安田女子大学紀要　50，241-250　2022.

「ゆっくり解説」手法を用いたオンライン授業コンテンツ作成において音声合成エンジン変更で想定すべき問題点

八　城　年　伸

Problems to be expected when Changing Text-to-Speech Engines for Online Class Content Creation Using Yukkuri Explanation Method

Toshinobu YASHIRO

安田女子大学家政学部造形デザイン学科

要　　旨

COVID-19によるオンライン授業においては、授業をする側、受ける側の双方が不慣れなため、様々な問題が生じた。筆者はYouTube等で公開されている、様々な事柄を解説している講座動画の手法を参考に、想定される問題を解消すべく試行を行った。この過程を情報処理学会、情報教育シンポジウム2021にて公開したが、より自然な発声のために音声合成エンジンにAmazon Pollyを使用してみてはとの提案が寄せられた。現状のソフトウェア環境においては、音声合成エンジンを変更することは多くの技術的課題が存在しており、その現状と課題をまとめる。

キーワード：オンライン授業、合成音声、
ゆっくり解説

1. は　じ　め　に

2020年に大規模な感染が発生した新型コロナウィルス感染症（COVID-19）の影響では、多くの教育機関において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、その期間も１年半に及ぼうとしている。安田女子大学においても、主に緊急事態宣言の期間において、Google Classroomを用いたオンライン授業を行っている。

限られた準備期間で、学術的な知見を求める時間的な余裕もなく、主としてインターネット上に公開されている資料や手法を参考にせざるを得なかった。自らの授業コンテンツの作成に際し、従前よりe-ラーニングのコンテンツや授業アーカイブにおける問題と感じていたことに対する、実験的な要素を加えてみることを試みた。その中核となるのが、各種ノイズを抑制し、受講者のアクセシビリティを改善するために、ニコニコ動画やYouTubeに見られる「ゆっくり解説」の手法を用いることである。「ゆっくり解説」の手法でオンライン授業のコンテンツを作成する過程で、様々な問題が明らかになった。

一連の試行の過程と考察については、情報処理学会第83回全国大会［*1］、および情報処理学会情報教育シンポジウム2021［*2］にて公開した。その際、より自然な発声が得られるAmazon Polly等の音声合成エンジンを使用する提案が寄せられた。現状のソフトウェア環境においては、音声合成エンジンを変更することは多くの技術的課題が存在するが、その現状と課題をまとめる。

1-1. 元となる問題意識

まず筆者は、e-ラーニングや授業アーカイブについては否定的な立場である。現状の収録方法においては以下のような問題点があり、単純な収録では解消が困難なためである。

・各種ノイズの混入が避けられない
・理解度のリアルタイムの把握が難しく、資料の

参照部位を積極的に指示する必要がある
・理解度の確認や関心をひくための雑談、特に時事ネタが時間の経過で関連が薄まる

収録時のノイズに関しては、マイク等を含めた収録環境の改善により低減が可能であるが、その物理的な負担は小さくない。またノイズと共に気になることが多いのが、講義の端々に入りやすい「えーと」や「あー」などのフィラーワードである。過去に聴覚障害のある学生を指導した際に、各種ノイズ、特にフィラーワードを耳障りだと嫌っていたが、アーカイブの際のノイズに着目した研究はなされてない。近年、着目され始めたアクセシビリティの観点からも、各種ノイズを減らすことが必要であると考える。

また、雑談や余談を減らし、授業の内容へと限定することは、雑談に含まれる教員の個々の知見が表れにくくなり、表面的には放送大学やNHK Eテレの教育番組に近づくため、教育機関としての自殺行為になりかねない危惧がある。

1-2. 改善が必要な事項と仮説設定

先に述べた問題意識を元に、学生が授業に集中しやすい環境とするためには、聞き取りやすさの改善と、授業へ集中しやすさの改善が必要であるとの仮説を立て、実験的な手法で対処することとした。

聞き取りやすさの改善に関しては、ノイズの抑制が中核となる。スタジオなど防音設備を使用しない環境での収録では、音声のS/N比が30db程度しか確保できず、かなり聞き取りづらいものである。これには、発声を大きめにする、マイクを使用する、動画編集ソフトウェアを用いた暗騒音の除去、音声のノーマライズ処理を施すことで対処が可能であった。また、環境騒音の除去には「nVIDIA RTX Voice」を用いることで、騒音が何なのか判別できないレベルまで低減することができた［*3］。

