

「ゆっくり解説」を用いた教材作成における 音声合成エンジン変更の試み

八城年伸¹

概要：COVID-19 対策のオンライン授業において、筆者は YouTube 等で公開されている講座動画の手法を参考に、予想した様々な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 にて公開したところ、音声合成エンジンに Amazon Polly を使用してはどの提案が寄せられた。現状においては、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、変更することは技術面および操作面に多くの課題が存在していることが明らかとなった。

キーワード：オンライン授業、合成音声、ゆっくり解説、音声合成エンジン [**]

An Attempt to Change the Text-to-Speech Engine in the Creation of "Yukkuri Explanation Method" Contents

TOSHINOBU YASHIRO^{†1}

Abstract: In the online class for COVID-19, the author tried to solve various problems that could be expected. At the core of this project was the method of course videos available on YouTube. When I presented this process to the public at SSS2021, I received a suggestion to use Amazon Polly as the text-to-speech engine. Currently, the speech synthesis engine is closely related to the software, and it is clear that there are many technical and operational challenges in changing it.

Keywords: Online Lecture, Synthetic voice, Yukkuri Explanation, Text-to-Speech Engine [**]

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イルス感染症(COVID-19)の影響で、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た。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た。

限られた準備期間で、学術的な知見を得る時間的余裕がなく、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、自らの授業コンテンツを作成した。従前より授業コンテンツにおいて課題と考えていた、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」の手法を実験的に取り入れてみた。一連の過程と、明らかになった問題点、それに対する解決策と考察については、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、より自然な発声を得られる Amazon Polly を使用しては、との提案が寄せられたため、それらを変更するための検討と試行を行ったが、実現が容易であるとは言いがたい結果となった。一連の過程については安田女子大

学紀要[*3]に投稿したが、その後の追加の試行により、ソフトウェアに密接に関係する音声合成エンジンを変更することは、技術面および操作面に多くの課題が存在しており、かかるコストに見合う効果が得られにくいことが明らかとなった。

1.1 元となる問題意識と仮説

現状のコンテンツ作成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である。

- ・フィルターワードを含む、各種ノイズの混入や残留が避けられない
- ・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

過去に聴覚障害のある学生を指導した際に、各種ノイズにより集中が削がれると訴えていたが、授業コンテンツ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た。コンテンツの長期に渡る使い回しを考えなければ、雑談や時事ネタについては問題としての重要度が低いため、主としてアクセシビリティの改善

¹ 安田女子大学
Yasuda Women's University

を重要視することとした。

各種ノイズの混入対策は、スタジオなど防音設備を使用することが望ましいが、収録後であっても動画編集ソフトウェアや、環境騒音の除去ソフトウェアを使用することで対処が可能である[*4]。COVID-19の影響の長期化により、Zoomをはじめとした各種ソフトウェアも改善され、2021年上半期には特殊な環境を用いなくてもノイズの低減が可能となっている。

資料参照の改善に関しては、Google Classroomのような蓄積型オンデマンドシステムでは、リアルタイムの状況確認が困難であることから、教員側が積極的に指示する必要がある。音声による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。こちらも各種ソフトウェアの改善により、2021年上半期には特殊な環境を用いなくても字幕の提示が可能となった。

ソフトウェアの改善で対処済みである問題点に対し、人が主要因となるフィルタ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、合成音声を用いるなど、発声方法そのものを変えることになる。

2020年の時点において、これらの問題点を同時に解決するために参考にしたのが、「ゆっくり解説」の手法である。原理的にノイズが発生しないこと、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、アクセシビリティの改善が期待できる。

学生が日常的に視聴しているYouTubeなどの動画サイトのコンテンツは、スマートホンや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」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。

1.2 「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[*5]。

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る模様。

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と共に前提条件も変化しており、前稿において「ゆっくり解説」は以下のように定義した[*2]。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、定義に含めた。

- ・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[*6]。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 コンテンツ作成手順

