

情報技術活用による NCPR 講習の実施支援システムの開発とその効果検証

西本騰^{†1} 安藤潤人^{†1} 野間春生^{†1}

概要: 本研究では、新生児蘇生法 (Neonatal Cardio Pulmonary Resuscitation : NCPR) 講習の支援を目的とし、情報技術を利用して、その実施を補助するシステムを開発した。全出生の内、約 5% の新生児が人工呼吸を始めとする積極的な処置を必要とする中、新生児の蘇生を体系化した NCPR 講習の普及は重要である。しかし、シミュレータを用いたシナリオ実習やビデオを用いたデブリーフィングにはいくつかの課題が存在した。第一にシミュレータの価格と性能のトレードオフ、第二にビデオを用いたデブリーフィングの運用コスト、第三にデブリーフィング時の状況把握の難しさである。そこで、本研究では模擬医療機器やスマホアプリを用いたシミュレータの拡張や、ビデオからの教材自動生成、要約システムを開発した。検証実験の結果、シナリオ実習のリアリティの向上やデブリーフィングにかかる時間の短縮、要約機能の有効性が確認された。本研究成果は、NCPR 普及の一助となることが期待される。

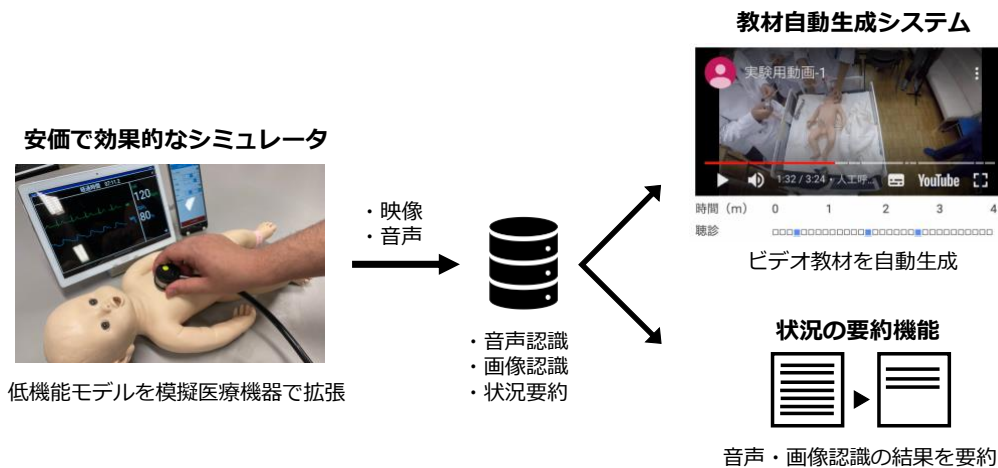


図 1: 提案システムの全体図

1. はじめに

出生時に子宮外生活に必要な呼吸循環動態の移行が順調に進行しない事例は、全出生の約 15% にみられ、約 5% が人工呼吸をはじめとする積極的な新生児蘇生法 (NCPR: Neonatal Cardio Pulmonary Resuscitation) 処置を必要としている。そこで、日本周産期・新生児医学会では「すべての分娩に新生児蘇生法を習得した医療スタッフが新生児の担当者として立ち会うことができる体制」の確立を目標に、2007 年から NCPR 普及事業を開始した[1]。NCPR 普及事業は、資格を有する専門的な講師によって講習形式で実施される。講習は、座学・実習・デブリーフィングから構成される。座学では、NCPR の基本的な理論や背景知識を習得する。実習では、NCPR 技術習得を目的として行われ、生徒はシミュレータを使用し、実際に NCPR を実施する方法を学ぶ、デブリーフィングでは、実習後の反省やフィードバックを行い、生徒は知識や行動の改善を行う。

実習で使用されるシミュレータには、高機能・低機能モデルの 2 種類が存在し[2]、その学習効果と導入コストはトレードオフの関係にある。高機能モデルは、モータやスピ

ーカーなどの状態再現機構を有し、学習効果が高いとされる[3]。しかしながら、高機能モデルは導入コストが高いため、普及しているシミュレータの多くは、再現機構を持たない低機能モデルである。このため、低機能モデルを使用した実習では、講師が心拍数を再現するために机を一定のタイミングで叩くなど (図 2)、実習の質を高機能モデルと同様に保つために講師は指導とは直接関わりのない部分で努力をする必要がある。しかしながら、現実とかけ離れた再現を行うために、リアリティの欠如、それに伴う学習効果の低下が課題となっていた。

