


卒業研究概要

提出年月日 2021年1月29日

卒業研究課題 マルチモーダル情報に基づく就職面接練習システムの開発	
学生番号 C17-060	氏名 竹内 直
概要 (1000字程度)	指導教員 神田 智子 教授 印
<p>コロナ禍により、人と人との接触を控えるという観点から、対面での模擬面接が避けられようになった。その為、一人で就職面接練習が可能なシステムの必要性が増していると考えられる。</p> <p>対話中に視覚的な非言語情報によって伝えられるメッセージは全体の 55%とされており[1], Washburn らは非言語行動が言語行動よりも被面接者の評価に影響すると指摘している[2]。近年、マルチモーダル情報を用いた社会的信号処理技術が対話分析に用いられており[3], 面接採用システム[4]や面接練習システム[5]にも応用されている。しかし、視線・表情を認識するシステムはあるが姿勢を認識対象にしておらず、フィードバックにCG エージェントを用いたシステムは我々が知る限り存在しない。また、[4, 5]は高価である為、多くの人が使える普及に適したシステムではない。その為、本研究では視線・表情・姿勢の非言語情報に特化し、安価な普及型 HW やオープンソース、CG エージェントを使用した面接練習システムの開発を目的とする。</p> <p>開発は、Unity, FaceAPI [6], OpenPose [7], Tobii EyeTracker 4C [8], Web カメラを組み合わせて、システム構築した。本システムは、属性入力・模擬面接・フィードバックフェーズの三部構成である。属性入力フェーズでは、使用回数・性別入力を行う。模擬面接フェーズでは、実際に模擬面接を約 2 分間受け、その際の被面接者の非言語情報の取得、動画撮影を行う。フィードバックフェーズでは、取得情報に従って CG エージェントが随時動画を一時停止しながら指摘内容に対するフィードバックを行う。</p> <p>非言語情報に関して、視線は 1 秒間隔、表情・姿勢は 3 秒間隔で取得し、凝視割合、面接官の顔から 5 秒以上注視を外したタイミング、右上・左上に注視点が移動した回数、笑顔・真顔度、表情 (6 種類)、前傾・後傾姿勢、足の開き、足が徐々に開いたタイミング、首のぶれ、肘の張り出しを検出する。姿勢検出は就職部指導の下作成した基準姿勢モデルとの対比で判断し、視線検出は当たり判定、表情検出は笑顔度の閾値判定と結果を用いて判断する。検出情報は、秒数ごとに視線・表情・姿勢配列それぞれに格納後、一つの配列にまとめ、フィードバック時に参照する。一つの配列にまとめる際、検出の重複や同じ種類に偏った指摘を避けるため、変化の多い①視線、②表情、③姿勢の順で重みづけを加えた優先順位設定を行った。</p> <p>初期評価を行うために、21~22 歳の大学生 5 名 (男性 3 名、女性 2 名) に実際に使用してもらい、ヒアリングを行った。その結果、OpenPose の認識精度、指摘に関する精度が 50%程度で十分でないこと、フィードバック回数が一人につき 2~6 回程度で指摘が少ないこと、CG エージェントとのインタラクションが一方的で、対人間に比べ劣るなどの改善点が挙げられた。今後の展望として、指摘され発覚した修正点をピンポイントに練習できる新たな機能の追加、フィードバック内容の優先度設定を行うアルゴリズムの改良、2 つ以上の非言語情報を組み合わせた検出方法の追加の必要があると考える。また、CG エージェントもしくは人によるフィードバックの傾聴効果を対比検証する必要がある。</p>	
<small>[1] Mehrabian A: Silent messages: Implicit communication of emotions and attitudes. Wadworth Publishing. Co., California. (1981)</small>	
<small>[2] Washburn P.V., Hakel M.D.: Visual cues and verbal content as influences on impressions after simulated employment interviews. Journal of Applied Psychology, 58, 137-140. (1973)</small>	
<small>[3] 岡田将吾, 松嶋良広, 中野有紀子, 林佑樹, 黄宏軒, 高瀬裕, 新田克己: マルチモーダル情報に基づくグループ会話におけるコミュニケーション能力の推定. 人工知能学会論文誌, Vol. 31, No. 6 A130-E (2016)</small>	
<small>[4] マイダスアイティ社. https://www.inair.co.jp/</small>	
<small>[5] 合田七穂, 石原圭太郎, 小尻智子: 非言語情報の特徴分析に基づいた就職面接練習支援システム. 信学技報, Vol. 116, No. 517, ET2016-98, 25-30 (2017)</small>	
<small>[6] Microsoft Azure. https://azure.microsoft.com/ja-jp/services/cognitive-services/face/</small>	
<small>[7] tf-pose-estimation. https://github.com/ildoonet/tf-pose-estimation</small>	
<small>[8] トビー・テクノロジー株式会社. https://www.tobiipro.com/ja/</small>	