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の字幕を使用した。非常勤を含む担当授業数が9コマあったため、上記の手順を週ごとに2500回強も繰り返すことになり、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に、必要なポイントに資料参照の指示の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ため、少し気が緩んだだけで見落とす可能性がある。調査を進めたところ、「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から、YMMを用いたコンテンツ作成を試行した。

同様のフリーソフトウェアとしては「ゆくサク」があるが、後で述べるアクセントや読みの修正の容易さから、YMMを使用した。

2.2 学生の反応

YMM を用いることで、1.1 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、ふざけた内容も少なくなく、加えて音声合成エンジンの AquesTalk 特有の不自然な響きがある。学生が受け入れられるかの危惧があったが、試行に対する学生の評価は、聴覚の障害の有無によらず好意的なものであった。

「ゆっくり解説」の手法を用いたことに対しては、違和感はあるが、単純な講義の録画ビデオより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、字幕を挙げた学生は86%に及び、音声安定していることを挙げた学生が28%いた。このことからアクセシビリティの改善については、一定の効果があったと考える。

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おり、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。

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、もう少し自然な声であればとする意見が1/4を越えている。これに関連しては「ゆっくり解説」動画での採用例が多い、(株)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。

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歌声と比較すると流暢とはいえないが、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。

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ため、それに見合う効果が得られるのか検証した。

3.1 音声合成エンジンの変更における評価項目

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「VoicroidUtl」を用いると、2.1 で述べた手順を自動的に行うことが可能である[*7]。そうした手順の簡便さに加え、音声合成エンジンの変更で生じる問題点として、以下のポイントを比較検討した。

3.1.1 アクセントや読みの修正

アクセントや読みを変更したい場合、および修正が生じた場合、その手順は作業量に直接的な影響を与える。動画投稿サイトに趣味的に公開するコンテンツであれば問題になることは少なくとも、授業コンテンツの作成であればネットスラングの確認と修正が重荷になると予想される。YMM の同種のフリーソフトウェアである「ゆくサク」を用いなかったのは、この点を主に考慮したものである。

3.1.2 音声合成のライセンス体系

現在の音声合成エンジンは、自動応答音声、操作ガイド、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多くない。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、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。そのため、オンライン授業コンテンツに不向きなライセンス体系のもの、OEM 供給のみで市販されていないもの、特定のソフトウェア環境に依存するものは除外した。

3.1.3 1/f ゆらぎにおける差異

音声合成エンジンは、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り、オンライン授業として長時間の聴取に向くかは、受講する学生による比較評価が有効である。しかしながら相対的な評価になることから、定量的に比較するために、同一の原稿を発声し、それらの1/f ゆらぎを比較した。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、 $\lambda=0.661$ である。比較対象は、総じていずれも自然な発声であり、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。

3.2 音声合成エンジンの評価結果

3.2.1 AquesTalk

SoftTalk が使用する標準音声ライブラリである。(株)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り、SSML との互換性はないため、可読性と記述性、将来への互換性はよいとは言い難い。オンライン授業での利用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。

1/f ゆらぎによる評価は、元の動画は $\lambda=0.661$ であったが、センテンスごとに人の会話に近い間隔を空けると $\lambda=0.834$ となり、さらに 2 種類の声質による会話に近づけたところ、 $\lambda=0.839$ が得られている。他の合成音声と比べ、1/f ゆらぎによる評価が高く表れた。

3.2.2 VOICEROID (継星あかり)

(株)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の特徴である。音声合成エンジンは(株)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nten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。非常勤の場合のみ、個人ライセンスとなる。

VOICEROID シリーズは、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 1/f ゆらぎは $\lambda=0.456$ 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 VOICEROID 上における操作が必要であり、加えてライセンス料が高額であることを考えると、実用に向くかは疑問である。3.1 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。

3.2.3 声の職人 (かほ)

グラスバレー(株)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(株)エーアイの AITalk の OEM であり、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。

音声合成エンジンが同一 OEM である VOICEROID と声の傾向は似ており、1/f ゆらぎの分析はサンプル音声で行えなかったが、 $\lambda=0.596$ 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ため、導入のハードルが低いとは言い難い。

3.2.4 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、字数が多いと発声そのものに失敗することがあることから、安定的な商業利用には不安が残ると言わざるを得ない。

