7G-02

音声合成エンジンの変更が「ゆっくり解説」オンライン講義教材の評価に与える影響について

八城年伸†

安田女子大学　家政学部†

はじめに

COVID-19 対策のオンライン講義において、筆者は主としてアクセシビリティの改善のため、「ゆっくり解説」の手法を用いたオンライン教材を作成し、講義を実施した。講義に対する受講生の評価は、字幕があることへの評価は高かったものの、各種ノイズの排除を目的にした合成音声の使用については、発声がより自然であれば受け入れやすくなるとするものであった。

一連の試行について、情報処理学会第 83 回全国大会[1]および情報教育シンポジウム SSS2021[2]において報告したところ、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。

現在のソフトウェア環境においては、Amazon Polly を使用するためには主として作業コストの問題が存在している。国産の音声合成エンジンの開発が停滞している現状を鑑みると、選択肢の一つとして手法の開発は必要であると考え、第 162 回 CE 研究発表会[3]において報告した。作業コストに見合う改善効果が得られるかは、受講生の評価が一つの指標となる。音声合成エンジンの変更が評価にどのように表れるのかの調査を実施し、その結果と考察を報告する。

コンテンツ作成環境と問題点

合成音声のナレーションを用い、ビデオ教材に字幕を挿入するには、動画編集ソフトを使用すると一般的には以下の手順となる。

・ナレーションのテキストを作成する
・合成した音声をファイルとして保存する
・ファイル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

2021 年には YouTube、Adobe Premiere、Zoom 等に字幕起こし機能が実装されたが、フィラーワードを含めた話者の発声に大きく左右される、専門用語への対応、の点で現時点では実用に耐えるとは言いがたい。そのため、従前の手順でコンテンツを作成すると仮定すると、筆者の作成したオンライン教材においては、1 回の講義で平均 320 の字幕となり、最多では 1151 の字幕を作成していた。すなわち、わずかな手間が増えるだけでも、作業時間全体に与える影響は大きくなる。

年度	科目	受講生	1回	2回	3回	4回	5回	平均
2020	デザインと知的財産	96	322	795	1151	850	402	704
	情報社会論	289	263	713	322	534	300	319
2021	デザインと知的財産	77	402	248	332	280	411	334

試験的に２行のテキストで作業時間を分析したところ、ゆっくりムービーメーカーでは 58.5 秒であったのに対し、Amazon Polly を使用したところ 149 秒を要した。単純に比例すると仮定すると、それぞれ 2 時間 36 分と 6 時間 37 分となり、約 4 時間の余分な時間を要する。

このことは、４時間分の手間に見合った評価が得られない限り、Amazon Polly を使用するベネフィットが得られないことを意味する。

ゆっくり解説（AquesTalk）を用いた動画は、一説には 1000 万本以上存在するとされている。若年層を中心として、YouTube 等の動画サイトで視聴し慣れていること、自然とは言い難い発声ではあるものの「ああいう声」と認識されている点は強みであるため、Amazon Polly には不利な条件が揃っていると言える。

評価環境

受講生による評価は、2021 年 12 月に、広島市立大学において、筆者が担当する非常勤講義科目において実施した。有効回答は 93 件、出席登録に対する有効回答率は 53.8%であった。

この科目を対象としたのは、対面の講義科目であること、受講生が 200 名を超えていること、前年に作成したビデオ教材を副教材としてオンライン授業システムにおいて公開しており、対面の講義との対比も可能であるためである。ただし、副教材のビデオ教材を視聴している学生は、出席登録学生の 30%前後に留まる。

The Impact of Changing Text-to-Speech Engines on the Evaluation of Online Lecture Content