 卒業研究発表
**マルチモーダル情報に基づく
就職面接練習システムの開発**
 2021年2月15日
 大阪工業大学 情報科学部 情報メディア学科
 ヒューマンインタフェース研究室
 C17-060 竹内 直

1

背景

- 新型コロナウイルスの発生により、就職活動に影響が及んでいる
- 新しい生活様式により、人と人との接触を避けるという観点から対面での面接練習も避けられるようになった

↓

- 一人で就職面接練習が可能なシステムの必要性が増していると考え

2

関連研究

- Washburnらは、面接における非言語的コミュニケーションが言語的コミュニケーションよりも被面接者の評価に影響すると指摘[1]
- 対人コミュニケーションにおいて、視覚からの非言語情報によって伝えられるメッセージは55%[2]

視覚から得る非言語情報の重要性

[1] Washburn PV, Hakel M.D. Visual cues and verbal content as influences on impressions after simulated employment interviews. Journal of Applied Psychology, 58,137-140,(1973)
[2] Mehrabian A.Silent messages: Implicit communication of emotions and attitudes. Wadworth Publishing, Co., California,(1981)

3

関連事例

- 近年、マルチモーダル情報を用いた社会的信号処理が対話分析に用いられている[3]
- AIを用いた面接採用システム[4]や就職活動生が利用する面接練習システム[5]にも応用




[3] 岡田博典, 松坂昌之, 中野裕紀子, 林祐典, 廣野祥, 高瀬祐和, 新田亮己. マルチモーダル情報に基づくグループ会議におけるコミュニケーション能力の推定. 人工知能学会論文誌, 30(3), No. 6, 130-140 (2016)
[4] 大阪工業大学. inAIR. URL: https://www.inair.co.jp/
[5] 谷田七穂, 石野圭太郎, 小沢新子. 非言語情報の特徴分析に基づいた就職面接練習支援システム. 電子情報通信学会論文誌, 98(114), No. 117, 2016, 96-103 (2017)

4

システムの設計方針

- 広く普及する必要がある
- CGエージェントによるフィードバックを用いたシステムはない
- 視線・表情に着目するが、姿勢に着目していない

- ① 安価な普及型HWやオープンソースの使用
- ② フィードバックにCGエージェントを使用
- ③ 視線・表情・姿勢の3つの非言語情報に特化
- ④ 実用性の高いシステム

5

開発環境

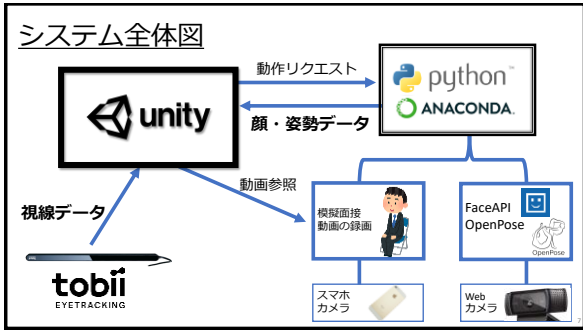
- PC: Mouse Computer GM5MPHW (GTUNE)
- OS: Microsoft Windows10 Pro
- CPU: Intel® Core™ i7-10875H CPU 2.30GHz
- GPU: GeForce® RTX 2070 SUPER™
- メモリ: 32GB



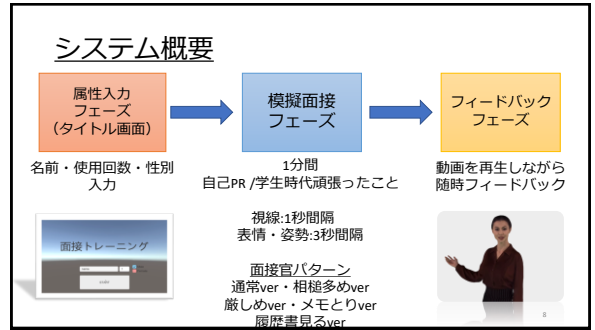

FaceAPI OpenPose

- 使用ソフトウェア:
 - Unity 2019.4.11f1 Microsoft Visual Studio 2019 Anaconda Prompt AI Talk
 - FaceAPI(Microsoft Azure) OpenPose (Tensorflow版 tf pose estimation) iVCam
- 使用CGエージェント: "NATASHA"(Metastage)
- 使用言語:
 - C# Python 3.8.3
- 使用ハードウェア:
 - Tobii Eye Tracker 4C Logitech Webカメラ C920n/C270 iPhoneSE

