発声に対して行為主体感が生じることが示唆される. また、「まるで講義を自分がしているかのように感じた」という設問に関しても強い傾向が見られることから、本システムにより、学習者に「自分が講義を行っている、あるいは講義内容を発話している」という感覚を抱かせ、学習プロセスへの能動的な関与を促進する可能性が示唆される.

RQ2-2: 行為主体感に伴い、講義内容への心理的所有感が生じるか

講義に対する行為主体感が高まったものの、講義内容に対する心理的所有感には統計的有意差が見られない. これは、自己聴取音によって行為主体感が得られても、講義内容そのものを「自分のもの」と感じるには不十分であったことが示唆される.

RQ3-1: 本システムによる講義体験が学習者の感情にポジティブまたはネガティブな変化をもたらすか

本システムは、学習者のネガティブな感情を増大させる. 介入群では「やる気が湧いた」という項目で統制群より低いスコアとなり、「不快だ」というネガティブな感情も有意に増加する. つまり、自分の声を聴くことによって、学習者の不快感が増大しやる気が削がれてしまっている. RQ1 より、講義の音声は自分のものであると感じられていることから、自分の声を聞く体験によって不快感が引き起こさ

れる原因として, voice confrontation 以外の原因があることが示唆される。

RQ3-2: 感情の変化に伴い, 講義内容に関する好感度に変化があるか

前述した通り, システムが学習者のネガティブ感情を喚起したにも関わらず, 講義内容に対する好感度に有意な差は見られなかった。この結果は, 自己聴取音による不快感が講義内容そのものの興味関心には影響を及ぼさなかったことを示唆する。つまり, 学習者は不快感を覚えつつも, 内容の有用性や面白さを独立して評価していた可能性が考えられる。

RQ4: 本システムが学習者の記憶力を向上させるか

テスト結果では, 介入群が統制群よりも高い正答率を示したが, 有意差は見られなかった。しかし, 介入群において得点が高くなる傾向が見られたことから, 合成自己聴取音が記憶定着に寄与する可能性が示唆された。自己関連付け効果や行為主体感が, 記憶プロセスをサポートしていると考えられる。

7. おわりに

本稿では, 学習者自身の合成自己聴取音声を用いた講義システム ELEVATE を提案し, その効果を検証した。結果として心理的所有感や行為主体感の向上が示唆される一方, 不快感が学習意欲を削ぐ課題が明らかとなり, 今後の改良や長期的効果の検証が必要とされる。