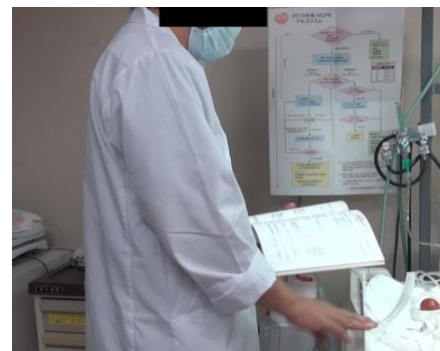


図 2: 心拍数の再現の様子

^{†1} 立命館大学
Ritsumeikan University

また、デブリーフィングでは、実習中に講師が紙のチェックリストを用いて生徒の行動を記録し、フィードバックを行う。しかしながら、講師は、実習中にシミュレータ操作や指導などの高度なマルチタスクの必要があり、チェックリストを記載する余裕がない。このため、デブリーフィングの実施に課題がある。そのため、ガイドラインにおいて、実習の様子を撮影したビデオの使用が推奨されているものの[1]、単にビデオを使用するだけでは、ビデオを全編再生する必要があるなど、デブリーフィングの効果的な実施は困難とされる[4]。

この様に、NCPR 講習の実施には、講師の多大な努力が求められ、情報技術による支援が必要である。そこで、本研究では、NCPR 講習の実施を情報技術によって支援し、NCPR の普及を助けることを目的とし、支援機能の開発を行った(図1)。実習に関しては、低機能モデルを模擬聴診器・模擬心電図によって拡張し、導入コストを抑えつつ学習効果を高める手法を提案した。これにより、実習のリアリティの向上、マルチタスク支援の効果が認められた[5]。デブリーフィングに関しては、撮影したビデオからデブリーフィング教材を自動生成するシステムを提案した。この提案により、デブリーフィングにかかる時間の短縮などの効果を得た[6]。また、これらの先行研究からのさらなる発展として、デブリーフィング時に生徒・講師の状況理解を促進するための要約機能を開発した。この要約機能は、画像認識・音声認識の結果を ChatGPT にて要約することで実装している。

本稿の構成は以下の通りである。はじめに、NCPR 講習における関連研究や、実習・デブリーフィングの課題を述べる。次に、その課題への解決策を設定する。そして、各システムの検証実験、実験結果を示す。おわりに、実験結果への議論、今後の課題について示す。

2. 関連研究

本章では、関連研究として、シナリオ実習で用いられるシミュレータ、デブリーフィングの実際について述べる。

2.1 シミュレータ

シミュレータには、大別して、高機能モデルと低機能モデルの二種類が存在する。両モデルの詳細について述べる。

2.1.1 高機能モデル

Laerdal 社の Sim-Baby[7]、高研の LM-111[8]のような高機能モデルは、新生児人形自体が、実際の新生児の状態を再現できるロボットである。高機能モデルは、新生児人形内部にモータやスピーカーなどの再現機構を備えているため、心音や手足の動き、泣き声、脈拍の再現が可能である。また、外部アタッチメントとして、模擬心電図も提供されており、実際の蘇生と同様の体験を提供可能である。さらに、センサーを内蔵し、センサ情報から手技の評価も可能なモデルも存在する[9]。

2.1.2 低機能モデル

低機能モデルは新生児の体の一部を再現したマネキン人形であり、内部に再現機構や外部のアタッチメントを持たないため、高機能モデルと比較して、低価格に抑えられている。しかしながら、再現機構を有さないために、講師は心拍数を再現するために、机を一定のタイミングで叩く、または、メトロノームを使用する。泣き声を再現するために泣き真似をする。心電図を再現するために、厚紙に記載された SpO₂ (血中酸素濃度) や心拍数を適切なタイミングでめくることによって心電図を模擬する(図3)などの代替手法を用いる必要がある。



図3：紙を用いた心電図の再現手法

2.2 デブリーフィング

NCPR 講習においては、自身の行動を客観的に振り返ることができるビデオを用いたデブリーフィングが推奨されているものの、単にビデオを再生するだけでは、NCPR 講習実施に支障が出る。そこで、ふりかえ朗[11]などでは、高機能シミュレータに内蔵したセンサと連携した行動の記録を行い、行動のブックマークを行うことにより、デブリーフィングを効率化するための手法が提案されている。

