

システム創造プロジェクト 最終レポート

2019 年度 第 5 班

2020 年 2 月 18 日

1 基本情報

1.1 プロジェクト名

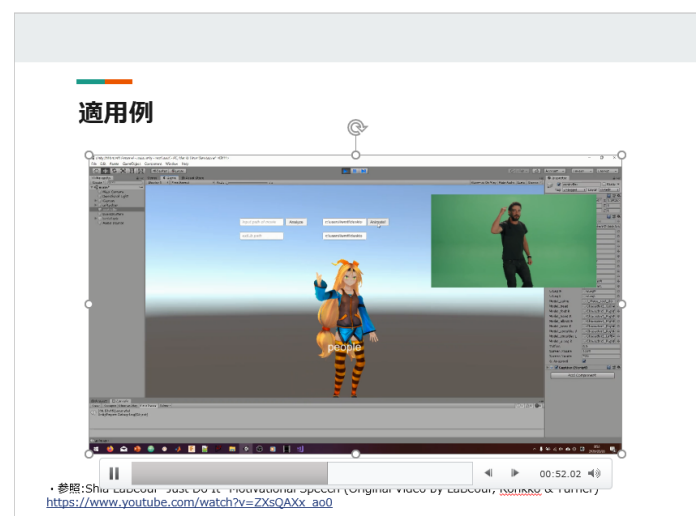
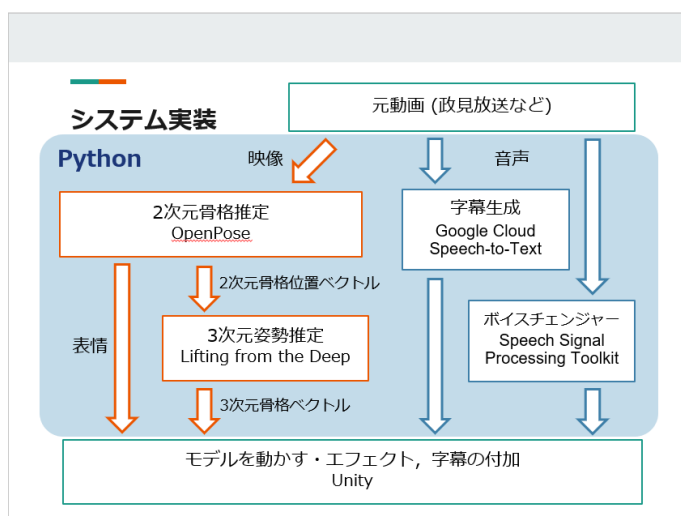
動画変換システム ～骨格推定等を用いた新時代の政治エンタメの提案～

1.2 メンバー

19M11893 岩本 拓也
19M12220 藤本 八雲
19M18323 倉林 快生
19M12065 田屋 裕輝
19M12160 林 良優
19M11982 岸 良樹
19M12295 森本 義弥

1.3 概要

現代の日本が抱える問題の一つに、選挙における若者の投票率の低さが挙げられる。本プロジェクトでは、政治への関心を高めることを目的として、骨格推定等を用いて政見放送等の動画を VTuber の動画に変換するシステムを作成した。



2 製作したソフトウェア

2.1 ソフトウェアの使用方法

2.1.1 環境のセットアップ

本製品の使用には以下のハードウェア環境が必要となる。

- Windows10(64-bit) machine
- CUDA 対応 GPU

加えて以下のソフトウェア環境が必要となる。

- Python3.X (Anaconda3 の最新バージョンに対応したものを推奨)
- Anaconda3
- Unity editor
- CUDA toolkit v9.0
- CUDNN 7.0.5
- Openpose v1.4.0
- Lifting from the Deep

同梱のバッチファイルを利用することで環境構成負担は大幅に軽減されている。紙面の都合上、実行環境のセットアップに関する詳細は <https://github.com/syspro5/iwamoto> を参照のこと。

2.1.2 実行方法

1. unity エディタを起動し、本製品のプロジェクトファイルを開く。
2. 解析したい動画のパスを指定する。
3. Analyze ボタンを押すと解析が開始され、結果ファイルが出力される
4. 解析終了後 Animate ボタンを押すことで字幕とモデルの動きの動画が出力される。

