

ビデオ教材の活用を目的とした再編集ツールの開発

土永 航太郎^{*1}, 佐々木 整^{*1}

^{*1} 拓殖大学工学部

Development of a Video Re-editing Tool for On-demand Learning Materials

Kotaro Tsuchinaga^{*1}, Hitoshi Sasaki^{*1}

^{*1} Faculty of Engineering, Takushoku University

今や、オンデマンド講義は、学生に限らず身近なものになっている。私たちはこのオンデマンド講義で使用されている講義動画を活用することは重要なことと考えており、講義動画を再編集するツールの開発に取り組んだ。このツールは、講義動画の内容を変更したり削除することなく、講義動画の再生時間を学習者の希望に添うよう可能な限り短縮したり、感情認識 AI を活用して講義動画で重要な箇所と思われる部分を抜き出してダイジェスト動画を作成するものである。実際にオンデマンド講義で使用された複数の動画で試したところ、動画再生時間を約 30%短縮することができた。また、教授者が特に重要と考える講義箇所の自動検出の実現に向けた調査を行った。

キーワード: 動画の再編集, オンデマンド教材, 学習教材, 文字起こし, 感情認識

1. はじめに

新型コロナウイルス感染症のパンデミックは、私たちの生活を劇的に変えた。変化した物の 1 つに、各種教育機関でのオンデマンド講義の増加が挙げられる。オンデマンド講義は、いつでも、どこでも、何度でも受講できるという大きなメリットがある。しかし、ネットワーク環境や動画を再生するための PC などの機器が必要なことから、かつては否定的な意見も多く見られ日本ではあまり普及はしていなかった。現在では、オンデマンド講義の環境が整備され、広く利用されているため、新型コロナウイルス感染症のパンデミックが終息した後も、オンデマンド講義による指導が引き続き使用されていくと考えられる。

本研究では、オンデマンド講義で使用されている動画をさらに有効活用することを目的として、動画を再カットおよび再編集⁽¹⁾するためのツールの開発を行っ

ている。

2. オンデマンド教材作成・利用の問題点

オンデマンド講義では、対面での講義時とは異なる問題点がある。

教授者は、講義動画を録画するが、録画時間が予定の時間を超過していないかを見返す必要がある。超過していた場合には、教授者自らが講義動画の編集を行い、時間内に収まるようにしなければならない。そのためには、講義動画を見返して削除可能な部分を見つけていくことが必要であるが、一般にはこの作業は決して容易なことではない。そのため、教授者の本来の目的である教育とは別の時間に多く費やすことになる。

一方で、学習者は、無音の時間が続いている場合や「あー」や「えー」、「えっと」等の意味のないつなぎ言葉が多いと動画をスキップさせてしまい重要な部分

を見逃してしまうことがある。

3. 問題点解決するための二つの案

このオンデマンド教材の問題点を解決するために本研究では、「再生時間の短縮」と「動画のダイジェスト版の生成」の2つを目的に動画を再編集するためのツールを開発する。

3.1 無音時間に着目した再編集

講義の内容を欠落させずに、できるだけ学習者の希望する時間に近い動画に再編集するために、無音時間や「えー」、「あの」などの間投詞を削除する。また、発話の速度を測定することで、機材の操作をしながら説明をするなどして、比較的ゆっくりと話している部分があれば、その再生速度を部分的に上げることで、動画の再生時間を短縮している。

3.2 動画のダイジェスト版の生成

学習の振り返りや動画の中で特に注目すべき部分を抽出することを目的として、動画のダイジェスト版を自動的に作成する仕組みを開発している。授業ごとの動画の復習ではなく、中間試験や期末試験などに向けて、これまでの学習内容の復習を目的としたものである。

4. 動画再生時間の短縮に向けた取り組み

実際に動画時間を短縮するためには、発話する時にはどのような無音時間があるのかを調べる必要があるため調査した。その調査結果を考慮し実際に無音を削除した結果を述べる。

4.1 無音時間の調査と検出

講義中の無音時間すべてが無駄ではあるとは言いきれず、聞き取りやすさなどのために、必要な無音がある。そのため、どのくらいの無音を削除対象にするか検討が必要であると考えた。そこで、2021年のオンデ

マンド講義で実際に使用された4つの講義動画を対象に無音時間の調査を行った。

表1は、無音時間の割合を0.10秒間隔に分割した結果を示している。すべての講義動画で、0.21秒から0.30秒の無音時間が多い結果となった。しかし、教授者の癖が原因である可能性が考えられるため、テレビのニュース報道でも同様の調査を行った。その結果、0.21秒から0.30秒の間に無音部分が多いことがわかった。これらの結果から、0.30秒より長い無音部分を0.30秒残して、削除することにした。

