

動画教材における音声合成の有用性評価に関する基礎的研究

A Basic Study to Evaluate the Usefulness of Synthesized Speech for Tutoring Video

齊藤雄路^{*1}, 松本慎平^{*2}, 寺西大^{*2}

Yuro SAITOH^{*1}, Shimpei MATSUMOTO^{*2}, Masaru TERANISHI^{*2}

^{*1} 広島工業大学大学院工学系研究科

^{*1} Graduate School of Science and Technology, Hiroshima Institute of Technology

Email: md21005@cc.it-hiroshima.ac.jp

^{*2} 広島工業大学情報学部

^{*2} Faculty of Applied Information Science, Hiroshima Institute of Technology

Email: {s.matsumoto.gk, teranishi}@cc.it-hiroshima.ac.jp

あらまし: 本研究では、オンデマンド型講義などでの配信が想定される動画教材における音声合成の有用性の検証を課題に設定し、そのための分析方針や分析結果の一部を示すことを目的とする。音声合成技術は既に実用化レベルに到達しており、音声合成エンジンを用いた教材作成及び学習支援システムの開発など教育分野で様々な形で音声合成技術が活用されている。最近ではリモート講義の普及もあり、音声合成による講義動画を利用したオンデマンド型講義なども実践されつつある。その一方で、音声合成を用いた講義動画の学習効果については十分に議論されていない。よって本研究は、音声合成の話者との差異を明らかにするための基礎的取り組みとして、プレゼンテーション型動画コンテンツにおける音声合成の有用性を評価するための実験方法を設計し、また、その結果を評価する方法や構想を明らかにする。

キーワード: 音声合成, 動画教材, ARCS モデル, POMS2 短縮版, 認知負荷

1. はじめに

人の声を人工的に作り出す音声合成技術が様々な分野で利用されている^(1,2)。音声合成技術の発展とそのソフトウェアの普及により、教育分野でも音声合成は近年に積極的に利用されている。例えば、外国語学習を対象としたシステムが開発され、その有用性が報告されている^(3,4)。また、音声合成による講義資料の読み上げなども一般的に広く行われている⁽⁵⁾。

教育分野では、最近ではリモート講義の普及があり、教材の中で音声合成技術は今まで以上に活用されつつある。音声合成を用いた講義動画を配布するオンデマンド型講義なども普及しつつある。動画教材は教授者の情報提示機能の代替や拡張する役割を果たすことができるため、比較的多くの授業や学習で利用されてきた経緯から、動画ベースのオンデマンド講義はスムーズに受け入れられた。また、その学習に対する有用性もいくつかの研究で示唆されている⁽⁶⁾。一方、このように講義動画での音声合成の利用は一般的になりつつある半面、音声合成の学習効果については十分に議論されてこなかった。今後教育のオンライン化が更に進められる動きがあることから、音声合成の有効性について検証することは意義のある取り組みだと言える。そこで我々は、音声合成は教授者自身による声と同程度の学習効果を有するか、あるいはそうではないのか、といったことの調査を目的に研究を進めている。本研究では、そのための実験方法及び評価方法の設計を目的とする。具体的には、一般的なプレゼンテーション動画を対象として、話者自身の声と、話者の声ではなく音声合成によって生成された声を合わせた動画の場合との2種類の教材を用意し、これらの学習に対する差を分析・評価するための方法を設計し明らかにする。

2. 関連研究

音声合成技術は実用化に対応レベルに十分に到達していると言える。教育において音声合成は従来から外国語教育で積極的に利用されてきた。また、その効果を調査する研究も行われている。例えば康らは、「とても聞きやすい」と評価されるネイティブ音声と同レベルのもの、「普通」と評価される違和感のないもの、ネイティブ音声と明らかな差があるものの3種類の音声合成を明らかにした⁽⁷⁾。教育全体を対象として、音声合成の有用性評価について議論した研究はいくつか行われている⁽⁸⁾。例えば Brouwer は、人工の音声表現が学習に与える影響について、また、これらがモチベーションや注意力に媒介する効果について分析した⁽⁹⁾。その結果、発話速度、ピッチ、感情表現が学習に対して有意に影響を与えている可能性が示唆された。この研究では学習に効果的な音声の合成方法について議論されているが、話者自身の声と合成された音声との差については言及されていない。話者自身とは異なる声の影響について研究した取り組みはいくつか存在する⁽⁹⁾。例えば、外国語映画等の吹き替えで時折知覚される人の顔と音声との間の違和感の原因の分析や、違和感の程度の定量化を目的とする場合が多い。一方で、音声合成による吹き替え自体の有用性やその適切性を評価した研究は行われていない。特に、学習といった観点から、当てられた音声の適切さを評価した取り組みも十分に行われていない。昨今、2D や 3D のアバターや音声合成を活用した学習コンテンツは動画共有サイトなどで急速に拡大しつつあり、またその教材としての採用事例も増えつつある現状を踏まえると、音声合成そのものの学習に対する効果を分析することは、意義のある取り組みだと言える。