実際の実行の様子は同梱されたビデオファイルを参照のこと。



図1 ミーティング内容の比較

2.2 ソフトウェアの機能・仕様

- 深度情報なしに骨格推定を行い、モデルに動きを反映させる一連の処理を画一化。
- 音声解析により字幕描画も実装。
- 推定可能人数は実装時点では1人

- 現時点では Unity editor 上での実行となる

2.3 ソフトウェアの設計・製作過程

2.3.1 設計

処理のフローチャートを以下の図 2 に示す。動画から音声などを抽出し、要素ごとに分けて処理を行っている。通常 Unity から直接 python の仮想環境へのアクセスはできないため、今回の実装ではシェルスクリプトを介して anaconda の仮想環境にアクセスする手法をとっている。仮想環境をセットアップするスクリプトも同梱しているため、利用者が一々必要なライブラリをインストールする必要はない。また、専用の仮想環境を作成することで、利用者の python 環境を改変することなしに複数のライブラリの利用が可能となることも利点となる。

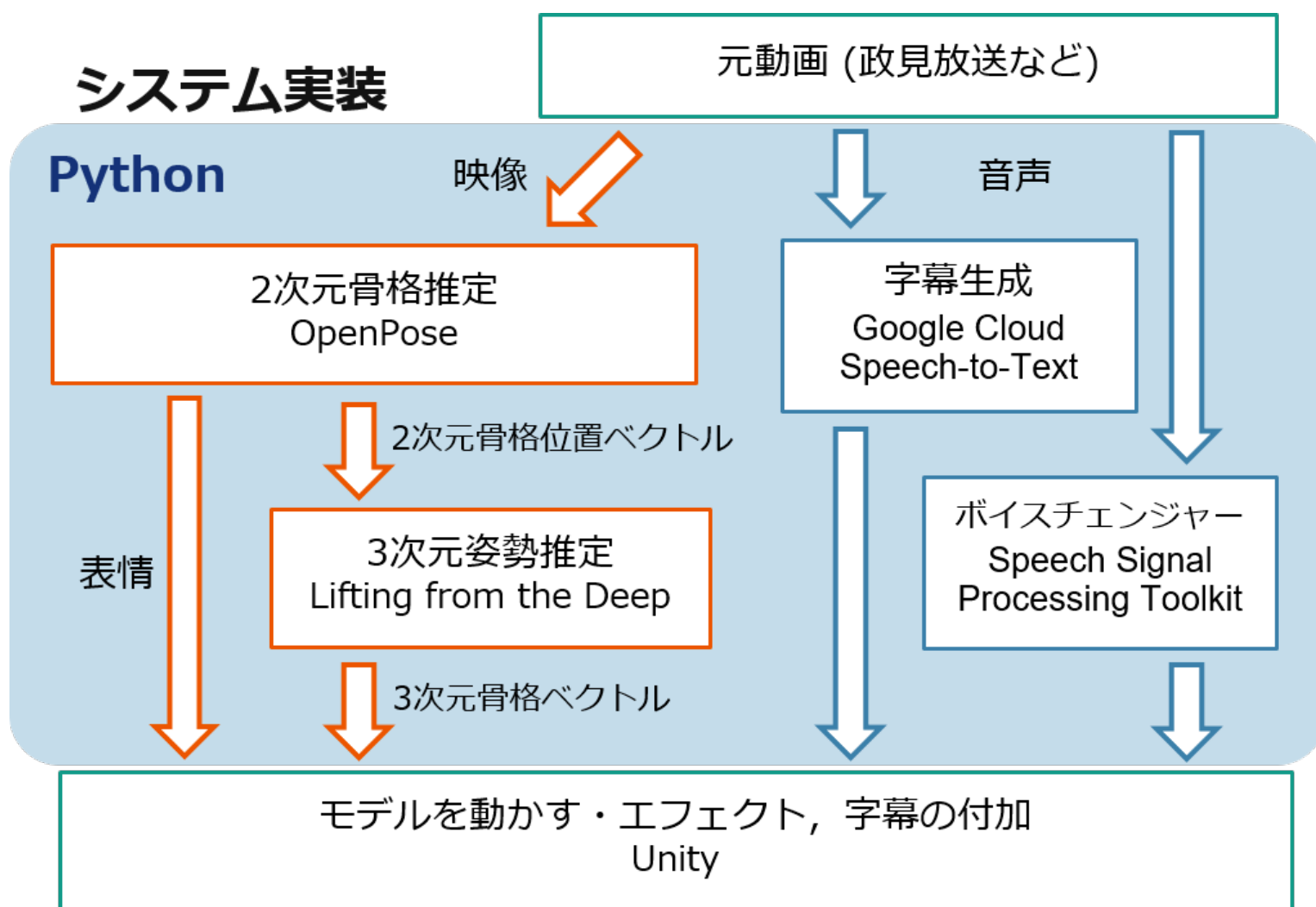


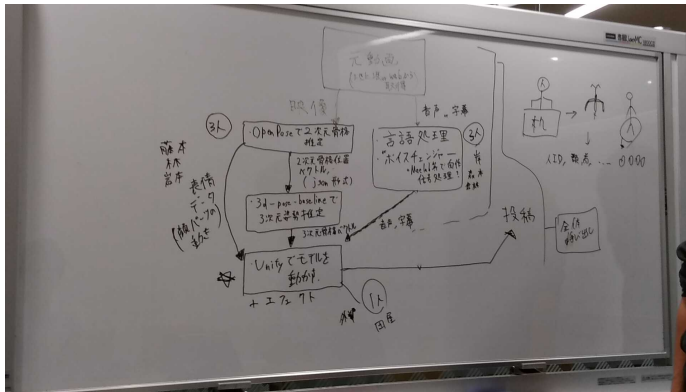
図 2 処理の概要図

2.3.2 製作過程

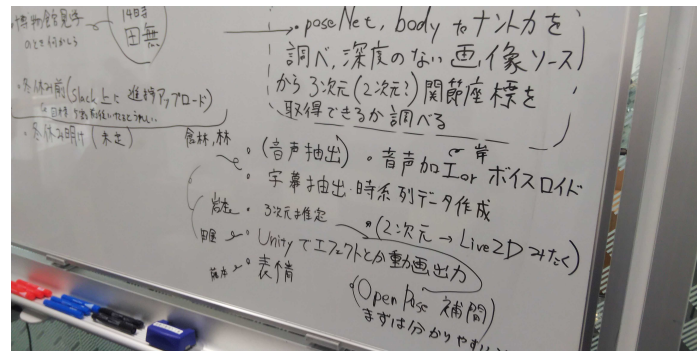
10 月時点と 12 月時点でのミーティング事項の比較を図 3 に示す。10 月時点では定まっていなかった「具体的に何を使用するか」が、12 月時点では明確に定まった。

3 製作したハードウェア

今プロジェクトでは、ハードウェアは作成していない。



10 月時点



12 月時点

図 3 ミーティング内容の比較

4 各種調査，分析レポート等

4.1 投票率低下の原因調査及び Vtuber 利用の妥当性の検討

図 4 に示すように現代の日本では選挙率の低下が問題となっており，特に若者の投票率が低い．図 5 に示された投票に行かなかった理由から，投票率の低下の原因の大きな一つとして政治への関心が低いことが考えられる．