人が主要因となるフィラーワードなどのノイズは、「気をつける」以外の対処法がなく、他にはアナウンスやナレーションのプロに依頼する、「ゆっくり解説」をはじめとした合成音声を用いるなど、発声方法そのものを変えることが必要となる。

集中しやすさの改善に関しては、学生に先回りをした提示が不可欠となる。Google Classroomのような蓄積型オンデマンドシステムでは、学生の理解度や反応をリアルタイムで探ることが、ほぼ不可能である。そのため、教員側がテキストや資料の参照部位を提示するなど、積極的に指示する必要がある。音声での指示は聞き逃しの可能性があることから、板書やスライド等の資料の他に、字幕等を併用することで改善できると考えた。

これらの要件を同時に満たす手法として参考にしたのが、「ゆっくり解説」の手法である。これを用いることで、原理的に環境ノイズおよび人が主要因となるノイズが一切発生しないこと、発声内容を字幕として画面内に表示することが可能なため、聞き取りやすさの改善と、集中しやすさの改善が同時に期待できる。

また学生が日常的に視聴するYouTubeなどの動画サイトにおける各種コンテンツでは、スマホやPCなど、各自の環境に合わせて自由なスタイルで視聴することが当たり前となっている。オンライン授業と娯楽は中身は異なるが、動画を視聴する行為は同じであることから、「ゆっくり解説」などでも真面目なコンテンツの構成や工夫は参考になると考え、取り入れることを考えた。

1-3.「ゆっくり解説」の定義

「ゆっくり解説」の発祥とされるニコニコ動画においては、以下のように述べられている［*4］。

> *ゆっくりボイスことsoftalk系の音声ソフトを使い、様々な事柄を解説している講座動画に付けられる。元は「ゆっくりしていってね！」のキャラクターが紹介する動画に付けられていたタグであったが、現在では別のキャラが紹介する作品（ゆっくりの大元となった東方Projectのキャラが紹介する作品もある）、更には紹介人物が登場せず、音声とスライドのみで内容の説明を行う作品にも広くこのタグが付けられている。*
> *（中略）*
> *また作ってみたや料理など、実況的な動画も後付けのアフレコという性質からこちらに統合されていっている模様。*

アバターが喋る形になっている動画が多いが、

ライセンスの関係から、アバターなしの動画も少なくない。明確な定義が存在しないだけでなく、動画サイトの流行の移り変わりや、ライセンス規約の変化から、時期によって前提条件が変化しており、本稿において「ゆっくり解説」は以下のように定義する。字幕に関してはオプション要件であるようだが、解説の理解促進に有用であることから、本稿では定義に含めることとした。

・合成音声でアフレコをした動画である
・様々な解説をする動画講座である
・解説内容が動画内に字幕として表示される

「ゆっくり解説」の手法を用いた動画は、ニコニコ動画とYouTubeを合わせると、一説には1000万件以上あるとされる［*5］。そのため、他の合成音声の動画と比べ、どこかで聞いたことがある、なじみがあるものであり、受講者の抵抗が比較的少ないと考えた。

2. コンテンツ作成環境と問題点

オンライン授業のコンテンツ作成において「ゆっくり解説」の手法を用いたが、その過程で明らかになった、学生の反応を含めた問題点についてまとめる。

2-1. 試験的なコンテンツ作成

動画編集ソフトを使用し、字幕を挿入するには、一般的には以下の手順となる。

・音声合成ソフトに元となるテキストを入力する
・合成音声を作成し、ファイルとして保存する
・動画編集ソフトでファイル化した音声をオーディオトラックに配置する
・字幕をテロップとして入力する
・テロップをタイムライン上に配置し、文字の書式を設定する

八城が作成した授業コンテンツにおいては、1回分の授業で平均320の字幕が発生するため、上記の手順を320回繰り返すことになる。週毎にコンテンツ作成が必要となるオンライン授業においては、一般的な手順のままでは作業量の点で非現実的と言わざるを得ない。講義を録画したビデオのノイズ低減を図り、必要なポイントにテロップを入れた方が「はるかに楽で効率的」であるが、受講生からするとコンテンツ内で字幕の有無が変化することになるため、無用な戸惑いや誤解を招くことになりかねない。