謝辞: 本研究は, JST 創発的研究支援事業 JPMJFR226V, JSPS 科研費 21H04900 の支援を受けて実施した。

参考文献

- [1] 平松あい et al., “コロナ禍での子どもの家庭学習におけるオンライン化の実態調査,” in *日本科学教育学会年會論文集 47*. 一般社団法人 日本科学教育学会, 2023, pp. 623–624.
- [2] 田中智之, “オンデマンド型授業の実践と課題,” *日本薬理学雑誌*, vol. 156, no. 6, pp. 330–334, 2021.
- [3] T. B. Rogers et al., “Self-reference and the encoding of personal information,” *Journal of personality and social psychology*, vol. 35, no. 9, p. 677, 1977.
- [4] 高橋麻衣子, “人はなぜ音読をするのか—読み能力の発達における音読の役割—,” *教育心理学研究*, vol. 61, no. 1, pp. 95–111, 2013.
- [5] 前田貴記, “Sense of agency: 自己意識の神経心理学,” *神経心理学*, vol. 35, no. 4, pp. 178–186, 2019.
- [6] J. B. Avey et al., “Psychological ownership: Theoretical extensions, measurement and relation to work outcomes,” *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, vol. 30, no. 2, pp. 173–191, 2009.
- [7] 中川優奈, 田中章浩, “自分と他人の声の境界は変化するか,” *電子情報通信学会技術研究報告; 信学技報*, vol. 121, no. 177, pp. 25–30, 2021.
- [8] I. Jussila et al., “Individual psychological ownership: Concepts, evidence, and implications for research in marketing,” *Journal of Marketing Theory and Practice*, vol. 23, no. 2, pp. 121–139, 2015.
- [9] J. L. Pierce et al., “Toward a theory of psychological ownership in organizations,” *Academy of management review*, vol. 26, no. 2, pp. 298–310, 2001.
- [10] —, “The state of psychological ownership: Integrating and extending a century of research,” *Review of general psychology*, vol. 7, no. 1, pp. 84–107, 2003.
- [11] 北澤涼平, “集団の心理的所有感—マーケティング研究への応用可能性—,” *マーケティングジャーナル*, vol. 42, no. 4, pp. 51–57, 2023.
- [12] R. Ohata et al., “I hear my voice; therefore I spoke: The sense of agency over speech is enhanced by hearing one’s own voice,” *Psychological Science*, vol. 33, no. 8, pp. 1226–1239, 2022, pMID: 35787212. [Online]. Available: <https://doi.org/10.1177/09567976211068880>
- [13] J. A. Fredricks et al., “School engagement: Potential of the concept, state of the evidence,” *Review of educational research*, vol. 74, no. 1, pp. 59–109, 2004.
- [14] M.-T. Wang, R. Holcombe, “Adolescents’ perceptions of school environment, engagement, and academic achievement in middle school,” *American educational research journal*, vol. 47, no. 3, pp. 633–662, 2010.
- [15] J. Parsons, L. Taylor, “Improving student engagement,” *Current issues in education*, vol. 14, no. 1, 2011.
- [16] J. Reeve, C.-M. Tseng, “Agency as a fourth aspect of students’ engagement during learning activities,” *Contemporary educational psychology*, vol. 36, no. 4, pp. 257–267, 2011.
- [17] N. Hon, N. Yeo, “Having a sense of agency can improve memory,” *Psychonomic Bulletin & Review*, vol. 28, no. 3, pp. 946–952, Jun 2021, epub 2021 Jan 7.
- [18] 全炳河, “テキスト音声合成技術の変遷と最先端,” *日本音響学会誌*, vol. 74, no. 7, pp. 387–393, 2018.
- [19] Y. Ren et al., “FastSpeech 2: Fast and high-quality end-to-end text to speech,” *ICLR*, 2020.
- [20] Y. Wang et al., “Tacotron: Towards end-to-end speech synthesis,” *interspeech*, 2017.
- [21] 卯城祐司, “英語リーディングの科学: 「読めたつもり」の謎を解く,” 2009.
- [22] M. Sugiura et al., “Passive and active recognition of one’s own face,” *Neuroimage*, vol. 11, no. 1, pp. 36–48, 2000.
- [23] P. P. Allen et al., “Neural correlates of the misattribution of self-generated speech,” *Human brain mapping*, vol. 26, no. 1, pp. 44–53, 2005.
- [24] J. Kim, H. Song, “My voice as a daily reminder: Self-voice alarm for daily goal achievement,” in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–16.
- [25] D. Kao et al., “The effects of a self-similar avatar voice in educational games,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CHI PLAY, pp. 1–28, 2021.
- [26] P. S. Holzman, C. Rousey, “The voice as a percept,” *Journal of Personality and Social Psychology*, vol. 4, no. 1, p. 79, 1966.
- [27] P. S. Holzman et al., “Voice confrontation: A bilingual study,” *Journal of Personality and Social Psychology*, vol. 7, no. 4p1, p. 423, 1967.
- [28] 森幹男 et al., “自己聴取音に占める気導音と骨導音の割合の推定,” *電気学会論文誌 C (電子・情報・システム部門誌)*, vol. 127, no. 8, pp. 1268–1269, 2007.
- [29] 落合裕貴 et al., “低域通過フィルタを用いた気導音からの自己聴取音の再現,” *研究報告音声言語情報処理 (SLP)*, vol. 2019, no. 66, pp. 1–5, 2019.
- [30] 伊藤敷 et al., “骨伝導マイクイヤホン,” *テレビジョン学会誌*, vol. 50, no. 3, pp. 351–357, 1996.
- [31] A. Radford et al., “Robust speech recognition via large-scale weak supervision,” <https://arxiv.org/abs/2212.04356>, 2022.
- [32] J. Kong et al., “VITS2: Improving quality and efficiency of single-stage text-to-speech with adversarial learning and architecture design,” *INTERSPEECH*, 2023.
- [33] I. S et al., “Development of a Japanese version of the psychological ownership scale,” *PeerJ*, vol. 10, 2022.
- [34] N. Ogawa et al., “Investigating effect of altered auditory feedback on self-representation, subjective operator experience, and task performance in teleoperation of a social robot,” in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’24, New York, NY, USA, 2024.
- [35] K. Kercher, “Assessing subjective well-being in the old-old: The panas as a measure of orthogonal dimensions of positive and negative affect,” *Research on Aging*, vol. 14, no. 2, pp. 131–168, 1992.
- [36] 川人潤子 et al., “日本語版 the positive and negative affect schedule (panas) 20 項目の信頼性と妥当性の検討; 論文,” *広島大学心理学研究*, vol. 11, pp. 225–240, 2012.