しかし、このような高機能シミュレータを導入できる医院は限られており、多くの現場では、紙の自己チェックシートなどを用いたデブリーフィングが主流である(図4)。ここでは、生徒、講師がシナリオ実習前後でチェックシートに記載を行い、デブリーフィングの際の反省材料とする。

チェック項目		実習前	実習後
知識	・人工呼吸の開始基準を述べることができる	<input type="checkbox"/>	<input type="checkbox"/>
	→蘇生の初期処置を行った後、無呼吸または100%/分未満の強弱を認める場合	<input type="checkbox"/>	<input type="checkbox"/>
	・初期処置に反応しない場合、おそくとも生後60秒までに、有効な人工呼吸を開始しなければいけないことを理解する	<input type="checkbox"/>	<input type="checkbox"/>
	・正期産児の場合、人工呼吸開始時の酸素濃度がわかる	<input type="checkbox"/>	<input type="checkbox"/>
	→21%、すなわち空気をを用いる	<input type="checkbox"/>	<input type="checkbox"/>
準備	・バッグは、新生児用を準備する	<input type="checkbox"/>	<input type="checkbox"/>
	・使用前点検をする(過剰加圧防止弁の作動・バッグ損傷等)	<input type="checkbox"/>	<input type="checkbox"/>
	・マスクのサイズを選択する基準を述べるができる	<input type="checkbox"/>	<input type="checkbox"/>
	→口と鼻を覆うが、眼にかからないサイズを選択する	<input type="checkbox"/>	<input type="checkbox"/>

図4：紙のチェックリスト項目

3. 課題

本章では、NCPR 講習において、IT 技術による支援が特に不足しているシナリオ実習・デブリーフィングを取り巻く3つの課題について述べる。

3.1 課題1：シミュレータの価格と性能のトレードオフ

シナリオ実習において、実際の新生児を使うことは、新生児の命を危険にさらしてしまうため、シミュレータが使用される。シミュレータには大別して、高機能・低機能モデルの2種類が存在し、その学習効果と導入コストはトレードオフの関係にある。高機能モデルは再現できる状況が多く、高機能モデルを用いた実習は学習効果が高い。

しかしながら、高機能モデルは高額であり、メンテナンスや運用のために、専門の人員を用意する必要もあり[12,13]、一般的に普及しているシミュレータは安価な低機能モデルである。低機能モデルは、高機能モデルの持つ状態再現機構を有さないため、低機能モデルを用いたシナリオ実習では、講師は状況再現のために代替手法を用いる。このため、リアリティの欠如、モチベーションの低下が指摘されている。また、シナリオ実習のリアリティが講師のスキルに依存し、属人性が高いことも課題である。

3.2 課題2：VADの運用コスト

デブリーフィングには大別して、口頭でのデブリーフィング（VD：Verbal Debrief）、ビデオを用いたデブリーフィング（VAD：Video Assisted Debrief）の2種類が存在する。VDでは、講師はシナリオ実習中に紙のチェックリストを用いて生徒の行動の記録を行い、記録を元にVDを実施する。しかしながら、シナリオ実習中に高度なマルチタスク（シミュレータの操作・生徒へのアドバイスなど）を求められる講師が、生徒の行動を観察し、チェックリストに記載をすることは難しい。

そのため、自身の記憶に左右されることなく、行動をビデオによって客観的に振り返ることができるVADがガイドラインにおいて推奨されている。VADの利点として、自己省察が促され、手技技術の向上[14]、蘇生ガイドラインの遵守[15]などの効果が認められている。しかしながら、多くの病院では、シナリオ実習の様子を撮影するビデオカメラ、ビデオカメラの録画開始・停止を行う人員・機材を用意することができず、その運用の知見も少ないため、VADを実施することは難しく、広く実施はされていない。

3.3 課題3：状況把握の難しさ

単純に撮影したビデオを用いてVADを実施しても、指摘したいシーンを記憶していなければ、ビデオを全編再生する必要がある。そのため、VADの実施時間はシナリオ実習と同じ、あるいはそれ以上となり得る。さらに、重要なシーンの視聴は、生徒の集中力を散漫させる恐れがある[16]。また、シナリオ実習中の生徒は緊張状態にあり、自身の行動や発言を記憶できないことも多く、デブリーフィング時に過去の自分の行動を指摘されても、それを適切に思い返すことは難しい。