4.2 無音時間の検出

なお、これらの無音時間を検出するために、Amazon Transcribe⁽²⁾を使用した。音声を自動的にテキストに変換することで、それを発話した開始時刻と終了時刻を得ることができる。この発話間の時間（ある発話の終わりとの次の発話の開始時間との差）を使用して、無音時間の長さを測定している。

4.3 削除結果

講義動画の一部を削除するシステムをAmazon Web Servicesを利用して構築した。動画が縮小した割合と再構成の処理に必要な時間を表2に示す。概ね元の動画時間の10%の処理時間で、約30%動画時間を削減をすることができた。この処理時間は、本システムが想定する利用場面にとって十分実用的な値であるといえる。

一方、時間短縮された講義動画が実際に学習に使えるかどうかを判断する必要がある。学習者が加工後の講義動画を見て違和感を覚え、学習者の負担となっはいけない。そこで、削除する無音時間の下限が異なる2種類の動画を用いて、15名の被験者を対象に違和感、不快感を感じるかの調査を行った。その結果、14名から0.20秒以上の無音を削除した動画に、言葉がつながって聞こえ不快に感じるという反応が確認できた。

表1 4つの講義動画の無音時間の割合

	～ 0.19s	0.20s ～ 0.29s	0.30s ～ 0.39s	0.40s ～ 0.49s	0.50s ～ 0.59s	0.60s ～
--	---------	---------------	---------------	---------------	---------------	---------

動画 1	5%	26%	14%	16%	13%	24%
動画 2	5%	28%	16%	20%	17%	14%
動画 3	8%	32%	15%	14%	10%	21%
動画 4	7%	27%	19%	12%	10%	25%

表 2 処理結果

	元動画時間	編集後の 動画時間	割合	編集 処理時間
動画 1	55 分 17 秒	38 分 42 秒	70%	7 分
動画 2	28 分 29 秒	21 分 55 秒	77%	3 分
動画 3	41 分 01 秒	32 分 48 秒	80%	5 分
動画 4	50 分 42 秒	42 分 04 秒	83%	5 分

4.4 部分的な再生速度の変更の検討

無音時間の削除で、一定の時間削減効果は見られたが、それだけでは不十分な場合には、他に比べて発話速度が遅い部分の再生速度を上げ、全体の発話速度を一定にすることを検討している。アップロードされた動画の音声は、無音時間検出の段階で Amazon Transcribe によって文字に書き起こされているので、一定時間内にどの程度発話がなされているのかは容易に検出できる。これを他の部分の発話と比べ発話速度を求める。

5. 自動カットによる違和感の調査

本研究では、動画の編集を自動で機械的に行っているため動画と動画の結合部に違和感がないかの調査を行った。この調査は、計 15 名を対象に、「編集前の元の動画」と、「0.21 秒以上の無音時間をカットした動画」、「1.0 秒以上の無音時間をカットした動画」の 3 つを試聴していただき、元動画と比べ、良かった点と悪かった点、意見や感想などを調査した。

5.1 自動カットによる違和感の調査内容

まず、「0.21 秒以上の無音時間カット」と「1.0 秒以上の無音時間カット」の動画にした理由として、0.21 秒は聞き手の違和感を考慮せずに、とにかく無音をカ

ットして動画を短縮した場合として考えた。また、4.1 章の「無音時間の調査」を行った調査より、聞き手が聞く際に、0.61 秒以上の無音カットが、違和感なく聞くことができる基準であると考え、1.0 秒以上の無音カットをすればより違和感が残らないと考え、違和感が多い 0.21 秒以上の無音をカットした動画との比較に適していると考えたためである。「0.21 秒以上の無音時間をカットした動画」は編集後、元動画から 63% の動画時間に短縮されている。同様に、「1.0 秒以上の無音時間をカットした動画」は編集後、元動画から 73% の動画時間に短縮されている。

5.2 自動カットによる違和感の調査結果と考察

調査で得られた悪い点として、大きく 2 つの意見が得られた。まず、6 名から編集後の動画は映像がカクカクして見えることが挙げられた。この点に関しては、講義動画では、パワーポイントなどを扱うことが多いため、マウスカーソルが飛ぶ程度に抑えられるが、ニュース番組のような映像が多い動画ではカクカクした動画になり違和感が残る結果となった。そして 2 つ目の悪い点として、14 名から「0.21 秒以上の無音時間をカットした動画」では、言葉と言葉が繋がって聞こえることが多く、違和感を覚える箇所が多いという意見が挙げられた。

調査で得られた良い点として、元動画と比べスムーズに見ることができ、振り返る際や移動中での試聴に効果的であることが挙げられた。これは、オンデマンド講義の、場所や時間を問わず受講ができる特性を活かすことに繋がった。そして、「1.0 秒以上の無音時間をカットした動画」は言葉と言葉の繋ぎ部分の違和感も少なく、尚且つ元動画よりもスムーズに聞くことができるといった意見もあった。

得られた意見からの考察として、やはり聞き手が違和感なく聞くためには、必要な無音時間は存在すると