調査を進めていくと、「ゆっくり解説」の作成には「ゆっくりムービーメーカー（以下YMMと略記）」というフリーソフトウェアを使用する方法があり、一連の作業を一元的に行えることが判明した。YMMの環境構築およびコンテンツの作成方法については、「ゆずゆろぐ。」のサイトを参考にした［*6］。試験的なコンテンツ作成では、６分足らずの映像資料の作成に対し、その字幕（セリフ）の入力には２時間を要し、さらには発声内容の確認や修正に１時間以上を要した。

2-2. 学生の反応

YMMを用いることで、1-2で述べた問題点を解消できることを期待した。しかしながら「ゆっくり解説」の手法が用いられた動画は、その数が多いだけに、ふざけた内容も少なくなく、加えてAquesTalk特有の不自然な響きがあるため、実験と称しても学生が受け入れてくれるのか、という危惧があった。そのため、大学所定のフォーマットに合わせた通常のビデオ映像と、「ゆっくり解説」の手法を用いたビデオ資料を併用し、「ゆっくり解説」を用いた理由も述べた。

試行は八城の担当する授業のみで行い、その内訳と受講生、作成したビデオ資料の時間（単位:分）を表１に示す。

試行に際しては「ゆっくり解説」を継続するかの判断材料とするために、アンケートを実施した。「違和感はあるが、ありかなしかで言えばあり」とする回答が最も多く、危惧していた「止めて欲しい」とする回答は概ね５％前後であった。大学のサポート窓口にも、苦情や問合せがなかったため、２週目以降は発話部分の全てに「ゆっくり解説」の手法を用いることとした。

聴覚障害の学生については、合成音声の不自然さへの拒絶反応が予想されたが、特に気にはならず、それよりも字幕があることが嬉しいとの反応であった。また、聴覚過敏の学生についても、ノイズがないために授業に集中しやすいとの反応であったことから、実施への障害となることはなかった。

表1 受講生数と再生したビデオ資料の時間

年度	年次	科目	区分	受講生	1回	2回	3回	4回	5回		平均
2020	1	デザインと知的財産	講義	96	22	53	35	41	35		37
	1	まほろば教養ゼミⅠ	演習	37	31	18	25	29	31		27
	2	平面形状計測	演習	65	6	21	22	15	26		18
	3	製図演習Ⅱ	演習	39	8	10	18	18	17		14
	3	メディア造形演習Ⅰ	演習	66	17	69	25	26	29		33
	3	造形CADⅡ	演習	35	6	15	15	17	19		14
	4	メディア造形演習Ⅱ	演習	16	8	19	11	24	26		18
	非	情報社会論	講義	289	28	46	42	37	42	…	39
2021	1	デザインと知的財産	講義	77	51	30	32	28	39		36
	2	製図演習Ⅰ	演習	72	30	20	16				22
	2	まほろば教養ゼミⅡ	演習	35	35	13	7				18
	3	造形CADⅡ	演習	12	15	18	18	29	10		18
	4	生活総合造形Ⅰ	演習	60	4	3					4
	4	メディア造形演習Ⅱ	演習	26	17	27	4				16

最終の授業においてオンライン授業全体へのアンケートを実施した。合成音声の違和感はあるが、メリットを考えれば「ゆっくり解説」の手法が望ましいとする回答は半数を超えており、条件付きを含めて肯定的な回答は9割を越している。その理由として、字幕があることを挙げた学生は58%、音声が安定していることを挙げた学生が28%いる。字幕については、対面授業でも欲しいとする意見もあり、必要と回答した学生は86%に及ぶ。このことから、1-2で述べた仮説については、肯定的な評価が得られたと考える。

その一方で、ネットスラングやイントネーションに関する不満が多いが、これはYMMの仕組みに原因がある。入力したセリフは、読みとアクセントやイントネーションをまとめた辞書に従って音声が合成される。辞書は有志により10年以上に渡り作成・編集されたもので、その多くに「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。ネットスラングの例としては、「YouTube」が「ようつべ」、「上の方」が「うえのかた」、「主に」が「ぬしに」になるため、普通の言葉遣いを入力したつもりでも確認と修正が欠かせない。しかしながら漏れがあることは否めず、この点が不満として表れたと考える。

