

ボイスロイドを用いたオンライン講義コンテンツ作成の現状と課題

八城 年伸†

安田女子大学 家政学部†

はじめに

2020年度は多くの教育機関において、新型コロナウイルス感染症(COVID-19)の感染拡大防止のため、オンライン講義の実施など様々な試みが行われた。本学においても、4月5日にGoogle Classroomを用いたオンライン講義の試行を開始し、同14日の説明会を経て、同20日には講義を開始した。のちに緊急事態宣言の解除を受け、6月1日より対面講義を再開している。

限られた準備期間で実施されたオンライン講義であったが、学生の反応は、教員の負担への同情的な意見がある一方で、声が聞き取りにくい、板書が見えにくい、など芳しくない意見も少なくない。筆者は過去の講義における感触から、これらの問題が顕在化することに加え、資料の煩雑化が問題になると考えた。板書や配付資料、参考図書、口頭での説明や補足など、それらの全てを資料にあらかじめ盛り込むことは現実的ではないためである。

想定される問題を解決するため、「ゆっくりムービーメーカー」による合成音声^{*1}を用いたオンライン講義コンテンツを作成したが、その過程で様々な問題が明らかになった。

字幕を入れるための手法

本学におけるオンライン講義は、各教員の講義を録画したビデオ教材を用いることを主とした。そのため補足説明を行うには、板書、テレビ番組におけるフリップ様のものを使用する、ビデオ編集ソフトによりワイプまたは字幕を挿入する、などの方法が考えられる。単に字幕を入れるだけであれば、自動翻訳機を簡易字幕生成器として使用することも有効である。いずれも単純に手間が増えることになり、ビデオ教材の問題点の根本的な改善は期待できない。

筆者の講義においては、板書は従前よりPowerPointを用いた資料提示に置き換えており、オンライン講義用に改変したスライドを映像化することで対処可能であった。

音声については、「えーと」などの間合いのための発声、各種の操作音、周辺の環境音、呼吸など生命活動によるノイズが問題になると考えた。のちにnVIDIA RTX Voiceノイズキャンセリングなど様々なノイズ低減手法が短期間に投入されていることからしても、同様の悩みは各国に共通であったと言える。

そのため、音声と字幕を同時に解決するために、YouTubeおよびニコニコ動画において様々な分野の解説や検証・紹介を行なっている動画である「ゆっくり解説」の手法を参考にすることを考えた。「ゆっくりムービーメーカー」を使用することで、一連の作業を一元的に行うことができるが、そうでない場合は合成音声と字幕を個々に作成する必要があり、手間と入力間違いの可能性の点で非現実的であると言える。

ライセンス

合成音声は個人が趣味で使うには廉価であるが、ビデオ教材の作成に際しては、適したライセンスが適価で購入できるかが問題となる。肉声を元にして自然な音声で作成できる株式会社AHSのVOCAROID2は、講義で使用するためには900千円の法人ライセンスが必要となる。

それに対して「ゆっくりムービーメーカー」が標準で使用する、株式会社アクエストのAquesTalkは、合成音声の不自然さはあるものの、商用コンテンツライセンスが年額5800円と、個人的な実験としてもお手頃感がある。以上からAquesTalkを使用することとした。

講義までの流れ

ビデオ教材の作成は、概ね以下の手順となり、セリフの入力にかなりの時間を要する。

- スライドを画像または動画で出力する
- タイムラインに配置してセリフを入力する(セリフは字幕+合成音声になる)
- ムービーとして出力する

Current status and issues of creating online lecture content using Synthetic Speech

†Toshinobu YASHIRO

Aesthetic Design and Technology, YASUDA Women's University

*1 VOICEROIDは株式会社AHSの登録商標であるが、音声合成技術を指す一般名詞としてボイスロイドが用いられることがある。本稿においては商標と区別するために、合成音声と記すこととする。

一例であるが、36分のビデオ教材では、セリフ入力に3時間、発声の確認と修正に2時間を要した。筆者は動画コンテンツの作成経験を有しており、作業に慣熟したオンライン講義最終週の頃の所要時間であるため、不慣れな人であれば相応に余分に時間が必要になると思われる。