また、講習中にマルチタスクを行う必要がある講師も同様に緊張状態にあり、指摘したいシーンを記憶することは難しい。仮にシーンを記憶していたとしても、シナリオ実

習中の具体的な状況や、生徒の行った判断のプロセス、発言内容のすべてを完全に記憶することは容易ではない。これは、VADの効果的な運用における課題となる。

4. NCPR 講習実施支援システム

本章では、前章で述べたシナリオ実習・デブリーフィングにおける課題を解決する新機能について述べる。なお、NCPRの普及を支援する観点から、各機能は普及している低機能シミュレータを物理的に改造することなく、市販の電子部品やスマホ、PC、プログラムから実装している。

4.1 解決策1：安価で効果的なシミュレータ

シナリオ実習においては、再現機能の不足によるリアリティの欠如、それに伴うモチベーションの低下が課題である。先行研究では、スピーカーやアクチュエータによって、新生児の泣き声や心拍数、手足の動きを再現している。しかしながら、この手法は、コストやメンテナンスの必要性から、導入コストが高い。そこで、本研究では、普及している低機能モデルをスマホと一般的な電子部品で作成した模擬医療機器によって拡張し、講習に必要な機能を持つシミュレータを実現する。これにより、導入コストを抑えつつ学習効果の向上を目指した[5]。また、現在の試作機では、iOS・Android両端末への対応、模擬聴診器のモジュール化、遠隔講習のためにハブ端末の開発を行った。

4.2 解決策2：VAD教材自動生成システム

デブリーフィングにおいては、VADを実施のための実施・導入コストの課題が指摘されている。そこで、本研究では、画像認識技術・WEBサービスを用いて、状況を認識し、認識結果から教材を自動生成するVAD教材自動生成システムを開発した。これにより、VAD実施のための実施・導入コストを抑えながら、学習効果の向上を目指した[6]。

4.3 解決策3：講習内容の要約機能

デブリーフィングの際、講師は、シナリオ実習中の生徒の具体的な行動、発言、および判断のプロセスに基づいて、指導を行うのが理想的である。しかしながら、単にビデオを再生するだけでは、シナリオ実習中の具体的な流れを理解することは難しい。この解決として、本研究では、音声認識・画像認識・ChatGPT[17]を組み合わせたアプローチを採用した。具体的には、シナリオ実習中に撮影したビデオの音声・映像から、シナリオ実習中の状況を要約する機能を開発し、シナリオ実習の理解の補助を目指した。

5. 解決策1：安価で効果的なシミュレータ

本章では、シナリオ実習における課題「シミュレータの価格と性能のトレードオフ」を解決するシミュレータの設計、検証実験について述べる[5]。

5.1 シミュレータの設計

本節では、開発したシミュレータの設計・機能の詳細について述べる。シミュレータは、コントローラ、模擬聴診

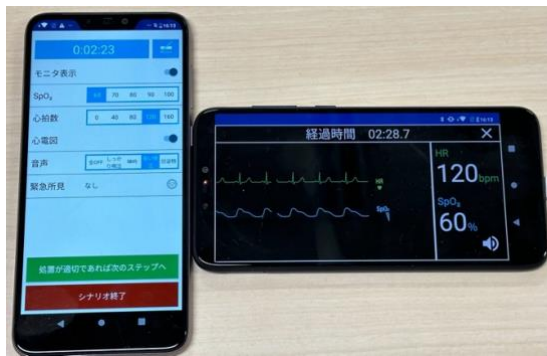


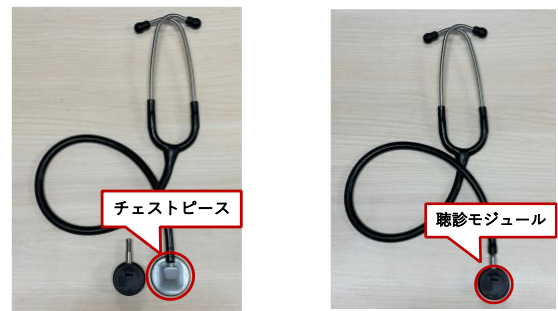
図 5：コントローラと心電図

器、模擬心電図、ハブ端末からなる。コントローラは新生児のバイタル（心拍数や SpO₂、泣き声）、模擬心電図の状態変更が可能である。本機能はワンタッチで状態変更を可能とし、シミュレータ操作の負担軽減を狙った。各端末とのデータ共有には、クラウド DB である Firebase を用い、複数端末を操作できる（図 5）。

