

音声を再生時に合成する携帯端末向け講義ビデオの提案

斉藤 秀[†], 黒田 晃弘[†], 古川 文人^{†‡}, 佐々木 茂[†]

帝京大学理工学部[†] 帝京大学ラーニングテクノロジー開発室[‡]

1. はじめに

ネットワーク上にある講義ビデオは、有効な学習コンテンツの一つである。一方で、携帯端末向けのネットワークにおいては、講義ビデオの視聴によるネットワークへの負荷が危惧される。

そこで、本研究では、スライドと口頭説明のための音声から構成される講義ビデオを対象として、音声合成技術を用いることで携帯端末へ配信する際の講義ビデオのデータサイズを削減する手法を提案する。

2. 視聴時音声合成型講義ビデオの提案

講義ビデオの収録方法は、一般に、以下の3種類に大別できる。1つ目は、板書、スクリーン、教授者などの映像と教授者による口頭説明の音声を含め講義の様子をそのまま収録したもの(図1左側)。2つ目は、MicrosoftのPowerPointなどによるスライドとその提示のタイミングと同期した口頭説明の音声を収録したもの(図1右側)。3つ目は、1つ目と2つ目を組み合わせたものであり、例えば、1つ目の教授者が口頭説明をする映像、スライド、音声を収録するものである。

本研究では、2つ目の方法により収録された講義ビデオを対象とする。この種の講義ビデオは、一般にはスライドの情報を保持する画像データと口頭説明の情報を保持する音声データ、およびそれらを同期して提示するための情報から構成される。視聴時には、これらのデータを配信サーバからダウンロードする(図2上部)。

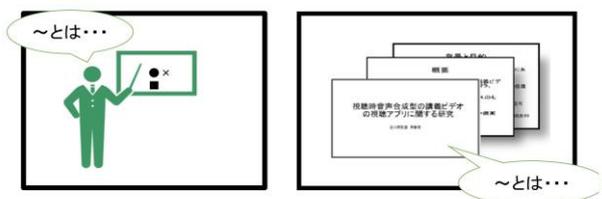


図1 講義ビデオの種類

これに対して本研究では、口頭説明の情報を文字情報(テキストデータ)として配信サーバに保持しておき、視聴時に端末上でテキストデータを基に音声合成をすることを提案する(図2下部)。これにより、講義ビデオのデータサイズを削減し、ネットワークへの負荷を軽減することが期待できる。

以降では、図2上部のように口頭説明の情報を音声データとして保持するビデオを従来方式、図2下部の本研究で提案する視聴時音声合成型講義ビデオを提案方式とよぶ。

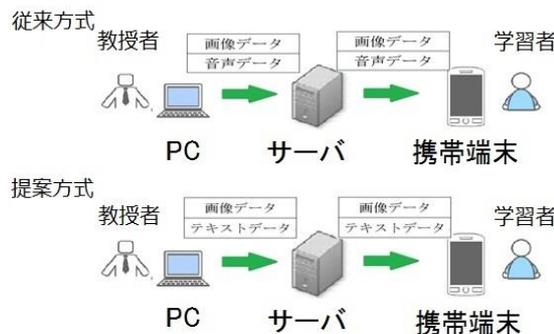


図2 従来方式と提案方式の比較

3. 提案方式によるデータサイズ削減効果の確認

提案方式により、講義ビデオのデータサイズを削減できることを実験的に評価した。評価に用いた教材は、帝京大学理工学部情報電子工学科1年次の「プログラミング2」において、反転授業のために作成した約13分間の講義ビデオである。

本方式がサイズ削減の対象とするのは、音声データである。音声データのサイズは、ビットレートにより大きく変動する。実験に用いた講義ビデオの音声データのビットレートを8kbps、16kbps、56kbpsの3種類のを聴き比べたところ、8kbpsでは音質の劣化がみられ、16kbps、56kbpsでは音質に特段の差はなかった。そこで、本実験では、ビットレート8kbpsと16kbpsの音声データを用いることにした。なお、スライドの画像は、解像度960×720のJPEG形式のものである。

Proposal of lecture video system for mobile terminals
speech synthesis when watching

Suguru Saitou, Akihiro Kuroda, Fumihito Furukawa,
Shigeru Sasaki, Teikyo University

