

音声認識を用いた自己評価型面接練習アプリの開発

Development of a Self-Assessment Interview Practice Application using Voice Recognition

人間情報システム工学科 福住 和真

指導教員: 大隈 千春 教授

Abstract: The global pandemic caused by COVID-19 has changed our lifestyle and communication. Various IT communication tools for work-from-home and online-class have been developed. Nowadays, there are also changes in job interviews and many companies introduce online job interviews. This paper describes the development of a self-assessment interview practice application to support practice for online interviews, which are different from traditional face-to-face interviews.

1. 序論

2019 年に発生した新型コロナウイルスは世界中に多大な影響を与えた一方で、IT 関連の技術開発や人々の IT コミュニケーションツールには革新的な影響も与えている。その一つとして就職活動においての変化が見られる。従来の就職面接の形態は対面が基本であったが、近年ではオンラインの面接を取り入れている企業の増加が見られるようになった。求職者が面接に向けての準備を行うために面接練習用のアプリやサイトも存在している。

しかし、従来の面接練習アプリではカメラ目線やマイクとの距離といったオンラインによる面接での考慮すべき点への対応が充分ではない。そこで本稿では従来の面接練習アプリに比べ、オンライン面接に主眼をおいた面接練習アプリの開発について述べる。

2. 面接練習アプリの現状

序論で述べた通り、近年ではオンライン面接を取り入れている企業が増加している。HR 総研が行った 2021 年卒及び 2022 年卒採用活動動向調査^[1]の結果では、対面でのみ実施した企業が 27%、オンラインのみを実施した企業が 19%、オンラインと対面のハイブリッドで実施した企業が 34% などとなっており、オンライン面接を実施した企業が 53% という結果が示されている。

オンライン面接は対面の面接とは異なる点に配慮する必要がある。例えば、画面に映る相手の顔ではなくカメラに視線を向ける、カメラの角度、マイクと自分の位置などが挙げられる。

現存している面接練習アプリは、従来の対面での面接を前提に作られている部分が多く存在する。その形態の多くは、質問が画面に表示されると回答時

間を与えられ、与えられた時間内に利用者は回答するというものである。この形態の利点として、実際の面接の流れに沿って練習を行うことが可能な点や、時間内に答える練習が可能な点が挙げられる。

しかし、この形態では自分の回答内容や回答している様子が記録として残ることがない。そのため自分がどのような発言を行っていたかの確認や、回答時の自分の様子を見返すことが出来ないという問題点がある。また、現存する面接練習アプリでは先ほど挙げたオンライン面接において配慮しなければならない点についての練習を行うことが難しい。

3. 面接練習アプリの開発

3.1 目的と機能分析

本研究では、前章で述べた現存する面接練習アプリの問題点を改善する Web アプリの開発を目的とする。必要な機能として、動画から音声を抽出し、その音声から音声認識を用いて文字起しを行う機能、評価項目に従って自己評価を行う機能、動画と文字起しされたテキスト、自己評価の内容を保存する機能、保存したデータを見返すことが出来る機能が考えられる。またこの機能を実装するためにユーザ管理を行うためのアカウント作成機能とログイン機能が必要である。以上の機能を実装した自己評価型の面接練習アプリの開発を目的とする。

3.2 システムの動作説明

本システムの大まかな動作の流れは以下のようになる。最初に利用者のアカウントを作成し、システムにログインする。次に利用者は練習を行いたい質問を選択し、質問の回答を自身のカメラアプリで撮影する。撮影した動画をサーバーにアップロードし、サーバー側で文字起しの処理を行い、文字起しの結果と受け取った動画を表示する。その後用意した評

価項目に応じて動画と文字起しのテキストを見ながら利用者は自己評価を行う。自己評価を行った後、回答している動画、文字起ししたテキスト、自己評価を保存する。

回答を見返したい場合は保存された動画、保存されたテキスト、自己評価を表示するページを表示する。

3.3 フレームワークの選定

Web アプリの開発を行う際は一般的にフレームワークが用いられる。フレームワークとは、ソフトウェアを開発するうえでの典型的・定型的な処理やデータ構造などをまとめた、枠組みやテンプレートとなるものである。開発には Python を用いたため、Python 向けのフレームワークを使用する。Python 向けのフレームワークとして代表的なものは Django^[2] と Flask^[3] の二つがある。

