

AI 生成動画によるオンデマンド授業の実践事例 —教材コンテンツ制作の報告と学修者評価の分析—

Teaching with AI-Generated Videos Report on the Production of Educational Material Content and Analysis of Learner Evaluations

細江 哲志

Satoshi Hosoe

常葉大学

Tokoha University

Email: sat.hosoe@sz.tokoha-u.ac.jp

あらまし：本研究はオンデマンド型の動画教材の制作において、機械学習によるリップシンク技術と合成音声を活用した事例を報告し、生成されたコンテンツの学修者評価について考察する。8割以上の学修者が AI 生成による動画を受け入れる一方で、音声や顔表情の精度に違和感を感じる者が一定数存在した。本取り組みは、機械学習による生成コンテンツが高等教育に有効であることを示唆している。

キーワード：機械学習、ジェネレーティブ AI、オンデマンド教育、学習行動分析、高等教育

1. はじめに

本研究では、AI 技術を活用したオンデマンド教育動画の生成が、高等教育の現場にてどのような効果をもたらしているのか事例報告する。本研究では 2020 年に公開された Wave2Lip と呼ばれるリップシンク技術と、筆者自身の合成音声を組み合わせ、いわゆるディープフェイク動画をオンデマンド授業に組み込んだ。この動画コンテンツに対し学修者からは一部否定的な反応があったものの、多くは受け入れる態度を示し、AI 技術によって生成されるコンテンツが高等教育に有効であることが示唆された。本文では教育実践の取り組みの過程を報告し、学修者の評価についてアンケートから考察する。

2. 背景

2020 年から 2022 後期頃までの COVID-19 パンデミックをきっかけに、数々のオンライン授業の手法が実践されたことは記憶に新しい。筆者は本学にて担当する 4 つの授業（「情報科学概論 I」「情報科学概論 II」「e コマース論」「経営情報システム論」）にて、スライド資料ならびに PDF テキスト資料とその内容を解説する動画を各回に制作し、オンデマンド形式で配信した。いずれの授業も 200～300 人程度が受講する大規模クラスであった。授業動画は 60 分から 70 分程度の内容を、視聴しやすいように 7～10 分程度に区切り、さらに 20～30 分程度を要する演習や、リフレクション課題と組み合わせた学修コンテンツを設計した。

配信には Microsoft Teams と YouTube の限定配信を組み合わせる仕組みを用いた。必ずしも学修者がリアルタイム動画配信に対応するネットワーク環境を持ち合わせていないことが想定されたためである。複数のストリーミングシステムを比較し、PC でもスマートフォンでも過不足なく利用できるこれらプラットフォームの組み合わせが最適だと判断した。

3. 授業動画制作の課題と AI による解決策

筆者は動画コンテンツの制作について相応の経験とスキルがあり、授業動画の収録・編集・配信作業は問題なく進めることができた。しかし事前の解説用台詞の準備や、大勢に視聴されることを意識したセッティング（服装や、撮影する自室や研究室の背景の様子等）を整え、また学修者にとって理解しやすい動画として整合性を保つための細かい編集作業を継続することは、肉体的にも精神的にも負担となった。解説する筆者自身の様子を収録することに拘ったのは、表情やジェスチャーによる親近感の演出が、継続的な動画視聴や、理解度の向上につながると考えたからである。

2022 年度の授業において、制作済みの動画の一部を新しい事例等に合わせて修正する必要があり、効率的な編集のために AI による生成動画およびに合成音声による加工・編集を試みた。

4. 自動生成動画の制作の流れ

動画の生成に用いた AI 技術は次の二つである。Wave2Lip はメルスペクトログラム（人間の聴覚に基づいた音高の知覚的尺度による音の時間的変化情報）を用いて音声情報を解析し、映像中の任意の顔画像情報に対して、唇の動きや歯の表出を中心とした口部分の変化を自動生成する機械学習モデルである。Python で実装され GitHub にて公開されている。

合成音声の生成には株式会社 CoeFont が提供するクラウドサービス (<https://coefont.cloud/>) を用いた。事前に 150～200 分ほどの設定用台詞を読み上げることで、その人の声に似せた音声を合成することができる。完成後は Web ブラウザ画面からテキスト入力やピッチ・速度の調整等の編集をしたり、音声ファイルとして出力することができるようになる。

動画生成の流れは次の通りである。1) 過去に配信済みの動画の一部から、自然な目線や仕草で話している筆者自身の 2～3 分程度の動画クリップを事前

に抽出しておく。2) 話す内容を CoeFont サービスにテキストで入力し、合成音声を生成する。3) Wave2Lip を用いて動画クリップと生成された音声データを処理し、筆者が合成音声に合わせて喋っているかのような動画を自動生成する。