「VOICEROID (継星あかり)」よりはハキハキとした調子のナレーションで、 $\lambda=0.606$ と AquesTalk を用いた一般的な「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。

3.2.5 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、ほぼ瞬時であり、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[*8]。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り、競争による今後の発展が期待できる[*9]。

元がビジネス用途であるため、ネットスラングや読み間違いが少なく、修正作業を大幅に減らすことが可能である。専門用語や略語が少ない授業であれば、ほぼ無修正での運

用も可能であると思われる。しかしながら 2021 年 11 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、「音読さん」と同じく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。

2021 年 9 月の時点で日本語ではニューラル音声を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。1/f ゆらぎの分析は $\lambda=0.388$ 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考える。

3.3 1/f ゆらぎによる評価の妥当性

音声合成エンジンの比較に際し、過去に作成した授業コンテンツについても検証した。短時間では聞きやすいと感じた音声、長時間の比較をすると、しつこく感じたことがきっかけである。

八城の朗読は $\lambda=0.633$ 、問いかけを混ぜた発声では $\lambda=0.850$ であったが、オンライン授業用の 45 分のビデオでは $\lambda=0.387$ と、かなりの低値であった。既に述べたように、ナレーションのように短時間では聞き取りやすいと感じる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音声とは言い難い。概ねコンテンツの長さに反比例するかのように低下する傾向が見られた。

反面、AquesTalk の発声は授業 20 回分の平均で $\lambda=0.451$ 、標準偏差 0.157 であった。他の音声合成エンジンも同様の傾向であるのかは追試が必要であるが、実験には多大な時間を要すること、ゆらぎアナライザー[*10]による解析結果を評価に用いることの妥当性には、他の事項の検討も必要となるため、稿を改めたい。

3.4 AquesTalk の優位性

自然とは言い難い AquesTalk の発声であるが、他の音声合成エンジンと比較をしてみると、欠点の少なさが明らかになった。

- ・一説に 1000 万件以上あるとされる動画の多さで聞き慣れている
- ・ライセンス料が比較的安価である
- ・YMM などのソフトウェア環境が揃っている
- ・作り方の差異が表れにくい傾向がある
- ・ほどほどに聞きやすい 1/f ゆらぎが得やすい

東方 Project のキャラクターと結びついており、若年層を中心に、「ああいう声」と認識されている点は強みと言える。

これは諸刃の剣であり、最新バージョンの AquesTalk10 においては声が変わるため、実質的に今後の発展が期待できず、現時点の技術水準に留まり続けることになる。他の音声合成エンジンの改良が進むと、その変化に取り残される危険性がある。しかしながら、改良に要する時間と、既存コンテンツが置き換わるのに要する時間を考えると、当面は優位性を保ち続けると考える。

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、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、手間に合う効果が得られたのかについては疑問が残った。

今回、Amazon Polly を使用してはとの意見により、音声合成エンジンを変更することで、もう少し自然な声にすることを試してみた。結論としては、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 6 項目について総合的な判断をする必要があると考える。

- ・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
- ・作成されたコンテンツの視聴者が、合成音声が妥当であると評価するか

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 秒の追加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」のコンテンツのフォーマットに合致した形となったことから、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未知数である。

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、作成者による差異が表れにくい、

複数の担当者による分担制作が容易であり、頻繁なアップデートを必要としない補助教材作成としての可能性が考えられる。この利点を活かすべく、今後も効率よく自然な音声の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、情報処理学会情報教育シンポジウム 2021にて意見を頂いた皆さまに、謹んで感謝の意を表す。

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] 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おいて音声合成エンジン変更の現状と課題』「印刷中」、安田女子大学紀要、2021
- [4] 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)
- [5] 『ゆっくり解説とは』、<https://dic.nicovideo.jp/a/ゆっくり解説>、(参照 2021-8-1)
- [6] 『"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)
- [7] 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)
- [8] Amazon Polly、<https://aws.amazon.com/jp/polly/>、(参照 2021-8-30)
- [9] 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)
- [10] 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)