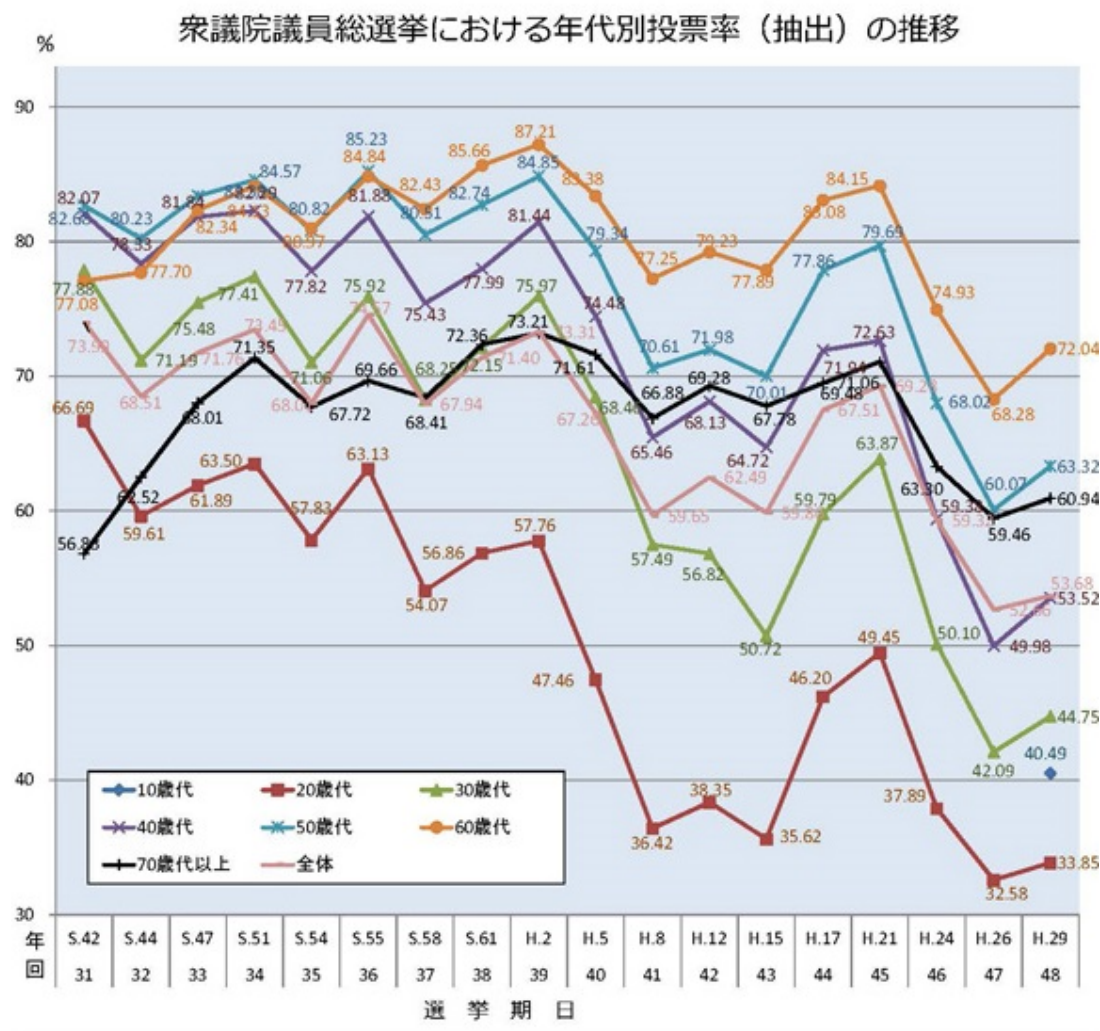


図 4 投票率の推移（出典 総務省 [1]）

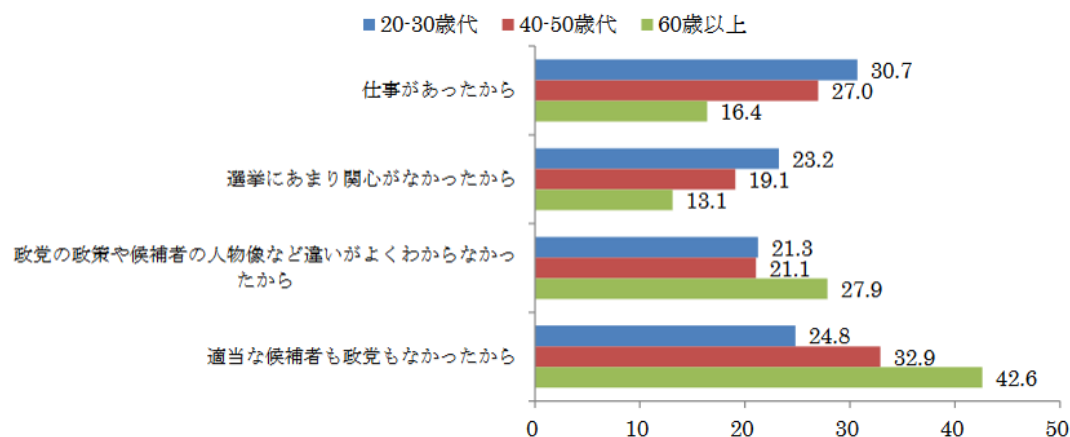


図5 投票に行かなかった理由（出典 財団法人明るい選挙推進協会 [2]）

若者の投票率を上げるためには、若者に認知度があり興味のある Vtuber と選挙を紐づけることが効果的であると考えた。図6に示すように Vtuber は若者の間で圧倒的な認知度を得ている。このことから Vtuber を利用することで、若者の選挙への関心度の増加が見込める。