また、従来の講習では、講師が口頭で心電図の情報を伝えるなどの代替手法を用いている。そこで、実際の心電図と同様の体験を提供する模擬心電図を開発した。模擬心電図は、コントローラで設定した心拍数などの値を元に、心電図を再現する。さらに、先行研究では対応環境が iOS のみであったが[5]、本研究にて Android への対応を追加し、対応プラットフォームを拡大した。従来の高機能モデルは、泣き声などの状態再現のために、内部にスピーカーやバッテリー、無線機器、専用の計算機が必要であり、コストが増大していた。本研究では、スマホを計算資源として捉え、スマホにアプリを DL するだけで利用できる。このため、新たに新生児モデルや計算資源の購入が不要であり、従来の高機能モデルと比べて、コスト面で優位である。

さらに、心音を再現するために模擬聴診器を開発した（図 6）。内部構造では、フォトランジスタと Bluetooth イヤホンを組み込み、聴診時の接触判定を通じ、心拍を再生する。先行研究では、チューブ部分からチェストピース部分への配線が必要であり、聴診器を改造する必要があったが[5]、本研究では、モジュール化を実施し、チェストピース部分を交換するだけで使用可能な設計とし（図 6 a, b）、汎用性を高めた。さらに、接触判定を変更し、低照明環境での誤動作を防止した。なお、模擬聴診器の基本的な構成はセンサ内蔵の Bluetooth イヤホンと同様であり、大量生産で製造コストを抑えられる。故障時も聴診モジュールのみを交換すればよいため、修理コストを抑えられる。

感染症対策、離島や僻地での NCPR 講習実施のために、Zoom 等を用いた遠隔講習が求められている。本シミュレータはクラウド DB を用いており、模擬聴診器を除く端末は遠隔地から操作できる。先行研究では、模擬聴診器とコントローラは Bluetooth 接続だったため、遠隔講習では聴診



a) モジュール交換前

b) モジュール交換後

図 6：模擬聴診器

機能を使用できなかった[5]。そこで、ハブ端末を開発し、遠隔地でも聴診機能を使用可能とした。

5.2 シミュレータの検証実験

シミュレータの効果検証として、医師、医学生を対象として実験を行った。以下、その詳細を述べる。

5.2.1 評価実験

実験対象者は、講師 4 名、生徒 9 名である。なお、講師は NCPR インストラクター（NCPR に関する技術・知識を有し、NCPR 講習の開催資格を持つ）であり、なおかつ、勤続年数 10 年以上の医療関係である。効果検証のために、従来方式のシナリオ実習（低機能モデルを用いた実習）、提案方式のシナリオ実習（本提案モデルを用いた実習）を実施させた。シナリオの難易度は、ルーチンケアのみで終わる低難易度のものと、人工呼吸や胸骨圧迫が必要な高難易度のシナリオの二種とした。その後、インタビューを実施し、インタビュー内容の主題分析を行った。

5.2.2 実験結果

主題分析の結果は以下の通りである（表 1）。以下、主題に繋がった個別のコードについての詳細を述べる。

生徒の自己判断の想起

従来方式では、情報の提示が、口頭での説明となっており、生徒に対する適切なフィードバックが欠けており、生徒は行動の結果、取得した情報に基づく判断が困難であった。例えば、心拍数を机を叩いて再現した場合、机を叩くタイミングで心拍数を推測できるので、生徒の主体的な判断が行われない。一方で、提案方式では、生徒は自分自身の聴診行動の結果、実際の新生児と同様の心音を聴き、情報に基づく判断が可能になり、自己判断の想起に繋がる。

生徒・講師のモチベーションの向上

従来方式では、バイタルの再現で机を叩き心拍数を再現する、講師が、泣き声の再現のために泣き真似をするなど、現実とはかけ離れた再現手法を用いておりリアリティの低下だけでなく、生徒・講師のモチベーションの低下に繋がっていた。しかし、提案方式で現実に近い再現が可能となり、生徒・講師のモチベーションが向上したと考えられた。

