

# 言語と非言語情報を用いた医療面接の自動評価システム

佐伯 蒼士郎<sup>†</sup> 西田 昌史<sup>†</sup> 田中 悟志<sup>‡</sup>

静岡大学<sup>†</sup> 浜松医科大学<sup>‡</sup>

## 1. はじめに

医療従事者が患者に対して行う医療面接は、あらかじめ決められた項目について人手で評価を行っているため、コストがかかっている。現在の評価を自動で行い適切なフィードバックを返すシステムを構築し現場に導入することで、医療従事者の業務負担の軽減や効率的に医療従事者の医療面接の指導を行えると考えられる。

関連研究として、藤倉ら[1]は模擬患者ロボットである SAYA を利用して、医療面接を行い、その中で医療従事者の発話に含まれるキーワードから正しい質問を行えているかを評価するシステムの構築を行っている。また、本間ら[2]は模擬患者ロボットを用いて医療面接の練習を行えるような対話システムを提案しており、医療従事者の顔きから評価を行う方法を提案している。さらに、タケロボ株式会社提供「医療面接・鑑別診断トレーニング AI システム」では、模擬患者役のチャットボットと医療面接の練習を行い、問診事項ごとに評価を行うシステムを提案している。従来研究では言語情報のみに着目しているが、医療従事者が患者とのコミュニケーションを行う上で、韻律などの非言語情報も重要であると考えられる。医療面接の評価では、韻律などの非言語情報を評価項目に加えることでより正確な評価を行うことができるのではないかと考える。

そこで、本研究では医療従事者と患者との医療面接の動画を収集し、これらの音声に対して音声認識を行うことで得られる言語情報に基づく評価と、韻律に基づく非言語情報も考慮した評価を行う手法を提案し、有効性を評価する。

## 2. 提案システム

提案手法の構成図を図 1 に示す。提案手法である評価システムは医療面接を撮影した動画から言語情報と非言語情報を抽出する。言語情報の評価は、実際に現場で使用されている評価項目に沿って評価する。非言語情報の評価は、患者の韻律に医療従事者が同調しているかどうかを評価対象とする。

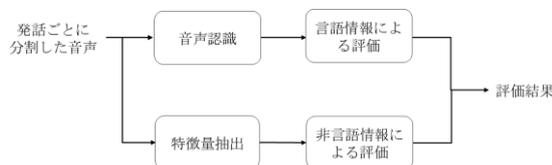


図1 提案システムの構成図

### 2.1. 言語情報による評価

言語情報の評価は、発話ごとに音声を分割し、音声認識を行うことで言語情報を抽出し、言語情報による評価への入力とする。音声認識には AmiVoice API を使用する。言語情報の評価項目は表 1 の通りである。

表 1 言語情報による評価項目

番号	評価項目
1	医療従事者の感情分析
2	患者の感情に気付いているか
3	専門用語を発話していないか
4	オウム返しをしているか
5	相槌を行っているか
6	オープンクエスチョンやクローズドクエスチョンを行っているか

評価項目 1 では、感情分析を用いた oseti[4] を使用する。評価項目 2 では、「～なはずですよね」や「～でしょうね」などの発話が含まれているかを評価する。評価項目 3 では、DMIME 医療用語変換辞書を用いて専門用語の検出を行う。評価項目 4 では直前の患者の発話で現れたキーワードが医療従事者の発話で検出されるかを評価する。評価項目 5 では「うんうん」などの相槌の発話が含まれているかを評価する。評価項目 6 では疑問詞を使った質問か「はい」や「いいえ」で答えられる質問をしているかを評価する。

### 2.2. 非言語情報による評価

非言語情報の評価は、発話毎に分割された音声から非言語情報を取得し、評価システムへの入力とする。患者と医療従事者の非言語情報がどの程度同調しているかを評価する。評価対象の非言語情報は声の大きさ、声の高さ、発話速度、ポーズ区間長である。声の大きさと高さは librosa を用いて特徴量の抽出を行う。発話速度は、分割した発話毎で発話速度を算出する。

## 3. 評価実験

言語情報の評価実験では、あらかじめ評価項目に対する正解ラベルを人手で設定し、システムが行った評価結果と正解ラベルを比較し、その正答率からシステムで評価した。非言語情報の評価実験では、非言語情報毎に医療従事者が

Automatic Evaluation System Using Verbal and Nonverbal Information for Medical Interviews

<sup>†</sup> Soshiro Saeki • Shizuoka University

<sup>†</sup> Masafumi Nishida • Shizuoka University

<sup>‡</sup> Satoshi Tanaka • Hamamatsu University School of Medicine

患者に同調できているかを、熟練度合いを考慮に入れ評価した。今回の実験では、浜松医科大学で撮影した医療面接の動画である、熟練者を模した医療従事者と患者との動画(33分)、熟練者でない医療従事者と患者との動画(18分)、熟練医療従事者と患者との動画(20分)の3つを使用した。使用する3つの動画の音声認識精度は、熟練者ではない動画が77.43%、熟練者を模した動画が77.94%、熟練者の動画が76.39%であった。

3.1. 言語情報の評価結果

表1で示した各々の評価項目に対して性能を検証するため、3つの動画を使用して検証を行った。各々の評価項目の正解ラベルを用いて求めた正答率を示した結果を表2に示す。表2中の「書」と「認」は、音声ファイルを書き起こしたものと音声認識したものである。