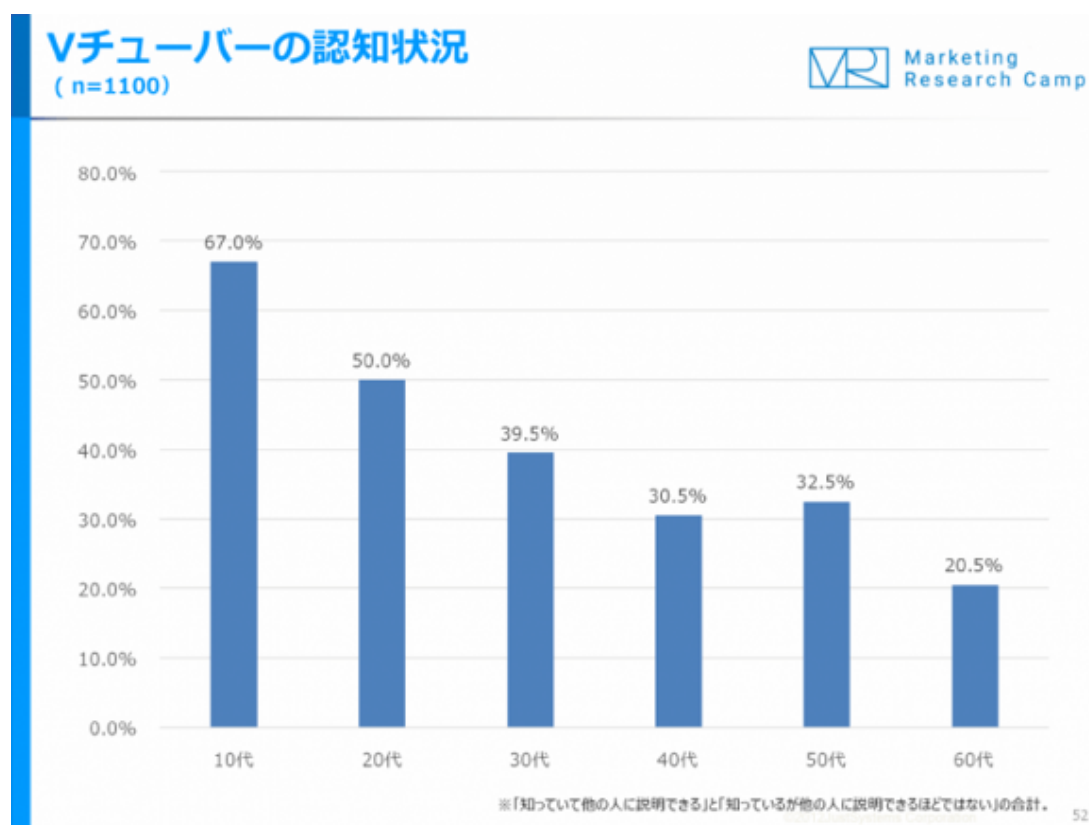


図6 VTuber の認知状況（出典 MoguLive[3]）

4.2 政見放送に関する権利等の調査

本システムでは政見放送の動画編集を行うため、政見放送の著作権及び公職選挙法を考慮する必要がある。

政見放送は著作権法第40条の定める「公開して行なわれた政治上の演説又は陳述」にあたり、「いずれの方法によるかを問わず、利用することができる」と解される。つまり、二次的利用（27条）も許され、同一性保持権の侵害にならない限り翻訳や要約も許される。また、他に複製物を譲渡すること（26条の2）ができる [4]。選挙期間中に政見放送が YouTube 等の動画共有サイトにアップロードされることが問題になることがあるが、これは著作権法上ではなく公職選挙法上の問題である。例えば、特定の候補者の政見放送のみをアップロードすることにより公平性が

担保できなくなる，などの理由が挙げられる [5]．

つまり，内容の同一性が確保された範囲での編集は可能であり，公平性が担保されていれば政見放送を利用することは可能であると考えられる．

4.3 Google Cloud Speech-to-Text の使用料金

本システムで使用した Google Cloud Speech-to-Text は 60 分まで無料で使用できるが，超過すると課金が必要でありその内訳を表 1 に示す [6]．ここで示す標準モデルは音声を対象としたモデルであり，プレミアムモデルは動画，拡張音声電話にも対象を拡張したモデルである．本システムでは一番安価なコースを用いた．約五分の政見放送を対象とすると一回の使用料金は約 8 円となる．なお，開発の際は無料期間を利用したため使用料金は発生しなかった．