講師のタスク減少

講師から提案方式では、従来方式と比べ「頭の仕事量が

表 1：主題分析の結果

主題	コード
リアリティの向上	生徒の自己判断の想起
	生徒・講師のモチベーション向上
マルチタスクの支援	講師のタスク減少
	シナリオの複雑化への寄与

減っている」との発言を得た。このことから、提案方式では、バイタルの再現をシミュレータが代行したため、シナリオ実習中に行うタスクが減ったと考えられた。

シナリオ複雑化への寄与

提案方式では、従来方式と比して、シナリオを複雑化できたことが、講師・生徒双方の発言から認められた。複雑なシナリオの場合、再現すべき状況や、確認事項も多く、実施が難しい。シナリオの複雑化ができたのは、シミュレータがマルチタスク支援したことにより、生まれた余裕をシナリオの複雑化に割くができたためと考えられた。

6. 解決策 2：VAD 教材自動生成システム

本章では、デブリーフィングにおける課題「VAD の運用コスト」を解決する VAD 教材自動生成システムの設計、効果検証のために行った実験について述べる[6]。

6.1 VAD 教材自動生成システムの設計

NCPR ガイドライン 2020 では VAD が推奨されている[1]。しかし、導入・実施コストの課題から、VAD の実施は困難である。そこで、本研究ではプログラムにより VAD 教材 (図 7) を自動生成することで、課題の解決を目指した。

6.1.1 手技の認識

NCPR において、頻出する 4 種類の基本手技 (気管挿管・人工呼吸・聴診・胸骨圧迫) (図 8) を対象に画像認識を行った。具体的には、Google Inc の提供する Google AutoML Video Intelligence (以下 GVI) [18] のショット分析機能を用いた。学習に用いたデータセットは、実験者自身が 300 時間分の NCPR 講習の動画から作成した。データセットの内訳は、各カテゴリで 100 本以上の動画データである。

6.1.2 教材の自動生成

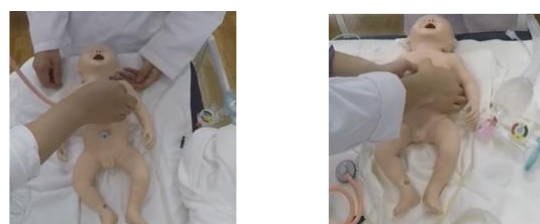
VAD 教材自動生成システムは、生徒の行う処置を GVI によって認識し、その結果とカメラ映像にマルチメディア処理を加え、動画・テキストサイトに投稿し、教材を自動生成する (図 9)。これにより、教材作成にかかる時間を短縮する。生成された教材は、動画内で実施された手技の継続部分が ■ によって可視化され、■ をクリックすると、手技の開始時点から動画を再生できるジャンプ機能を有する。

6.2 VAD 教材自動生成システムの検証実験

VAD 教材自動生成システムの効果検証として、医療関係者を対象として検証実験を行った。以下、その詳細を述べる。実験対象者は、NCPR インストラクター (NCPR に関する技術・知識を有し、NCPR 講習の開催資格を持つ) か



図 7：VAD 教材自動生成システム



(a) 聴診

(b) 胸骨圧迫

図 8：認識対象とする手技の一部

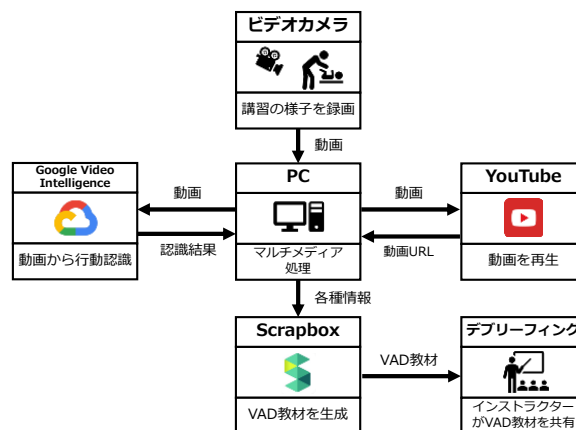


図 9：VAD 教材自動生成システムの構成

つ、勤続年数 10 年以上の医療関係者 10 名である。検証実験は、VAD 教材自動生成システムによって作成した教材を使用した後、アンケートを行い、その結果を分析した。

6.2.1 評価実験 1：画像認識の精度検証

画像認識の精度検証のために、客観評価とユーザーによる主観評価実験を行った。具体的には、客観評価では 4 種類の基本手技が出現する動画に対して処置の認識を行い、比較検証を実施した。主観評価では、四段階の選択肢応答にて手技の開始シーンのタイミングが講師の認識と一致しているかを検証した。客観評価の結果、各手技の再現率・適合率はそれぞれ 80%以上を示した。また、ユーザによる主観評価の結果、肯定的な反応が 80%を超え (表 2)、客観評価・主観評価の結果からも、画像認識の精度が講師の認識と相違なく、実用に足る精度であることが分かる。

