

# 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

Development of a Job Interview Training System with Multi-modal Behavior Analysis

竹内 直, 神田 智子

Nao Takeuchi and Tomoko Koda

大阪工業大学情報科学部情報メディア学科

Department of Media Science, Faculty of Information Science and Technology,  
Osaka Institute of Technology

**Abstract:**本研究では、被面接者の視線・表情・姿勢の非言語情報を視線追跡装置、カメラ画像を用いて認識し、被面接者の模範的な非言語行動モデルとの比較を行うことで、振り返りを行いながら改善点を指摘する面接練習システムの開発を行った。開発の方針は、普及型 HW とオープンソースを用いた安価かつ容易なシステム構築が可能であること、および CG エージェントによる振り返りが可能であることである。本システムの初期評価結果として、現段階では、精度が 50%程度で十分でないこと、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物足りないこと、フィードバックする回数が一人につき 2~6 回程度で指摘回数や内容が少ないことなどの改善点が挙げられた。

## 1. はじめに

コロナ禍の影響が就職活動に及んでおり、人々との接触をできるだけ控えるという観点から、これまで就職支援の一環として行われていた対面での模擬面接が避けられようになった。このことから、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。

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している[2]。また、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], AIを用いた面接採用システム[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]は高価なシステムである為、多くの人が使える普及に適したシステムではない。これらのことから、本研究では視線・表情に姿勢を加えた 3つ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、そして CG エージェント(図 1)を使用した面接練習システムの開発を目的とする。

## 2. 就職面接練習システム

### 2.1 システム概要

開発は、Unity 2019.4.11f1 をベースに、FaceAPI[6], OpenPose[7], TobiiEyeTracker4C[8], Web カメラを組み合わせて、システム構築を行った。本システムは、属性入力フェーズ、模擬面接フェーズ、フィードバックフェーズの三部構成となっている。属性入力フェーズ(図 2)では、使用回数・性別入力を行う。模擬面接フェーズ(図 3)では、実際に模擬面接を約 1~2 分間受け、被面接者の視線・表情・姿勢の情報を取得し、模擬面接の様子を斜めから撮影する。フィードバックフェーズ(図 4)では、撮影された動画を再生し、取得した情報に従って 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。



図 1 使用 CG エージェント

図 2 属性入力フェーズ



図3 模擬面接フェーズ (左モニターに面接官) 図4 フィードバックフェーズ (右モニターにCGエージェント)

## 2.2 検出可能な非言語情報

非言語情報に関して、視線は1秒間隔、表情・姿勢は3秒間隔で取得する。取得情報から凝視割合、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、笑顔度、真顔度、表情(6種類)、姿勢(前傾・後傾)、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。

## 2.3 検出方法

姿勢検出に関しては大阪工業大学就職部の指導のもと、OpenPoseを用いて正しい姿勢モデルを男女別で作成し、そのモデルと被面接者から取得した姿勢の対比で判断する。また、視線検出に関しては、面接官の顔を対象とし、対象から5秒以上離れたなどの当たり判定、表情検出に関してはFaceAPIによる笑顔度の閾値判定、および表情認識結果より判定する。

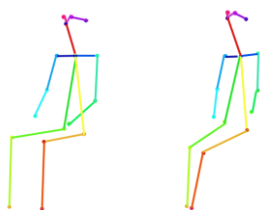


図5 基準姿勢モデル(左:男性, 右:女性)

## 2.4 フィードバックアルゴリズム

検出した情報は、秒数ごとに視線・表情・姿勢配列それぞれに格納し、一つの配列にまとめ、CGエージェントがフィードバックする際に参照する。一つの配列にまとめる際、同じ秒数で複数の検出が重複してしまうことや、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付けを行った。

## 3. 初期評価実験

本システムの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)に実際に本システムを使用してもらい、インタビュー形式でヒアリングを行った。なお、被験者実験にあたり、大阪工業大学よりヒト対象実験実施の承認済みである。

### 3.1 精度に関する結果と考察

ヒアリング結果として、「(姿勢に対する)指摘の誤りがあったと思う」、「指摘されると思ったが指摘されなかった」というような精度に関するコメントが挙げられた。また、コメントより総合的に判断すると50%程度の精度であったと考えられる。

この原因として、まずOpenPoseの検出精度が低かったのではないかと考える。本システムでは、姿勢(前傾・後継)の角度を取得するために、斜め45°にWebカメラを配置した。OpenPoseの性質上、斜めからだけでは精度が下がる為、その対策として、正面や横にWebカメラの増設もしくはKinectでの代用で、精度向上が見込めるのではないかと考える。