機能	標準モデル	プレミアムモデル
音声認識（データロギングなし）	\$0.006/15[s]	\$0.009/15[s]
音声認識（データロギングあり）	\$0.004/15[s]	\$0.006/15[s]

表 1 Google Cloud Speech-to-Text 使用料金

5 発表資料

5.1 発表用スライド

添付ファイルを参照のこと．

5.2 課題設定発表での質疑応答一覧

質問	回答
骨格推定はできるか	Open Pose を用いることで可能
全員写すか, 1 人にフォーカスするか	1 人にフォーカス
各党でモデルを変えてはどうか	公平性担保のためあまり変えない
モデルは美少女のみか	はい
可愛さのみで投票する恐れはないか	各党でモデルに差異を付けない
VTuber 視聴割合のデータはあるか	認知度のデータはある
美少女でも討論は面白くないのでは	エフェクト等を工夫したい
音声認識は難しいのでは	字幕情報を使用する
既存の政治解説 VTuber との差別化は	リアルタイムの用語解説を付加
国会中継は動きがない	動きのあるコンテンツを探す
献血ポスターのように炎上する恐れ	政治的中立性に則った上で関心を高めることが目的

5.3 中間発表での質疑応答一覧

質問	回答
国会中継から政見放送に変えたのか	本システムを実装しやすいソースを選択した
Open Pose の代替案 (Pose Net 等)	検討する
政見放送の権利問題, 著作権	システムを政権に提供する形であれば問題なし
他のものをソースにしてはどうか	エンタメ要素として活用できるため, 十分検討する

5.4 最終発表での質疑応答一覧

質問	回答
作成にあたり大変だった点はどこか	開発言語が様々である各パートを組み合わせる点

6 自己評価書

我々の班で制作した, 「人物の動画から, 人物の姿と声を 3D キャラクターに置き換えてエンターテインメント性を高めた動画を出力するソフトウェア」についての評価を行う. まず評価の前にソフトウェアの目的について整理する. ターゲットは政治に関心がない若者で youtube などのインターネット文化に馴染んでいる者とした. こういった若者が我々のソフトウェアにより生成されたキャラクターアニメーションを見て, 愛着の湧くようなかわいらしさとキャラクター性による馴染みやすさ, 内容と見た目のミスマッチやアニメ的表現・表情により面白さを感じ, そこから少しずつ政治の用語や情勢が頭に入るようになって, 自分から積極的に調べるようになり, 若者が政治にどんどん関わるようになる, というようなストーリーを実現できるようなソフトウェアの開発が目的である.

この目的において制作したソフトウェアの評価を行う. 現在制作したソフトウェアの持つ機能は動画から 3 次元の空間上のキャラクターの動きと字幕を生成するところまでであり, ボイスチェンジャーについても統合が可能な段階まで完成している. 制作物の具体的な動作については同じく提出した動画及び仕様書を参照. 動画からわかるように, 動画内に写っている体のパーツについては上手くキャラクターに反映できており, 字幕もかなりの精度で再現できていることがわかる. また, 動き自体も滑らかに再現できており, キャラクターとしての馴染みやすさは表現できていて, 内容とのミスマッチ感の面白さも感じられる. 一方で, 表情・エフェクト (集中線や字幕の大きさ・色・フォントの変更) が現在つけられていないのでアニメ的な面白さについては十分に表現できていない. また, 音声についてはそのまま用いることはできないので, どのような音声をつけると面白く馴染みやすいのか検証する必要がある.

参考文献

- [1] 総務省—参議院議員通常選挙における年代別投票率の推移
https://www.soumu.go.jp/senkyo/senkyo_s/news/sonota/nendaibetu/
- [2] 財団法人明るい選挙推進協会—第 46 回衆議院議員総選挙全国意識調査 p39
<http://www.akaruisenkyo.or.jp/wp/wp-content/uploads/2013/06/070seihon1.pdf>
- [3] MoguLive—VTuber を 10 代の約 7 割が認知 ジャストシステムがリサーチを発表
<https://www.moguravr.com/just-systems-vtuber-research/>
- [4] 栗田隆: 著作権法注釈
<http://civilpro.law.kansai-u.ac.jp/kurita/copyright/commentary/Act40.html>

[5] Wikipedia—政見放送

<https://ja.wikipedia.org/wiki/政見放送>

[6] Google Cloud—Cloud Speech-to-Text ドキュメント-料金

<https://cloud.google.com/speech-to-text/pricing?hl=ja>