Django の特徴は Web アプリを開発する上で必要な機能がとても充実している点である。例えばテンプレートでユーザ認証機能を実装できる点や、管理者画面が最初から備わっている点、データベースとの接続のテンプレートが用意されている点などが挙げられる。また、ユーザ認証においてアカウントとパスワードを安全に管理する方法が提供されており、セキュリティ面でも信頼度が高い。しかしながら、機能が充実している分、導入や理解するのに時間を要する。

対して Flask は Django よりも機能が少ないが、自由度の高い開発を行うことが出来ることが特徴である。機能が少ない分理解するまでの時間はあまりかからないが、一から実装しなければいけないことが増えてしまう。

以上のことより、機能が充実おりデータベースも扱いやすい Django を用いて開発を行う。

3.4 面接質問内容と評価項目の選定

面接練習アプリを開発するにあたり質問項目と自己評価項目を検討した。質問項目はマイナビ転職のよくある質問の回答例^[4]と自分や友人の実際の経験をもとに、志望動機、学生時代頑張ったこと、企業に入ってやりたいこと、長所と短所、最も苦労したこと、自己 PR とした。

次に評価項目では、面接質問集^[5]の見極めのポイントと自分や友人の実際の経験を参考、にカメラに対する視線、上目使いや下目使いになっていないか、発言中の視線が不必要に動いていないか、表情、回答

のテンポ、声量と声色、姿勢、質問に対する回答の妥当性という評価項目を設定した。

3.5 開発環境

アプリの開発環境を表 1 に示す。

表 1 開発環境

開発環境 (OS)	Windows 10 home
開発言語	Python , html css , javaScript
データベース	MySQL
フレームワーク	Django
ツール	ffmpeg

本アプリは Python を用いて音声認識、文字起しを行うプログラムを実装する。Python を用いた主な理由としては、音声認識のライブラリが用意されている点とスクリプト中に外部ツールを導入することが比較的容易な点である。

動画から音声を抽出する際には ffmpeg というツールを用いた。このツールを用いることにより、mp4 ファイルから音声だけを抽出し wav ファイルに変換することが可能である。

3.6 データベースの内容

データベースには MySQL を用いた。作成したデータベースのテーブルを表 2 に示す。またそれぞれのテーブルのカラムを表 3 ～ 表 5 に示す。

表 2 作成したテーブルと説明

テーブル名	説明
account	ユーザー情報の保存先
question	質問の保存先
data	動画、テキスト 評価内容の保存先

表 3 account のフィールドと属性

フィールド名	属性	説明
id	bigint	ユーザ番号
last_name	varchar(100)	苗字
first_name	varchar(100)	名前

表 4 question のフィールドと属性

フィールド名	属性	説明
q_id	int	質問番号
question	varchar(100)	質問文

表5 data のフィールドと属性

フィールド名	属性	説明
id	bigint	回答 id
video	varchar(255)	動画の保存場所の URL
text	varchar(255)	文字起しされたテキストの URL
hyouka_1~8	int	評価項目 1~8 の評価
hyouka_total	double	総合評価
user_id	int	回答したユーザの ID
q_id	int	回答した質問の ID

3.7 画面設計

本アプリは、アカウント登録画面、ログイン画面、ホーム画面、質問に回答している動画のファイル送信と文字起しを行う画面、回答検索画面で構成されている。画面遷移図を図 1 に示す。

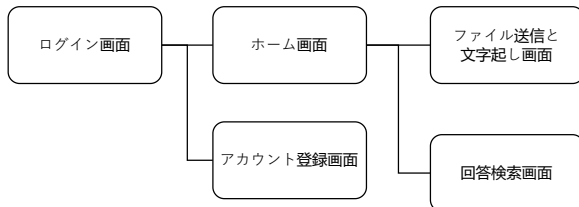


図 1 画面遷移図

3.7.1 ログイン画面

アプリのサイトにアクセスすると図 2 のようなログイン画面が表示される。ここでアカウントを持っていない新規ユーザの場合はアカウント作成を押すことでアカウント作成画面に進むことが出来る。既存ユーザは、登録しているユーザ ID とパスワードを入力し、認証が通れば図 1 のようにホーム画面に遷移する。

図 2 ログイン画面

3.7.2 アカウント登録画面

ログイン画面で、アカウント作成のボタンを押すと図 3 の画面に遷移する。ここで新規ユーザは、ユーザ ID、パスワード、メールアドレス、名前の情報を入力する。入力されたユーザ ID、メールアドレスは Django に用意されている user というテンプレートに登録される。パスワードはハッシュで暗号化された後、ユーザ ID、メールアドレスと同じ user テンプレートに登録される。入力された名前の情報は表 3 に示した account フィールドの last_name と first_name にそれぞれ登録される。登録が完了するとログイン画面に誘導する画面が表示されるので、入力したユーザ ID とパスワードを用いてログインを行うことが可能である。