また、合成音声による授業コンテンツの増加に対する意見は様々で、もっと増やして欲しいとする意見が1/3を占めたが、もう少し自然な声であればとする意見も1/4を越えている。これに関連しては㈱AHSのVOICEROIDを使用する改善策を考えていたが、情報教育シンポジウム2021において、Amazon Pollyを使用する提案が寄せられた。

2-3. 発声間隔の調整と1/fゆらぎにおける差異

「ゆっくり解説」の動画のように、発声間隔なしで一気に喋るのは、10分程度のYouTube動画であればともかく、30分前後のオンライン授業の教材となると学生の集中力の低下が予想される。またノートを取るための一時停止や、聞き直しをしやすくするために、発声間隔を「ゆっくり解説」としては長めの2秒程度とした。2021年度においては、少し短くした1.5秒程度を基準としたが、これでもSSMLにおける段落後の発声間隔1.25秒に対し長めの間隔である［*7］。発声間隔に対する学生の反応は、ちょうどよいとする回答が、2020年度においては76%、2021年度においては87%であった。もう少し短くてもよいとする回答も10%あるため、SSMLにおける1.25秒を基準としてもよいと考える。

発声間隔を調整した際、間隔を機械的に揃えると外国人の朗読のように聞こえる違和感があった。そのため、一律の発声間隔とするのではなく、前後に続くセリフがある時には短めに、続くセリフがない時には長めにと、普通に日本人が喋

るときの間隔に近づけるようにした。

こうした発声間隔の変化により、音声を聞いた時の印象が変わった。この現象について、時間あたりの周波数成分が変わることから、時間軸における差異が表れるのではないかとの仮説を立て、分析を行ったところ、1/fゆらぎが大きく変化することを発見した。

適切な発声間隔かを比較するために、同一の原稿を用いて、発声間隔と発声方法を変えながら朗読させた音声を分析した。用いたのは、YouTubeにある「ゆっくり解説」動画から、BGMのないセリフだけの部分の147文字である。1/fゆらぎの分析には「Art Studio まほろば」の「ゆらぎアナライザー v1.16」を用い、各パラメータはデフォルト値とし、分析結果のλ値に注目した。

> *一般に「1/fゆらぎ」と呼ばれているのはλ=1の場合に相当します。λの値がおよそ0.9～1.1の範囲に入っていれば、その楽曲は1/fゆらぎを持っているということができます。λの値が大きい（一般的には2を超えることはほとんどありません）ほどゆらぎの程度が小さく、規則性が高いことを示します。またλの値が0に近いほどゆらぎの程度が大きく、ランダム性が高いことを示します。*［*8］

元の動画は「ゆっくり解説」に多い一気に発声するタイプの動画で、λ=0.661である。これを2秒間隔の発声にするとλ=0.834となり、さらに2種類の声質を用いて会話に近づけたところ、λ=0.839が得られている。また筆者の声では、単純に朗読した際はλ=0.633であったのに対し、授業に近い発声にするとλ=0.850が得られている。

以上のことから、発声間隔の調整することにより、人が心地よいと感じるか否かが変化する結果となった。このことは、ナレーションのように、聞き取りやすい声であったとしても、一方的に淡々と喋るだけの内容は心地よいとは言えず、弾んだ会話の調子に近づけることで聞き心地のよさが増すと考えられる。

3. 音声合成エンジンの変更

PCにおける声の合成は、1980年代後半になり、IBM PC/AT互換機におけるフリーソフトウェアの形で、様々なソフトウェアが作成されてきた。それらは、映画「2001年宇宙の旅」のHAL9000をモチーフとしたもので、金属的な破裂音を多く含む。現在の音声合成エンジンにおいても、サンプリングに依らないパラメータ合成式では、同様の傾向がある。

合成音声が大きく変化したのは、2007年に発売された初音ミクである。ヤマハが開発した音声合成システム「VOCALOID」を用い、歌声を合成することが可能になった。初音ミクのしゃべり方は流暢とは言い難いが、これはソフトウェアのターゲットが声なのか歌なのかで、音声合成アルゴリズムが大きく異なるためである。そのため、歌のエンジンに喋らせる、声のエンジンに歌わせるなどの使い方をしても、手間の割に質がよいものは得られにくいとされる。音声の場合には、音声に特化した合成エンジンを使用することが必要である。