表2 言語情報の評価結果

		感情分析			感情に気付いているか	専門用語	オウム返し	相槌	質問の種類	
		ポジティブ	ニュートラル	ネガティブ					オープンクエスチョン	クローズドクエスチョン
熟練者でない動画	書	100% (8/8)	100% (59/59)	100% (12/12)	-	-	100% (9/9)	100% (36/36)	100% (18/18)	60% (9/15)
	認	75% (6/8)	95% (56/59)	75% (9/12)	-	-	78% (7/9)	58% (21/36)	100% (18/18)	47% (7/15)
熟練者を模した動画	書	100% (17/17)	100% (58/58)	100% (19/19)	100% (3/3)	-	100% (19/19)	100% (39/39)	80% (33/41)	68% (13/19)
	認	82% (14/17)	95% (55/58)	68% (13/19)	100% (3/3)	-	53% (10/19)	51% (20/39)	80% (33/41)	68% (13/19)
熟練者の動画	書	100% (24/24)	100% (40/40)	100% (10/10)	100% (1/1)	-	100% (5/5)	100% (10/10)	100% (9/9)	67% (12/18)
	認	71% (17/24)	88% (35/40)	60% (6/10)	100% (1/1)	-	40% (2/5)	50% (5/10)	67% (6/9)	56% (10/18)

評価結果から、相槌や質問の種類の評価精度は高い一方、オウム返しや感情に気付いた発話を行っているかの評価精度が低い傾向にあった。

3.2. 非言語情報の評価結果

撮影した3つの動画から、非言語情報を抽出し、平均と相関を表3~6のように求め、評価を行った。

表3 声の大きさの平均(RMS)と相関

	平均		相関
	医療従事者	患者	
熟練者ではない動画	0.018	0.009	-0.332
熟練者を模した動画	0.014	0.011	-0.151
熟練者の動画	0.017	0.019	0.069

表4 声の高さの平均(Hz)と相関

	平均		相関
	医療従事者	患者	
熟練者ではない動画	200	186	0.142
熟練者を模した動画	267	305	-0.015
熟練者の動画	220	249	0.343

表5 発話速度の平均(モーラ/秒)と相関

	平均		相関
	医療従事者	患者	
熟練者ではない動画	5.424	5.613	-0.098
熟練者を模した動画	5.738	5.066	0.126
熟練者の動画	5.797	5.595	0.085

表6 ポーズ区間長の平均(ミリ秒)と相関

	平均		相関
	医療従事者	患者	
熟練者ではない動画	332	420	-0.185
熟練者を模した動画	453	505	0.107
熟練者の動画	348	399	0.022

さらに、発話速度の変化を可視化したグラフを図2と3に示す。

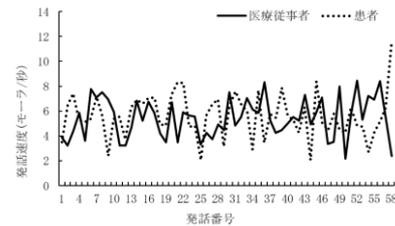


図2. 熟練者でない医療従事者の発話速度の可視化

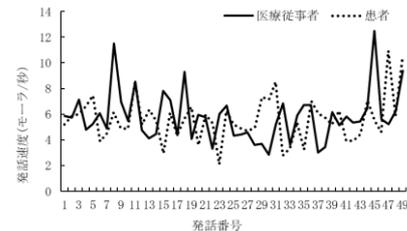


図3. 熟練医療従事者の発話速度の可視化

結果からどの韻律でも、相関が高いことから熟練医療従事者の方が患者の韻律に同調していることが分かる。また、図2と3のように韻律をグラフ化して可視化することで、医療従事者に対してフィードバックが可能になる。

4. おわりに

本研究では医療従事者への業務負担の軽減とフィードバックを目的として言語情報に基づく評価と、韻律に基づく非言語情報も考慮した評価を行う手法を提案した。結果から、言語情報の評価システムを用いることで、人手で評価した場合と同じようなフィードバックを返すことを評価できた。また、熟練医療従事者の方が患者の韻律に同調している結果が得られた。今後の課題として、今回発話を手動で分割したので、発話の分割を自動で行い、より多くの医療面接の動画で評価を行う予定である。

参考文献

[1] 藤倉輝道, 内藤知佐子, 羽場政法, 高橋優三, "人工知能 (AI) をいかにしてシミュレーション医療者教育に活かすか?", 日シミュレーション医療学会誌, 第9巻, 2021.

[2] 本間悟史, 橋本卓弥, 藤倉輝道, 早坂明哲, 竹下俊行, 伊藤保彦, 大久保公裕, 竹村裕, "模擬患者ロボットの音声対話システムの構築と面接者の頷きの抽出", 日本機械学会, ロボティクス・メカトロニクス講演会, 2021 巻, 2021.

[3] 医療面接・鑑別診断トレーニング AI システム, [http://www.takerobo.co.jp/medical\\_interview.html](http://www.takerobo.co.jp/medical_interview.html).

[4] Ikegami yukino, 2023, oseti, <https://github.com/ikegami-yukino/oseti>.