評価の結果、本方式により、8kbps、16kbpsの音声からなる講義ビデオの総サイズの、それぞれ14%、24%を削減できることがわかった(図3,4)。

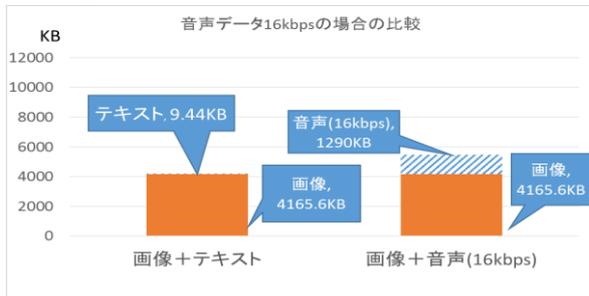


図3 音声16kbpsの場合の比較

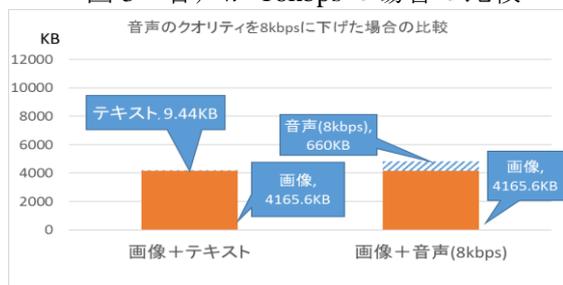


図4 音声8kbpsの場合の比較

4. 提案方式のためのシステムの構築例

提案方式のビデオ作成から視聴までの全体像を以下の図5に示す。教授者がMicrosoft Power Point等のプレゼンソフトで作成したスライドとノート部分に記述されたテキストを変換し、複数のスライド画像とテキストデータで構成されたビデオコンテンツデータを作成する。作成したものは、サーバにアップロードされ、学習者が端末にダウンロードすることで視聴することができる。本研究では、Android OSを対象として提案方式の視聴環境をAndroid Studio[1]を用いて作成した。

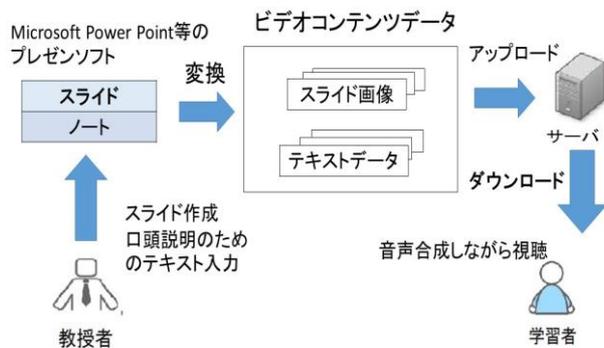


図5 ビデオ作成から視聴までの全体像

5. 考察

提案方式では、音声合成の導入による学習への影響が考えられる。音声合成では、抑揚をつけることが教授者の発声行為に比べて難しいため学習者は、例えば、学習のポイントを捉え難い状況が考えられる。また、言語教育などにおける発音が学習の対象になる場合は、音声合成の質に注意する必要がある。

一方で、音声合成を導入することにより講義ビデオの作成コストを軽減できる場合がある。教授者の言い間違いや雑音などによる再収録および編集作業を回避できるためである。

提案方式では、視聴時に音声合成技術を導入しているのに対し、ビデオ作成時に同技術を導入している類似のシステムが既にある[2, 3]。ビデオ作成時にMicrosoft Power Pointのノート部分のテキストからスライドと同期する音声を合成するものである。以下では、これらの類似システムとの比較をする。

提案方式の長所は、ネットワークへの負荷が軽い点である。類似システムでは、口頭説明の情報を配信サーバから音声データとして端末にダウンロードするのに対し、提案方式ではテキストデータであるためである。

短所は、端末に視聴環境を整備する必要がある点である。類似システムでは、MPEG-4やWindows Media Videoなどの標準的な形式で配信できる。

6. おわりに

本稿では、コンテンツのデータサイズを削減するために視聴時音声合成型の講義ビデオを提案した。また、大学の教育現場で実際に用いられているビデオを用いて、提案方式によりデータサイズを削減できることを確認した。

参考文献

[1]Android Studio : Android デベロッパー | Android Developers, <https://developer.android.com>, 2016/11/02.
 [2]Real Narrators2 : パワーポイント活用ー合成音声付動画作成ソフトウェアリアルナレーターズ, <http://www.skyfish.co.jp/rn2/index.html>, 2016/12/01.
 [3]STORM Maker : 音声合成付きeラーニング・e-learning教材作成ソフト【STORM】, <https://suite.logosware.com/storm-maker/>, 2016/12/01.