現状のソフトウェア環境においては、2-1で述べた手順を効率化してあるソフトウェアであるほど、特定の音声合成エンジンとの結びつきが強い。そのため、音声合成エンジンを変更することは多くの手間を要する。手間をかけてまで改善する価値があるか、2-3で述べた1/fゆらぎを踏まえて検討する。

3-1. 音声合成のライセンス体系

音声合成エンジンに関しては、ゼミの学生が参加したアイデアコンテストの指導で調査を行っていた。アイデアの内容が、方言での観光案内を合成音声で行うものであったため、発声の自然さ、合成の所要時間、イントネーションなどの変更方法、ライセンス体系など、商業利用を前提とした調査とした。

現在の音声合成エンジンは、自動応答音声、操作ガイダンス、朗読を主な用途としており、OEM供給が多く、単独の製品として利用できるものは少ない。さらにはナレーションや朗読などの長時間の使用を想定してない製品もあり、選択肢は限られる。

オンライン授業コンテンツ作成において最も問題となるのが、そうした向き不向きに加え、オンライン授業が商業利用に該当するのか、その際のライセンス期間とライセンス料である。生成した

音声に対して、ライセンス元が所有権を有するケースもあるため、ライセンス内容の精査は重要である。以下に個々の調査結果を記すが、オンライン授業コンテンツに使用できないライセンス体系のもの、市販されてないもの、特定のソフトウェア環境でのみ使用できるものは除外した。

3-1-1. AquesTalk

SoftTalkが使用する標準音声ライブラリである。㈱アクエストのAquesTalkは、パラメータによる合成音声で、自然な発声とは言い難いが、バージョンによっては32KBのライブラリで動作するなど、負荷が低いため組み込み機器などで用いられている。合成に要する時間は、ほぼ瞬時である。アクセントやイントネーションを変更するための記述方法はドキュメントとして公開されているが、その多くが記号であるため可読性と記述性はよいとは言い難い。オンライン授業には商業コンテンツライセンスが必要となり、年額で￥6,380である。

3-1-2. VOICEROID

㈱AHSのVOICEROIDシリーズは、サンプリングベースの合成音声であり、多様な声優を用いたキャラクターボイスとしての音声が特徴である。音声合成エンジンは㈱エーアイのAITalkのOEMである。発声が自然であるが、システム負荷は大き目である。合成に要する時間は１～２秒程度であり、ワンテンポのタイムラグがある感じである。基本的にソフトウェア画面での操作となり、イントネーションの変更は専用の画面にて、音ごとに高低を調整する。オンライン授業コンテンツとしては、ソフトウェアの購入の他に商用ライセンスが必要となる。教育機関に所属する場合は法人ライセンスとなり、永年ライセンスで￥990,000である。

3-1-3. 声 の 職 人

グラスバレー㈱のビデオ編集ソフトであるEDIUSに、合成音声によるナレーションを付加するためのソフトウェアである。音声合成エンジンは㈱エーアイのAITalkのOEMである。イントネーションなどの調整はVOICEROIDと同様である。基本ライセンスで商業利用が可能であり、EDIUS X Proが￥24,800、声の職人２ for EDIUSが￥33,000となる。

3-1-4. 音 読 さ ん

COMOMOの音読さんは、Webサービスのためブラウザのみで使用可能であり、特定のソフトウェアのインストールを必要としない。発声が人の声と大きく変わらない自然なものである。合成に要する時間はサーバーの負荷にも依るが３～４秒程度であり、やや待たされる印象がある。イントネーションの変更はSSMLを用いる。オンライン授業コンテンツとしては、クレジットを記載することで無料での使用も可能であり、料金プランにより月間の発声可能字数が異なる。他の音声合成エンジンと比べてライセンス内容に曖昧な点が多く、商業利用には不安が残る。

3-1-5. Amazon Polly

アマゾンのAmazon Web Service（AWS）に含まれる。文章を音声に変換するサービスで、ブラウザから利用する他に、公開されているAPIを用いてアプリケーションを作成することも可能である。合成に要する時間は、ほぼ瞬時である。イントネーションの変更はSSMLを用いる。処理した字数による従量課金となり、単純な発声は廉価なものの、SSMLによるマークアップも字数にカウントされるため、凝った発声をすると高額になる可能性がある。作成した音声の継続利用には追加の料金は発生しない［*9］。