次に、姿勢に関する検出誤差があったと考える。姿勢検出に関して、被面接者の姿勢と基準モデルを対比し、逸脱した箇所を閾値判定により検出する方法を用いた。その基準モデルが男女1パターンずつで個人差を配慮していなかった為に、姿勢の逸脱箇所の検出の誤差が見られたのではないかと考える。その対策として、基準モデルを数パターン用意し、ユーザの身長や体格にできるだけ近いものを使用することで精度向上が見込めるのではないかと考える。

最後に、姿勢・表情に関する検出漏れがあったと考える。表情・姿勢検出に関して、FaceAPIの無料版が1分間に20トランザクションのみの使用しかできない。また、OpenPoseのFPSを上げると処理が重くなることから、3秒間隔で取得を行った。その為、3秒の間で基本モデルから逸脱している箇所の検出ができていない部分があったことが考えられる。

### 3.2 CGエージェントに関する結果と考察

ヒアリング結果として、「動画が止まって(CGエージェントが)指摘やアドバイスを都度言ってくれるので、どの場面でどこが良かった/悪かったのかが明確に分かる」という良い点が挙げられたのに対し、「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、個人に合わせて助言している感じがあまりしない」、「気軽さは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」というような悪い点が挙げられた。

この原因として、CG エージェントとのインタラクションが一方的であり、対人間に比べ劣ることが考えられる。その対策として、同じ指摘内容でも複数音声パターンを用意、またCG エージェントとのインタラクションの機会を増やす必要があると考える。例えば、模擬面接の感想や自己評価を話すフェーズを作成し、CG エージェントがその内容に対して、褒める・勇気づける・励ますといった応答をするなどの実装が考えられる。

### 3.3 フィードバックに関する結果と考察

ヒアリング結果として、「フィードバック内容（指摘の種類、回数）が少ない」というコメントが挙げられ、実際に一人あたり2~6回程度であった。

この原因は、検出精度とも関わりはあるが、他にフィードバック配列を作る際の重みづけを加えた優先順位アルゴリズムにも原因があると考えられる。本システムでは、面接時重要視され、時間経過で変化の多い①視線②表情③姿勢の順で優先順位付けを行い、検出された種類の順位を下げる形式に、重みづけ（順位がしばらく更新されなかった時に順位付けを再構築する）を加えた。実際検出はしているが、フィードバック配列に反映されなかったというケースも考えられるため、重みづけを加えたアルゴリズムの改良が必要である。

また、検出する箇所の追加も必要であると考えられる。「手、指の動き、体の揺れ、声の大きさ、接続詞、質問に対する回答の内容、話し方の指摘があればいい」というコメントも挙げられたように、非言語情報に加え、言語情報の指摘も必要であることが指摘されている。

### 3.4 ユーザビリティに関する結果

ヒアリング結果として、「自身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」、「また練習が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）」というような好評価のコメントが挙げられた。

## 4. おわりに

本研究では、被面接者の視線・表情・姿勢の非言語情報を視線追跡装置、カメラ画像を用いて認識し、被面接者の模範的な非言語行動モデルとの比較を行うことで、振り返りを行いながら改善点を指摘する面接練習システムの開発を行った。開発の方針は、普及型HWとオープンソースを用いた安価かつ容易なシステム構築が可能であること、およびCG エージェントによる振り返りが可能であることである。

初期評価実験を行い、現段階では、精度が50%程度で十分でないこと、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物足りないこと、フィードバックする回数が一人につき2~6回程度で指摘回数や内容が少ないことなどの改善点が挙げられた。

今後の展望として、初期評価実験で挙げられた改善点を解決し、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や被面接者が自身の面接動画をシステム使用後振り返ることのできる機能の追加、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が必要であると考えられる。また、CG エージェントによるフィードバックと人によるフィードバックの傾聴効果を対比し、検証する必要がある。

## 謝辞

本研究の一部は、科研費「基盤(C)20K11926」の交付を受けて実施した。

## 参考文献

- [1] Mehrabian A: Silent messages: Implicit communication of emotions and attitudes. Wadworth Publishing.Co., California.(1981)
- [2] Washburn P.V., Hakel M.D.:Visual cues and verbal content as influences on impressions after simulated employment interviews.Journal of Applied Psychology, pp.58,137-140.(1973)
- [3] 岡田将吾,松儀良広,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)
- [4] 株式会社マイダスアイティジャパン, <https://www.ina-ir.co.jp/>

- [ 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)
- [ 6 ] Microsoft Azure, <https://azure.microsoft.com/ja-jp/services/cognitive-services/face/>
- [ 7 ] tf-pose-estimation, <https://github.com/ildoonet/tf-pose-estimation>
- [ 8 ] トビー・テクノロジー株式会社,<https://www.tobiipro.com/ja/>