5. 学修者の評価 (1)・アンケート結果より

2022 年度後期の経営情報システム論 (二年次生以上が履修、n=207、授業 A) の第 13 回、及びに情報科学概論 II(一年次生以上が履修、n=238、授業 B)の第 15 回のオンデマンド授業において、計 20 分程度の AI 生成動画を配信した。いずれも情報通信技術をテーマにした授業であり、最新のテクノロジーについて事例紹介する目的も兼ねていた。授業後の振り返りアンケートの結果は表 1 の通りである。

表 1 アンケート結果

	授業A (n=207)	授業B (n=238)
(AIで生成した) 授業動画を視聴してどう感じたか		
いつもと同じような授業動画だと感じた	10%	11%
どちらかという、いつもと同じような授業動画だと感じた	34%	26%
どちらかという、いつもとは少々異なる授業動画だと感じた	26%	29%
いつもとは異なる授業動画だと感じた	29%	32%
過去の (リアルな教員が映っている) 授業動画と比べてどの程度理解できたか		
過去に提供されてきた動画と同様に、理解できた	26%	26%
どちらかという過去に提供された動画と同様に理解できた	52%	51%
どちらかという過去に提供された動画より理解しづらかった	15%	18%
過去に提供されてきた動画より、理解しづらかった	7%	4%
(AIで生成した) 授業動画に違和感を感じたか		
全く違和感を感じなかった	11%	12%
どちらかという違和感を感じなかった	33%	27%
どちらかという違和感を感じた	30%	33%
強い違和感を感じた	25%	28%
(AI生成による) 動画を今後もオンデマンド授業にするとしたら		
特に問題無いと思う	57%	47%
どちらかという問題無いと思う	29%	34%
どちらかという問題があると思う	11%	16%
問題があると思う	3%	3%

問い「授業動画を視聴してどう感じたか」「授業動画を視聴して違和感を感じたか」の集計結果から、「(どちらかという～)の回答も合わせると)相違点や違和感を感じた学修者が半数以上いたことがわかる。生成動画の視聴を通じて、およそ通常の動画と同じに理解できたと回答する割合は計 7 割以上いた。「このような動画を今後もオンデマンド授業にするとしたら」の問いに対し、授業 A では約 57%、授業 B では約 47%の学生が「特に問題無いと思う」と回答し、「どちらかという問題無いと思う」との回答と合わせると 8 割以上の学修者が AI で生成された授業動画を受け入れていることになる。

6. 学修者の評価 (2)・自由記述回答より

自由記述回答による感想や意見も募った。特に意見やコメントがない場合は「特になし」と入力する

よう指示した結果、445 人中 286 人 (64.3%) がそのように回答した。表 2 は「特になし」を除外した自由記述回答について、TF-IDF 法によるテキストマイニングを行った結果の一部である。

表 2 自由回答のテキストマイニング (一部)

名詞	スコア	類出頻度
「ai」	189.28	43
「音声」	82.99	40
「違和感」	73.45	55
「抑揚」	47.51	11

形容詞	スコア	類出頻度
「聞きづらい」	16.8	4
「ごこちない」	9.12	3
「残りにくい」	8.93	2
「聞き取りやすい」	7.82	3
「分かりやすい」	1.7	4

結果から、映像における独特の AI っぽさや、音声についての違和感、抑揚のなさ、聞きづらさやごこちなさを指摘するネガティブな反応を示している様子が伺える。興味深い回答の一部には「いつもの授業動画より話す内容が決まっていたスムーズだった」「何の違和感もなかったが自動生成と知って驚いた」「音声、映像が不気味で焦燥感を感じ、動画をすべて視聴できなかった」「AI による授業でも内容は理解できるが先生がいる意味がないのもっと授業料を安くして欲しい」「(比較すると)いつもの先生が喋っている方が温かみがある」などがあつた。

7. まとめと今後の展開

本取り組みについてまとめる。まず、既存の機械学習モデルを組み合わせることで、教員自身のリアルな姿ではなく、デジタルペルソナを活用した教育コンテンツの制作が、技術的には充分可能になったことが挙げられる。学修者の多くはこのような動画コンテンツを受け入れる傾向にあるが、音声の抑揚や顔表情に違和感を感じる者が一定数おり、現状では学修の阻害要因になりかねない。

近年、機械学習による画像や文章の自動生成技術の発展がめまぐるしい。本取り組みから、一連の技術を連携させていくことで、動画配信のような一方通行の知識伝達型コンテンツだけでなく、大人数の学修者に個別にかつインタラクティブに対応するアバターやデジタルペルソナが台頭してくることが示唆される。コロナ禍や、情報の自動生成技術の台頭は、「100 年変わっていない」と批判されがちであった授業形態や教室における学びにも大きなインパクトをもたらした。今後も新しい技術を積極的に取り入れつつ、適切な学修環境の提供に努めていきたい。

参考文献

- (1) Prajwal, K. R., et al. "A Lip Sync Expert is All You Need for Speech to Lip Generation in the Wild.", Proceedings of the 28th ACM International Conference on Multimedia (2020)