入力したセリフは発音辞書に従って音声合成される。発音辞書のいずれもが、ゆっくり解説に即したネットスラングが多用されていること、収録されていない固有名詞や専門用語が少ないことから、確認と修正が不可欠である。しかしながら修正漏れを完全に防ぐことはできず、学生へのアンケートでも、イントネーションが不自然であるとする回答が29.4%、読みにネットスラングが多いとする回答が12.7%あった。

学生の反応

多大な時間で作成した真面目な内容のビデオ教材であるが、一見すると「ゆっくり解説動画」と大差ない。そのため実施に際しては、学生からの拒否反応が出ないかが気がかりであった。結果として「驚いたけど面白くてわかりやすい」という好意的なものであったことから、オンライン講義の期間中、合成音声を使い続けた。他の教職員の反応も、学生の理解が進むのであれば手法としてはあり得ると、概ね肯定的であった。

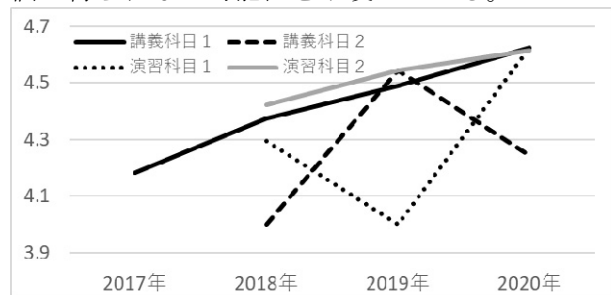
合成音声を用いることへの反応は、講義期間中に随時、そして前期終了時点でアンケートを実施した。前提として合成音声動画の視聴経験がない学生は20.3%であり、ゆっくり解説の手法を用いることについて「ふざけている」とした回答が、講義開始の時点で5.2%、講義終了の時点でも2.9%あった。反面、積極的に使って欲しいとする回答は72.8%である。

ゆっくり解説の手法を用いた目的である字幕については、58.1%の学生が最大の利点であると評価し、補助資料としての使用を含めると96.7%の学生が肯定的に捉えている。他に自由記述欄に、音声ノイズがないだけでも聞き取りやすいとした意見が7.5%あり、オンライン講義のストレスの一因となっていることが伺えた。他には眠くならなかったことを利点として挙げた意見も12.5%あった。

授業評価との関係

学生の反応が悪くないことから、講義をする側として期待するのは、学生の評価の向上である。過去の授業評価と比較できる4科目において、授業への満足度に特異な変化は見られな

かった。このことは、仮に講義改善の手法の一つとして用いたとしても、要する手間ほどには評価が得られない可能性を示唆している。



明らかになった問題点

合成音声を用いたオンライン講義コンテンツの作成を通して、大きく2つの問題点が明らかになった。

1点目は、コンテンツの再利用が難しいことである。これは動画編集ソフト全般に当てはまり、作成した動画の部分的な編集の難易度は高い。すなわち、年次の改定が必要な分野においてはビデオ教材による省力化は期待できず、改訂の頻度の低い、例えば小中高等学校の基礎的な教材には向いていると言える。合成音声を用いることで、複数の教員で分担しての教材作成が容易になると考えられる。

2点目は、できあがったビデオ教材の時間が短いことである。設定にもよるが1分間の発声文字数は約400字であるが、これは筆者の講義における発声の250字だけでなく、NHK番組制作基準である300字よりも速い。実際の講義では、発声の間合い、重要な点の繰り返し、ノートを取るためなどの待ち時間が生じるため、実効では100~120字程度になると考えられる。講義の中身が同一であっても、90分の講義を合成音声のビデオ教材にすると22~27分にしかならない。そのため、単純に時間だけで比較されると、手抜きにしか見えないことから、内容は同一であるとの証拠を示すことが必要になると考える。

まとめ

手探りで試行した、合成音声を用いたオンライン講義コンテンツ作成であったが、想定よりも好意的な反応が多く、特に字幕への評価が非常に高かった。しかしながら、今回明らかになったように、作成にかかる時間的および金銭的なコスト、コンテンツ再利用の難しさ、少数であるが拒否反応の存在という問題点を考えると、一般化できる手法とは言い難い。そのため現状においては、実験的な実施に留め、その際もシラバス等に明記することが必要と考える。