図 3 アカウント作成画面

3.7.3 ホーム画面

ログインしたユーザーは図 4 のホーム画面に遷移する。この画面には現在ログインしているユーザーの質問の回答数が表示されている。また、サイドバーを押すことでファイル送信と文字起しの画面、回答検索画面に遷移することが可能である。

図 4 ホーム画面

3.7.4 動画ファイル送信と文字起し画面

図 5 に動画ファイル送信画面を示す。ユーザーは練習したい質問を選択し、質問に対する回答の様子をユーザ自身でカメラアプリなどを用いて録画す

る。録画した動画ファイルを選択しサーバーにアップロードすると、アップロードされた動画と動画ファイルから文字起したテキストが画面に表示される (図 6)。ユーザーは動画とテキストを見ながら自己評価を行う。図 7 は自己評価を行った画面である。保存ボタンをクリックすることで、回答をデータベースに保存する。



図 5 ファイル送信と文字起し画面

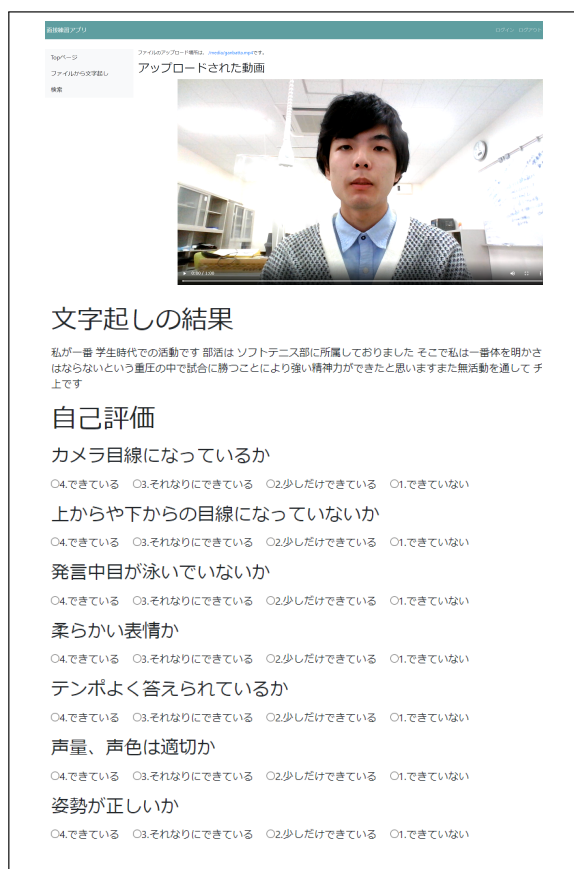


図 6 ファイル送信後の画面

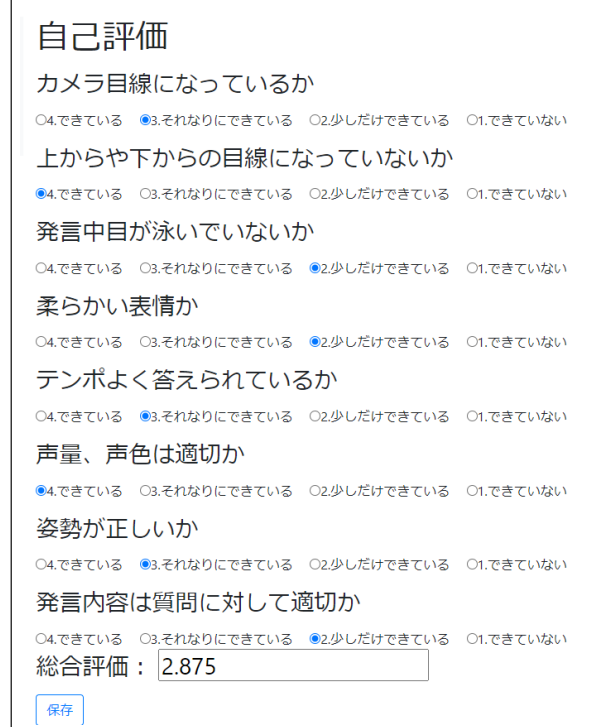


図 7 自己評価画面

3.7.5 回答検索画面

回答検索画面を図 8 に示す。この画面では現在ログインしているユーザーが自身のこれまでの回答を検索することが出来る。図 9 のように質問を選択し、対応した質問に対する回答を表示する方法と図 10 のようにユーザーの全ての回答を表示する方法がある。表示された回答の中からユーザーは見返したい回答の詳細を見ることが出来る。画面に表示されている詳細ボタンを押した場合、図 11 の画面に遷移し、ユーザは回答内容と評価項目の内訳を見返すことが出来る。