同様のクラウドサービスはMicrosoftもAzure Speech Serviceとして提供しているため、用途に応じて選択が可能である［*10］。

3-2. ゆっくりムービーメーカーにおける音声合成エンジンの変更

YMMで使用する音声合成エンジンを変更することは、幾多の方法が試みられてきた。その一つである「VoiceroidUtl」を用いると、2-1で述べた手順を自動で行うことができる［*11］。

YMMの標準的な操作に比べ、アクセントや読みの修正はVOICEROIDの操作で行う必要がある点が煩雑になる。YouTubeに趣味的に公開する動画であれば問題になることは少ないと思われるが、オンライン授業となると、ネットスラングの修正が相当な負荷になると予想される。加えてライセンス料金が高額であることから、オンライン授業コンテンツの作成には向いているとは言い難い。

ネットスラングや読み間違いの少ない、精度の高い、Amazon Pollyなどの音声合成エンジンであれば、修正作業を大幅に減らすことができる。しかしながら2021年８月の時点で、YMMでAmazon Pollyを使用するための連携ソフトウェアは存在しておらず、2-1で述べた手順を手動で行う必要があるため、作業負荷は極めて大きなものとなる。APIが公開されているため、連携ソフトウェアを作成することで作業負荷の改善が見込めるが、そうでない場合の利用は非現実的と言わざるを得ない。

結論として、音声合成エンジンはソフトウェアに密接に関係しており、変更することは容易ではない。仮に変更をするのであれば、以下の５項目について総合的な判断をする必要があると言える。

・ライセンス料の負担に耐えられるか
・アクセントや読みの修正が頻繁に発生するか
・ソフトウェアの開発能力があるか
・新たなソフトウェアの操作が習得可能であるか
・一連の作業がシームレスに行えるか

3-3. 1/fゆらぎにおける差異

3-1で述べた音声合成エンジンを用いて、2-3で用いた原稿を発声し、それらの1/fゆらぎを比較した。総じて自然な発声であり、それだけを聞くと好感が持てるが、1/fゆらぎには異なる結果が表れた。最終的には受講する学生による比較評価が必要ではあるが、カーナビの案内音声や自動応答音声のように短時間のアナウンス向けであって、オンライン授業で長時間を聞き続けるのに向いているかは疑問が残った。

3-3-1. VOICEROID2 紲星あかり

声優による一般的なナレーションに聞こえる。しかしながらλ=0.456と低値であり、長時間の利用には工夫が必要と言える。YouTubeにおける動画では印象の異なるものもあるため、パラメータの調整で改善される可能性がある。しかしながら、細かな調整にはVOICEROID2上における操作が必要であり、高額なライセンス料を考えると、実用に向くかは疑問である。3-2で述べたように、YMMと連携するためのソフトウェアが公開されており、比較的シームレスな操作が可能である。

3-3-2. 声の職人 かほ

音声合成エンジンが同一ＯＥＭであるVOICEROIDと声の傾向は似ている。サンプル音声の分析しかできなかったが、λ=0.596であり、VOICEROIDよりは聞きやすい印象であった。細かな調整にはソフトウェア上での操作が必要である他、字幕の挿入は別作業として行う必要がある。EDIUSそのものの操作方法を習得する必要があるため、導入のハードルが低いとは言い難い。

3-3-3. 音 読 さ ん

3-3-1よりはハキハキとした調子のナレーションで、λ=0.606と一般的な「ゆっくり解説」と変わらない結果となった。SSMLによる抑揚の付加などで、さらに改善できる見込みがある。基本的にブラウザによる操作になるため、使用するには2-1に述べた手順をそのまま行う必要があり、作業の煩雑さが問題となる。

3-3-4. Amazon Polly

2021年９月の時点で日本語ではニューラル音声が使用できないためか、Amazon EchoのAlexaがナレーションをしていると形容するしかない音声である。Alexaのナレーションを30分～１時間に渡って聞き続けることを想像してみると、受け入れられるか否かが人により大きく異なることが予

想される。λ=0.388と他の音声合成エンジンと比較しても低値であり、将来のニューラル音声の提供を期待するか、SSMLによる細かな調整が必要になると思われる。

4. 考 察

「ゆっくり解説」の手法を用いたオンライン授業コンテンツの作成については、YouTube等の視聴で慣れている学生が多く、大きな抵抗はないと予想しての試行であった。大きな抵抗はなかったものの、想定したほどに視聴し慣れている訳ではなく、作成にも多大な手間がかかるため、手間に見合う効果が得られたのかについては疑問が残った。

