# 音声解析とキーワードに基づく動画教材自動生成システム

# Automatic Video Tutorial Generation System Based on Audio Analysis and Keywords

中村　拓樹[*1]，小島　篤博[*2]
Hiroki NAKAMURA[*1], Atsuhiro KOJIMA[*2]
[*1] 大阪公立大学現代システム科学域
[*1]College of Sustainable System Sciences, Osaka Metropolitan University
[*2] 大阪公立大学大学院情報学研究科
[*2] Graduate School of Informatics, Osaka Metropolitan University
Email: ss22525j@st.omu.ac.jp

**あらまし**：3D アバターを用いた授業動画制作におけるモーション作成の自動化を目的とし，講義スライドと音声から動画を生成するシステムを構築した．本システムは，入力音声に対して AI によるタイミング解析を行い，事前に用意したアバター動作を，設定したキーワードの発話タイミングに合わせて自動で割り当てる．これにより専門的な 3D 操作を不要とし，スライドの内容に沿った必要最低限なアバター動作の自動化を実現した．
**キーワード**：授業動画，システム開発

## 1.　はじめに

昨今，オンライン授業の普及により，動画教材の需要が拡大している．しかしながら，実写映像やスライドと音声によって説明される映像を用いた動画教材では，教材作成にかかる時間や画面の単調さが問題となっている (1)．そこで，これらの問題点を補う動画教材の形式として，スライドと音声に加え，3DCG キャラクタを用いた動画教材に関する研究が行われている (2) (3)．

しかしながら，この形式の動画教材を作成するためのシステムでは，特にモーションを作成するという部分で手間がかかるうえに，3DCG の専門知識が必要となる事例もある．そこで本研究では，入力音声に対して AI によるタイミング解析を行い，事前に用意したアバター動作を設定したキーワードの発話タイミングに合わせて自動で割り当てる．これによりスライドの内容に沿った必要最低限なアバター動作のある動画教材の作成を自動化し，教材作成時間の短縮化を目指す．

## 2.　システムの概要

### 2.1　システムの設計

動画生成の自動化には，複数の異なる種類の処理を統合する必要がある．具体的には，動画生成のために必要なスライドの画像化と 3DCG レンダリング，モーションを割り当てるための AI による音声解析，音声解析用の環境を整える音声合成である．

これらの処理すべてを単一のプログラムで実装することは困難であるため，本システムでは Python プログラムを仲介して既存の外部アプリケーションや API を制御する設計とした．

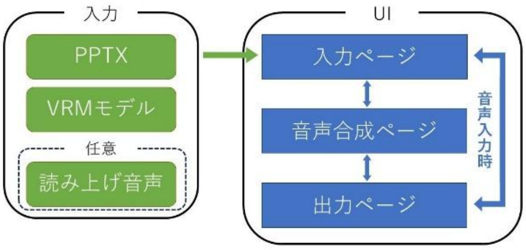


図 1　システム構成

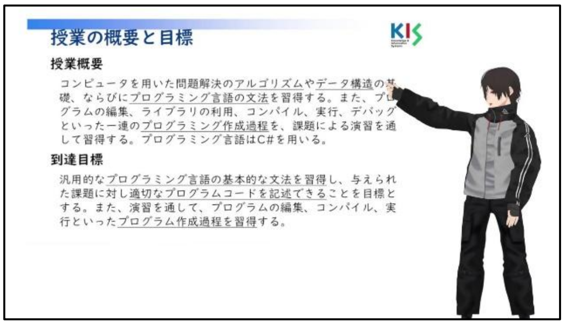


図 2　動画のスクリーンショット

### 2.2　システムの構成

本システムの構成を図 1 に示す．入力として必須のファイルは，読み上げるテキストがノートに記載してある PPTX ファイルと VRM モデルである．テキストを読み上げた音声ファイルの入力は任意で，1 枚のスライドにつき 1 つの音声を用意する必要がある．

システムは 3 つのページを遷移する形で実行され，音声が入力されているか否かで遷移の仕方が変わる．音声合成ページは音声が入力されていない場合に遷移し，音声合成を行う．出力ページではまずスライドの画像化を行い，音声解析や音声合成ソフトから