表 2：画像認識の主観評価

手技の開始シーンのタイミングはあなたの認識と一致していましたか？	選択数 (n=10)
一致していた	3
ある程度一致していた	5
あまり一致していなかった	2
まったく一致していなかった	0

6.2.2 評価実験 2：VAD 教材自動生成システムの学習効果

学習効果を検証するために、選択式応答にて評価を行った。10 名中 9 名の回答者から肯定的な評価が認められ、否定的な評価は認められなかった。評価の多かった項目として、「指摘したいシーンを探す・振り返りにかかる時間を短縮」、「生徒に対して指摘をしやすくなる」などの評価が見られた。このことから、NCPR 講習に対し、システムの悪影響は薄く、デブリーフィングの時間短縮・指摘を行いやすくなる等の学習効果があると考えられた。

7. 解決策 3：講習内容の要約機能

本章では、デブリーフィングにおける課題「状況把握の難しさ」を解決する講習内容の要約機能の設計、効果検証のために行った実験について述べる。VAD 教材自動生成システムを用いたとしても、手技の開始から終了までのタイミングが分かるだけであり、シナリオ実習における生徒の行動の流れ（どの用に状況を判断し、行動したのか）について理解することは難しい。そこで、本研究では、専門用語を含む会話の内容を書き下し、その結果を、5.1 節の画像認識の結果を合わせ、ChatGPT を用いて要約することにより、状況の理解の補助を目指した。以下にその詳細を示す。

7.1 専門用語を含んだ音声の認識

シナリオ実習中の講師・生徒の発話内容を書き下すために、音声認識技術 Whisper[18]を利用した。シナリオ実習中の生徒・講師の発話内容には、専門用語が多く含まれる。しかしながら、音声認識 API は学習データの関係から、専門用語の認識率が低いという問題点を持っている。このため、既存の音声認識 API を使用するだけでは、発話内容を正確に書き下すことは困難である。この問題の解決には、追加の学習が有効とされる。一方で、複数の発話者から音声データを収集して、学習用のデータセットを作成することは長い時間と手間が必要となる。そこで、本研究では、専門用語の認識率向上のために、実習において頻繁に発言される専門用語のリストを作成し、リストから音声データを生成し、生成したデータを用いて Whisper のファインチューニングを行った。これにより、データの収集と前処理の段階での作業負担を大幅に削減した (図 10)。

7.1.1 専門用語音声データセット構築とモデル最適化

本節では、専門用語を含んだ音声を認識するための、専門用語リストの作成方法、リストを用いた学習用データセットの作成方法、モデルの最適化手法について述べる。

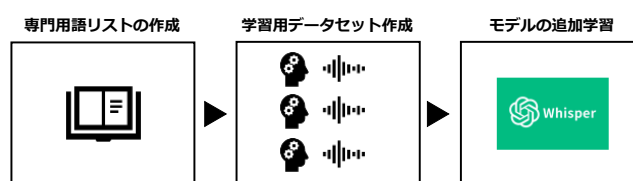


図 10：学習の概念図

専門用語リストの作成

学習用データセットを作成する過程で、まず NCPR インストラクター教本を参照し、頻繁に出現する単語を抽出、それらの単語の読みと書きを組み合わせました。これにより、最終的に約 500 種類の専門用語を含むリストを作成した。

学習用音声データセットの作成

Whisper の追加学習において、特定の音声データとその読みが必須となる。この要件を満たすために、Google Text to Speech[19]を使用し、先に作成した専門用語のリストを元に音声データを作成した。さらに、データの多様性を確保するために、4 種類の異なる日本語話者モデルを使用した。これにより、500 種類の専門用語それぞれに、4 人分の音声データを作成し、合計 2000 の学習データを得た。

モデルの追加学習

得られた学習用データセットを用いて、Whisper medium モデルのファインチューニングを行った。この時の学習に使用した主なパラメータを以下に示す (表 3)。