図 8 回答検索画面

検索結果			
ユーザー	質問	総合評価	詳細
kazunoko	学生時代頑張ったことは?	2.875	詳細


図 9 質問を選択した場合の検索結果画面

検索結果			
ユーザー	質問	総合評価	詳細
kazunoko	志望動機は何ですか?	3.125	詳細
kazunoko	あなたの長所と短所は?	3.0	詳細
kazunoko	志望動機は何ですか?	2.875	詳細
kazunoko	自己PRをお願いします	3.0	詳細
kazunoko	志望動機は何ですか?	3.25	詳細
kazunoko	入ってみたいことは何ですか?	2.625	詳細

図 10 全部の回答を検索した場合の検索結果画面

質問内容: 学生時代頑張ったことは?

アップロードされた動画



文字起しの結果

私が一番 学生時代での活動です 国語はソフトテニス部に所属しておりました。そこで私は一番得意な競技です。1番得意というチームリーダーの存在であるため、その責任は重たいという責任の中で試合に勝つことにより強い精神力が生まれてきました。また、運動を通して、チームメンバーと協力し、優勝は目指さなかったものの、目標が達成できたことです。

自己評価

カメラ目線になっているか	3
上からやや下からの距離になっているか	4
画面中央が適切に映っているか	2
姿勢がよい姿勢か	2
テンションが保たれているか	3
声量、声質は適切か	4
発音が正しいか	3
画面内は質問に対して適切か	2
総合評価	2.875

図 11 回答の詳細画面

4. システム検証

4.1 検証方法

開発したシステムの音声認識の一致率と文字起こしにかかる時間について検証を行った。音声認識の一致率の検証方法は、動画から手動で書き起こしたテ

キストと音声認識を用いて生成したテキストを一文単位で比較を行い、手動で書き起こしたテキストの文字数と音声認識のテキストが一致した文字数の割合を計算することで検証した。また、他の音声認識ツールとの一致率の比較を行いたいため Google ドキュメントの音声認識を用いた入力 wav ファイルの文字起こしを行い、同様に一致率を計算して比較した。

文字起こし処理にかかる時間に影響を与えるのは、動画の時間と発話者の発話量が考えられる。よって動画時間が長く発話量が少ないものと、動画時間が短く発話量が多いものを用意し、それぞれが文字起こしにかかる時間の比較を行った。

4.2 検証結果

4.2.1 音声認識から生成したテキストの一致率

検証には長さや発話量の異なる 3 つの動画を用意した。動画時間 20 秒程度の簡単な文を読み上げた動画、動画時間が 1 分程度の 200 字程度の文を読み上げた動画、動画時間 3 分程度の 550 字程度の文を読み上げた動画を用いて検証を行った。検証結果を表 6 に示す。

また、同動画を用いて、Google ドキュメントの音声認識機能を用いた文字起こしについての検証も行った。検証結果を表 7 に示す。

表 6 システムを用いた一致率の検証結果

動画時間	一致した文字数	一致率
20 秒程度	40/41	97%
1 分程度	185/212	85.3%
3 分程度	526/541	97.2%

表 7 Google ドキュメントを用いた一致率の検証結果

動画時間	一致した文字数	一致率
20 秒程度	38/41	92.3%
1 分程度	175/212	82.3%
3 分程度	488/541	90.2%

表 6 を見ると 20 秒程度の動画と 3 分程度の動画で一致率が高くなっていることがわかる。1 分程度の動画で一致率が低くなった原因として音声ファイルの声の大きさが小さかったことが考えられる。他の二つの動画に比べ、1 分程度の動画の音声ファイルはマイクとの距離が遠く、動画から音声を抽出し

た際声が小さくなってしまった。このことから一致率は動画時間や文字数に影響されず、入力音声の正確さだけに影響されることがわかる。更にシステムの検証結果を見ると、このシステムの一致率が平均90%を超えていることがわかる。

また表6と表7を比較すると、全ての動画において作成したシステムの一致率がより高くなっていることがわかる。以上のことより今回作成したシステムの音声認識の一致率は他のツールと比較しても劣っていないと言える。