3. 研究目的及び分析方法

本研究では、オンデマンド型の講義で一般的に教材として配布されているプレゼンテーション動画によるコンテンツを対象とする。そして、話者自身の声によるプレゼンテーション動画と、話者の声ではなく音声合成が当てられた動画の2種類のコンテンツを用意し、これらの学習効果を明らかにする。本研究では、2種類の教材を用意する。具体的には、コンテンツA、コンテンツBそれぞれ通常版と音声合成版を用意し、通常版をコンテンツA、B、音声合成版をコンテンツA'、B'とする。音声合成には感情やイントネーションなどといった特別な工夫を行わず、タイミングや一時停止の調整のみとした。全ての被験者は、通常コンテンツと音声合成コンテンツ両方で学習を行う。被験者を2つのグループに分ける。グループ1の被験者は、コンテンツA、コンテンツB'の順で学習を行う。グループ2の被験者は、コンテンツA'、コンテンツBの順で学習を行う。学習後、学習内容の理解度を問うテストを行うと共に、認知負荷⁽¹⁰⁾、感情プロフィール^(11,12)、動機付け⁽¹³⁾の3つの観点から学習効果を評価する。

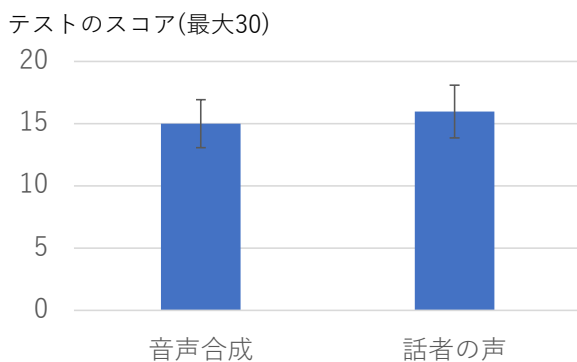


図1 理解度テストの結果

本研究では、動画教材を日本語のTEDトーク¹とした。被験者30人を動員し、15人ずつ2つのグループに分けて実験を行った。コンテンツの長さは20分、理解度テストの制限時間は10分、問題数は15問とした。理解度テストの結果は図1の通りとなった。ここで、エラーバーは標準誤差を表す。対応のあるウェルチのt検定(両側)を行った結果、2群の間に有意な差は見られなかった。このことから、音声合成は教材に不適切ではない可能性が示唆された。一方、音声合成のテストの結果(平均値)はわずかながら話者自身の声の結果よりも下回っていることから、話者の声よりも学習効果が低い可能性もある。この点については音声合成の工夫により改善の余地は十分にあると言え、このことはBrouwerの成果からも明らかである。他、認知負荷、感情プロフィール、動機付けの情報と紐付けて分析することで、属性に応じた効果を明らかにできると考えている。

4. おわりに

本研究では、音声合成は教授者自身による声と同程度の学習効果を有するか、といったことをリサーチクエスションとし、それを明らかにするための実験方法及び評価方法の設計を目的とした。音声合成と話者自身による声との差について、プレゼンテーション動画によるコンテンツを対象とし実験を行った結果、有意な差は見られなかった。このことから、教材において音声合成の有用性が示唆された。今後は、認知負荷、感情プロフィール、動機付けの結果と被験者とを紐付けて詳細な分析を行う。

謝辞

本研究は、独立行政法人日本学術振興会科学研究費助成事業(基盤研究(C)20K02810)の助成を受けて実施した成果の一部である。ここに記して謝意を表します。

参考文献

- (1) S. Lemmetty, History and Development of Speech Synthesis. Helsinki University of Technology (2004).
- (2) 広瀬啓吉 (2000) 21世紀に向けての音声合成の技術展望. 情報処理, Vol.41, No.3, pp.277-281.
- (3) 柏木治美, 康敏, 大月一弘, 外国語学習における音声合成利用の現状と課題. 神戸大学国際コミュニケーションセンター論集, 5, pp.10-19 (2008).
- (4) 鎗木誠, 上橋純子, 浅瀬純子, 加藤雅之, 康敏, 音声合成エンジンを用いた教材作成及び学習支援システムの開発, 日本教育工学雑誌, Vol.27, pp.141-144 (2004)
- (5) L. E. Gonzalez Brouwer, Effect of an artificial model's vocal expressiveness on affective and cognitive learning, Master Thesis of Eindhoven University of Technology (2018)
- (6) B. Means, Y. Toyama, R. Murphy, M. Bakia, K. Jones, Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies, U.S. Department of Education (2009).
- (7) 康敏, 柏木治美, 山名由貴子, 加藤雅之, 鎗木誠, 外国語教育のための音声合成評価, 日本教育工学会論文誌, Vol. 29, pp.161-164 (2006)
- (8) P. L. Witt, L. R. Wheelless, M. Allen, A meta-analytical review of the relationship between teacher immediacy and student learning. Communication Monographs, 71(2), pp.184-207 (2004)
- (9) 伊東裕司, 高山博, 日比谷潤子, 渡辺茂, 顔と声の関連性の判断: 人物の同一性について. 哲学, No. 98, pp. 123-129 (1995).
- (10) J. Sweller, Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257-285 (1988).
- (11) D. McNair, M. Lorr, L. Doppleman, POMS Manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Service (1971)
- (12) J. P. Heuchert, D. M. McNair, Ph.D, 横山和仁, 渡邊一久, Profile of Mood States 2nd Edition, 金子書房 (2015)
- (13) M. Keller, 鈴木克明, 学習意欲をデザインするARCSモデルによるインストラクショナルデザイン, 北大路書房 (2010)

¹ <https://www.ted.com/>