6



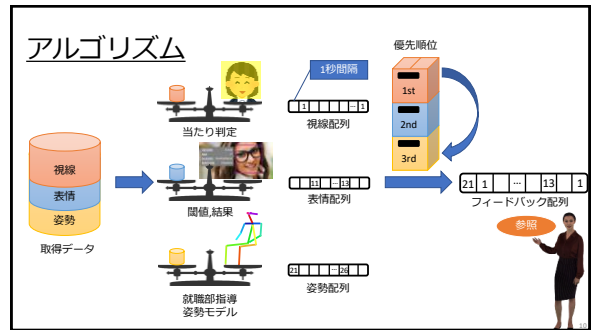
7



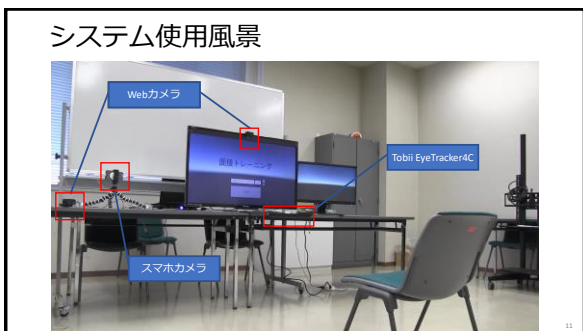
8

- ### 検出可能な情報
- **視線**
→ 凝視割合、右上・左上に注視点が移動した回数
面接官の顔から5秒以上外したタイミング
 - **表情**
→ 笑顔度、真顔度
6つの表情 (怒り、軽蔑、嫌悪、恐怖、悲しみ、驚き)
 - **姿勢**
→ 姿勢 (前傾・後傾)、足の開き、自然に足が開いたタイミング
首のぶれ、肘の張り出し

9



10



11



12

初期評価実験

- 実験参加者:21~22歳の大学生5名(男性3名,女性2名)
 - 一人あたり2,3回システムを使用してもらう
 - システム使用后、インタビュー形式でヒアリング
- ※実験参加者のうち3名はEyeTrackerでのトラッキングができなかった為、視線の計測を除外

13

13

設計方針①に関する結果と考察

- ①安価な市販HWやオープンソースの使用
- FaceAPI、OpenPose、EyeTrackerを用いて実現
 - 「**指摘の誤り**があったと思う」
- OpenPoseの認識精度
姿勢の角度を取得する為に斜めから撮影
⇒ Webカメラの増設、Kinectの使用
 - 検出精度
基準モデルの作成が不十分、表情・姿勢に関しては3秒間隔取得
⇒ 個人差に配慮した正確なモデル作成、1秒間隔の取得

14

14

設計方針②に関する結果と考察

- ②フィードバックにCGエージェントを使用
- 「動画が止まって都度言ってくれるので、**どの場面でもここ良かった/悪かったのかが明確に分かる**」
 - 「CGエージェントの場合、元々用意されたものしか再生されていないので、**個人に合わせて動いている感じがあまりしない!**」
 - 「**気遣いはCGエージェントだが人間の方が重く突き刺さる、受け止めやすい!**」
- CGエージェントとのインタラクションが一方的であり、対人間に劣る
⇒ 同じ指摘内容でも複数音声パターンを用意
⇒ 被面接者のアクションに対応したCGエージェント

15

15

設計方針③に関する結果と考察

- ③視線・表情・姿勢の3つの非言語情報に特化
- 「**フィードバック内容** (アドバイス、指摘内容、回数) が**少ない!**」
- 1回の使用あたり2~6回
⇒ 指摘内容の追加、優先順位 (重みづけ) アルゴリズムの改良
- 「手、指の動き、身体のゆれ、声の大きさ、接続詞、内容、話し方の指摘があればいい!」

16

16

設計方針④に関する結果と考察

- ④実用性の高いシステム
- 「**自信の面接動画を見る機会がないので、フィードバックの時に客観的に動画を見ながらアドバイスしてくれるのは良い!**」
 - 「**また練習**できるならしてみたい!」
 - 「**面接官がリアル**だった、本番のようで緊張感があり見られている感があった」
 - 「**一人でもできる!**」
(積極的に就職課に行って練習できない消極的な人でも使えそう)
 - 自身の面接動画をあまり**見たくない、恥ずかしい!**

17

17

今後の展望

- 視線・表情・姿勢以外の非言語情報や言語情報の取得
- 指摘され発覚した修正点をピンポイントに練習できる機能の追加
被面接者が自身の面接動画を振り返る機能の追加
- 2つ以上の非言語情報を組み合わせた検出方法の追加
- CGエージェント/人間によるフィードバックの傾聴効果の対比検証

18

18