4.2.2 文字起しの処理時間

速読情報館^[6]によると人が一般的に話す速度は1秒間に6~7文字程度と言われている。よって、30秒間に200文字の文を読む動画、90秒間に200文字を読む動画、30秒間に100文字読む動画、90秒間に100文字読む動画を用いて文字起しにかかる時間を検証する。検証結果を表8に示す。

表8 システムの文字起し処理時間

動画時間	文字数	処理時間	一致率
30 秒	200 文字	6.45 秒	93.1%
30 秒	100 文字	5.35 秒	98.4%
90 秒	200 文字	11.14 秒	100%
90 秒	100 文字	11.52 秒	100%

表8を見ると動画時間が増加すると、文字起しにかかる処理時間が増加していることがわかる。また、同じ時間でテキスト量を変えた動画同士を比較すると処理時間にあまり差が無いことがわかる。以上のことより、処理時間はテキスト量に影響されず、動画の時間のみに影響されることがわかる。

5. 追加機能の検討

5.1 Web アプリ上での録画機能

本アプリではユーザが録画ファイルを自身で用意し、アップロードする際にユーザに選択させている。この手法のメリットは、ユーザが好きな動画を選択できることである。しかし、質問を選択し録画を行う場合、一度アプリから離れなければならない。そこでWebアプリ内で録画を行えるような機能を追加すれば、ユーザはWebアプリから離れる必要がなくなり利便性が向上すると考えられる。

5.2 回答の削除機能

本アプリは、ユーザが使用すればするほどデータベースに回答が蓄積されていく。また、本システムの利用回数が増加するほどユーザは成長していくと考えられる。成長したユーザが回答を見返すとき、見返す必要が無い回答が残っていた場合邪魔になってしまう。しかし現在のシステムには削除機能を実施していないため、回答を削除したい場合はデータベースから削除しなければならない。そこで回答の削除機能を追加することにより、不必要になった回答を削除することが可能になり、ユーザは見返したい回答をスムーズに検索できることが可能になるためシステムの利便性が向上すると考えられる。

5.3 視線計測機能

本アプリには評価方法として自己評価を採用している。自己評価のメリットはユーザ自身で振り返ることで自身の回答をより注意深く観察することができ、自身で反省点を見つけることが出来る点である。しかし、視線の評価項目においてはどうしてもユーザの感覚に依存してしまう。そこで動画内でどのくらいカメラに視線が集まっていたかを計測する視線計測機能を実施することにより評価時の新たな指標となり、ユーザの感覚と組み合わせて評価を行うことが出来る考える。

6. 結論

本研究では2019年に発生した新型コロナウィルスより増加したオンラインの面接に対応できるWebアプリを開発することで、変化した面接様式に対応することを目標に取り組んだ。開発したアプリでは、動画から音声認識を用いて文字起しをする機能、自己評価を行う機能、動画と文字起しされたテキストと共に自己評価内容を保存する機能、保存した回答を見返す機能を実装することが出来た。これにより従来のものよりも現代で役に立つ面接練習アプリを開発できたと考えている。

2022年現在も新型コロナウィルスは猛威を振るっており、多方面に多大な影響を及ぼしている。しかしその影響は悪い影響ばかりではなく、このウィルスをきっかけに開発された新しいツールも存在するように、IT技術には良い影響を与えている面も存在する。技術者がウィルスに対抗する手段は新しい技術を生み出すことだと考える。この姿勢を忘れずにいけば技術は進歩していき、今までよりも良い生活が送れるようになると思う。このウィルス

の一連の騒動が落ち着いたとき, 飛躍的に技術が進歩していることを期待している.

謝辞

本研究において, 丁寧かつ親切なご指導を頂いた指導教員である大隈千春教授に心から感謝致します.

参考文献

- [1] HR 総研：2021 年卒及び 2022 年卒採用活動動向調査 結果報告
https://www.hrpro.co.jp/research_detail.php?r_no=273
- [2] Django
<https://docs.djangoproject.com/ja/4.0/>
- [3] Flask
<https://msiz07-flask-docs-jareadthedocs.io/ja/latest/>
- [4] 面接対策まとめ—よくある質問の回答例-マイナビ転職
<https://tenshoku.mynavi.jp/knowhow/mensetsu/>
- [5] 面接質問集
<https://partners.en-japan.com/special/mensetsu/>
- [6] 「聞く」「話す」「読む」「書く」早さはどのくらい?
<https://www.sokunousokudoku.net/media/?p=4575>