† Toshinobu Yashiro・Yasuda Women's University

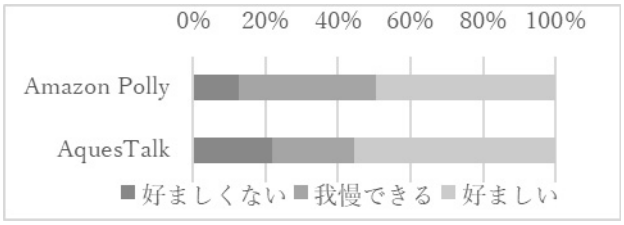
評価項目

評価は WebClass のアンケート機能を用いた無記名式とした。設問数は 11 で、Amazon Polly、ゆっくり解説（AquesTalk）のそれぞれに対する発声の印象、２種類の音声の使い分けの印象を尋ねた。また、どちらの音声が、授業に対する熱意を感じるか、自然に聞こえるか、オンライン授業にふさわしいか、を尋ねた。

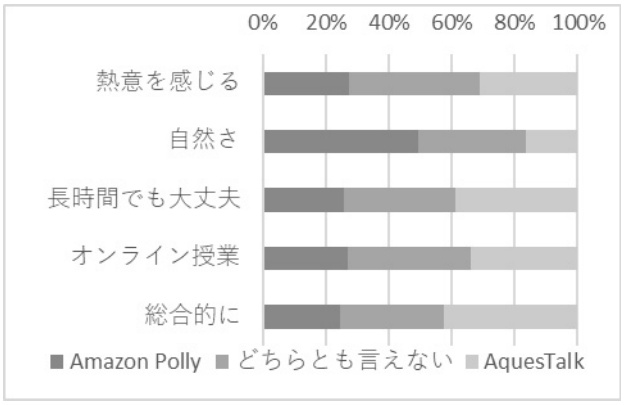
評価結果

前提として、YouTube 等における「ゆっくり解説」動画の視聴経験は、調査時期により多少異なるが、概ね６割の学生が日常的あるいは時々の視聴をしている。

発声そのものの印象は、特有の金属的な響きを持つ AquesTalk を好ましくないとする意見が優勢であるが、不自然さの割には健闘していると言える。



どちらが好ましいかとする設問では、傾向が大きく異なった。



Amazon Polly の方が自然な発声であると評価する一方で、オンライン授業向きであるかについては回答が拮抗している。さらには長時間でも聞き疲れないか、総合的にみて好ましいか、においては、AquesTalk の方を好ましいとする意見が優勢である。

自然な声であるのにオンライン授業には向かないとする原因としては、女性の声に対して、男性の声の不自然さが耳障りである、と自由記述にあったことが手がかりとなる。自然とされる女性の声も、Amazon Alexa が喋っていると形容するしかない、いかにも合成音声的なナレーションであることを苦手とする意見があった。

考察とまとめ

第 162 回 CE 研究発表会において、「ああいう声」と認識されている点に優位性があり、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したが、受講生の評価はそれを裏付ける形となった。

回答率の低さを考えると、Amazon Polly の評価も変わってくる可能性があるが、少なくとも現状においては、時間コストに見合う評価であるとは言い難い。一連の作業を効率化するソフトウェアを開発するなどして、時間コストを低減する工夫をしない限り、音声合成エンジンを変更することのメリットは少ない、と結論を出さざるを得ない。

それより今回の調査で気になったのは、どちらとも言えないとする回答が予想以上に多く、さらには自由記述や、他のアンケートにおいても、合成音声を用いることそのものが YouTube っぽくて授業としては好ましくない、とする意見が増加していることである。アクセシビリティの改善に対する評価も下がってきており、COVID-19 に対する緊迫感の差異も影響していると思われるが、合成音声の細かな差異などの目先の点を気にするより、肉声を用いないことそのもののメリットを訴求する重要性を再認識した。

現状ではキワモノ的な実験の域を出ないが、字幕、ノイズの少なさといったアクセシビリティの改善に加え、複数の制作者による差異を吸収できる、自学自習のための教材を作る手法の一つとして認識してもらうことが重要であると考える。「ゆっくり解説」は様々な事柄を解説する講座動画として発展してきた経緯もあり、中学校や高等学校の補助教材、実験や演習の予復習教材としての方向性を模索してみたい。

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

[3] 八城年伸、『「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』、情報処理学会 第 162 回コンピュータと教育・第 35 回教育学習支援合同研究発表会論文集、2021、No.29