の情報を基にモーション割り当てのタイミングを決定する．その後 3DCG レンダリングを行い，図 2 に示すような動画を出力する．

## 3.　システムの実装

### 3.1 開発環境

本システムの開発には，統合開発環境として Visual Studio Code を採用している．また，開発言語として Python を, 3DCG のレンダリングには Blender を，音声合成には VOICEVOX を，スライドの画像化に Microsoft PowerPoint を，音声解析には Whisper を，GUI の実装には PySide6 を，それぞれ使用している．

### 3.2 外部連携機能の実装

外部プロセス制御を行うために，音声合成処理を行う VoiceVoxWorker クラス，スライドの画像化を行う PptxToPngWorker クラス，音声解析を行う WhisperWorker クラス，3DCG レンダリングを行う BlenderRenderWorker クラスの 4 つのクラスを独自に実装した．

### 3.3 モーション生成ロジックの実装

テキスト解析および音声合成情報に応じて，アバターの適切な動作とそのタイミングを決定するロジックを実装するクラスとして，MotionPlanner クラスを実装した．このクラスの処理の流れを図 3 に示す．まず音声解析によってデータ化した発話文の各単語がいつ発せられるかという情報、または音声合成ソフトから取得した発話文の各文字（モーラ）をいつ発話させたかという情報を，タイミングマップという形式に変換する．その後特定のキーワードが検出されたタイミングをタイミングマップから読み取り，モーションスクリプトを生成する．

### 3.4 Blender 制御スクリプトの実装

Blender を制御するスクリプトはシステムから JSON 形式の設定ファイルを介して起動される．背景の生成，VRM モデルのインポート，スライド画像の平面オブジェクトへの貼り付け，およびモーションスクリプトに基づいた非線形アニメーショントラックの構築を自動で実行し，Blender への指示データを作成する．この指示データを基に，Blender でモーション付きの動画を生成する．

### 3.5 そのほかの実装

GUI で複数のページを用いるため，それぞれのページでデータを共有する必要があることから，そのことに特化した ProjectState クラスを実装した．

## 4.　評価・検討

実際に動画を生成した結果，おおよそ正確なタイミングでモーションが割り当てられていることを確認した．また，本システムによって，専門知識なしに 3DCG キャラクタを用いた動画教材の作成ができ，動画作成に要する時間の短縮が可能となった．

一方，モーションの数が少ないことから，待機モーションの時間が長くなってしまうケースや，同じ動きを繰り返し，単調となってしまうケースがあった．そのため，扱うモーションを多様にし，これらの問題の解決を図ることが課題として挙げられる．ただし，不必要なモーションはノイズとなるため，必要なモーションの選定を行う必要がある．

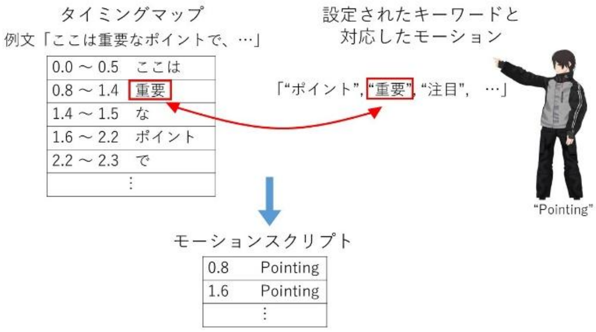


図 3 MotionPlanner クラスの流れ

また，事前に設定したキーワードのみでは，講義内容固有のキーワードに反応させることができず，動画制作者がその都度入力する必要がある．この問題の解決策として，AI によるキーワード自動選定機能の実装も課題として挙げられる．

## 5.　まとめ

本研究では，モーションの割り当てを自動で行うことで，3DCG キャラクタを用いた動画教材の作成時間が短縮した．

今後の課題として，必要なモーションの選定や多様化，AI によってモーション割り当てのキーワードを自動選定することなどが挙げられる．

### 参考文献

(1) デジタルナレッジ, "ビデオ教材 (映像コンテンツ) の教育利用に関する定点調査報告書," https://www.digital-knowledge.co.jp/archives/1702/ , 2014
(2) 天野由貴，隅谷孝洋，岩沢和男，西村浩二，"情報セキュリティ教育の動画教材における実写映像とアバター動画の比較," 大学 ICT 推進協議会 2015 年度年次大会，pp．1-5，2015.
(3) 高山伸也，酒澤茂之，愛澤伯友，"3D キャラクタを用いた教育コンテンツの有効性検証，"映像情報メディア学会冬季大会講演予稿集，2013 巻, pp. 2-5, 2013.