受講した学生へのアンケートでは、もう少し自然な声の方が望ましいとの意見も少なくないことからも、情報教育シンポジウム2021における、Amazon Pollyを使用してみてはとの意見に繋がったと考える。現在、「ゆっくり解説」などの動画作成を除くと、動画に合成音声のナレーションをアフレコするニーズは一般的であるとは言い難い。オンライン授業へ応用するための工夫の余地も限られているため、音声合成エンジンの変更の可否を検討してみた。

その過程で明らかになったのが、５項目について総合的な判断を行う必要があることである。特に作業がシームレスに行えるかは重要な判断基準となる。八城の場合、１回分のオンライン授業コンテンツを作成するために４～７時間を要している。一つのセリフにつき５秒の余分な作業が発生するだけで、所要時間が30分増加することを考えると、日常的なオンライン授業コンテンツの作成に耐えるためには必須と言える。

加えて、聞きやすい音声であるか否か、聞き慣れている声であるか否かも大きく作用する。八城の作成したオンライン授業コンテンツへは大きな拒絶反応はなかったが、これは「ゆっくり解説」の手法を取り入れただけでなく、多くの「ゆっくり解説」に合致したフォーマットとなったことから、学生側に違和感が少なかったためと考える。そのため音声だけを改善したとしても、それが全体的な評価に結びつくかは疑問が残る。

オンライン授業に限定すれば、より自然なコンテンツを作成するための改善と、COVID-19に社会環境が対応できるかの競争となる。実験を繰り返す意味と時間があるかは微妙であるが、しかしながら「ゆっくり解説」を用いた授業コンテンツには、小中高等学校における自習教材としての可能性が考えられる。今後も効率よく自然な教材を作成するための手法を研究していきたい。

謝 辞

コンテンツの作成ならびに分析に使用したソフトウェアの作者の皆さま、本稿の査読をして頂いた皆さまに，謹んで感謝の意を表する。

参 考 文 献

1．八城年伸、『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』、情報処理学会 第83回全国大会講演論文集(4)、2021、pp381 ～ 382.
2．八城年伸、『「ゆっくり解説」手法を用いたオンライン授業コンテンツ作成に係る考察』、情報処理学会 情報教育シンポジウム論文集、2021、pp53 ～ 60、（参照 2021-8-21）
https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=212340&item_no=1&page_id=13&block_id=8
3．中村真司、『NVIDIA、GPUで打鍵音や環境騒音を除去する「RTX Voice」』、https://pc.watch.impress.co.jp/docs/news/1248195.html、（参照 2020-4-20）
4．『ゆっくり解説とは』、https://dic.nicovideo.jp/a/ゆっくり解説、（参照 2021-8-1）
5．『“ゆっくり解説”のキャラクターって著作権は大丈夫？ ガイドラインを守れば誰でも使える“ゆっくり”の歴史を紹介してみた』、https://originalnews.nico/280583、（参照 2021-1-4）
6．ゆずゆろぐ。、『ゆっくり実況の作り方』、https://yuzuyu3.com/yukkuri-start/、（参照 2020-4-14）
7．Alexa Developer Documentation、『音声合成マークアップ言語（SSML）のリファレンス』、https://developer.amazon.com/ja-JP/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html、（参照 2021-8-30）
8．Art Studio まほろば、『解析結果の評価』、https://mahoroba.logical-arts.jp/archives/202、（参照 2021-6-13）
9．Amazon Polly、https://aws.amazon.com/jp/polly/、（参照 2021-8-30）
10．Microsoft Speech SDK、『Speech Serviceのドキュメント』、https://docs.microsoft.com/ja-jp/azure/cognitive-services/speech-service/、（参照 2021-8-

30）
11. わがまま趣味な自己啓発Blog、『ゆっくりムービーメーカー4にボイスロイドを連系導入してみよう！』、http://kozi001.com/2020/06/19/howto-use-voiceroid/、（参照 2021-8-30）

〔2021. 9. 16 受理〕

コントリビューター：谷口 和弘 教授
（造形デザイン学科）