表 3：学習パラメータ

Batch_size = 16	: バッチサイズ
Learning_rate = 1.0×10^{-5}	: 学習率
Warmup_steps = 500	: ウォームアップステップ数
Max_steps = 4000	: マックスステップ数
Evaluation_strategy = steps	: 評価戦略
Generation_max_length = 225	: 生成最大長
Metric_for_best_model = WER	: 最適モデル指標

7.1.2 書き下し機能の検証実験

書き下しシステムの認識精度を評価するために、検証データを用いた客観評価を行った。検証データには、NCPR 講習におけるシナリオ実習を撮影した 3 分間の動画を用いた。評価指標として、音声認識の結果を参照テキストとの間の文字レベルでの誤差率を計測する評価指標である CER (Character Error Rate)、単語レベルでの誤差率を計測する WER (Word Error Rate) の 2 つを用いた。追加学習を行わない Whisper medium モデルと、追加学習を行った Whisper medium モデルの認識結果の CER と WER を検証した。

7.1.3 書き下し機能の検証実験結果

追加学習を施した後の専門用語を含んだ会話の誤差率について、CER, WER ともに 0.2 ポイントの改善が確認された (表 4)。また、追加学習により、追加学習前では認識できなかった専門用語 (羊水混濁など) が、追加学習後に正確に認識された (表 5)。このことから、本アプローチによって、データ収集と前処理の段階での作業負担を削減しながら、専門用語の認識率が向上できることが示唆された。

表 4：専門用語を含んだ会話の誤差率

	CER	WER
追加学習なし	0.37	0.49
追加学習あり	0.11	0.13

表 5：書き下し結果の比較（抜粋）

正解	追加学習なし	追加学習あり
羊水混濁	陽性混濁	羊水混濁
聴診器	調子に気持ち	聴診器
帝王切開	定予設改	帝王切開

じゃ、今も着々やってもらってますけど物の準備は大丈夫ですかね？
マスクの動作確認をまずやってみてもらって大丈夫ですか。マスク OK、吸引 OK
聴診器も OK
じゃあ大丈夫ですか？
赤ちゃんはあんまり推定変わらないです。
39 週、推定 2700 グラムくらいで
指定の帝王切開で赤ちゃんが生まれます。
清潔野の情報では羊水混濁はないということです。

図 11：正解データ

じゃあ、今も、**シャクシャク**やってもらってますけど、**もの**の準備は大丈夫ですかね？
マスクの**シヨップ**、**ご正学**にもね、まずやってみてもらって。
大丈夫ですか**ね**?マスクも OK、**QA** も OK、
調子に気持ちも OK。
じゃあ大丈夫ですか？
赤ちゃんは、あんまり**ついて買わない**ですけど、
39 週、**すいてい** 2700 グラムくらいで、
経営の定予設改で赤ちゃんで生まれます。
清潔やからの情報では、**陽性混濁**はないということです。

図 12：追加学習前の書き下し結果

じゃ、今も、着々やってもらってますけど物の準備は大丈夫ですかね？
マスクの**汚染**をまずやってみてもらって大丈夫です。マスク OK、吸引 OK、
聴診器も OK。
じゃあ大丈夫ですか？
赤ちゃんは、あんまり**ついて買わない**です。
39 週、推定 2700 グラムくらいで
指定の帝王切開で赤ちゃんが生まれます。
清潔薬の情報では羊水混濁はないということです。

図 13：追加学習後の書き下し結果

正解データと追加学習前後の書き下し結果を図 11~13 に示す。なお、認識に差が出た部分を太字にて示す。専門用語の認識に関して、精度の向上を確認できたものの、一部の専門用語についての誤認識が生じていた。この誤認識の主な原因は、学習に用いた Google Text to Speech で作成された音声データの発音の違いと考えられる。具体的には、作成した音声データの発音が、実際の発音とは異なるため、適切に特徴量を学習できなかったためと考えられた。また、専門用語リストに未収録の専門用語に関しても、適切に認識することが出来なかった。この問題は辞書内の用語を拡

充することで改善可能と考えられる。

さらに、専門用語に特化した学習を行った結果として、一般的な用語に関する認識精度に若干の低下が確認された。これは、学習モデルが新しい知識を習得する過程で、元々持っていた一般的な知識の一部を忘れてしまう、いわゆる「忘却」が生じたためと推測された。

7.2 講習内容の要約機能

VAD 教材自動生成システムを用いたとしても、手技の開始から終了までのタイミングが分かるだけであり、シナリオ実習における生徒の行動の流れ(どの用に状況を判断し、行動したのか)について理解することは難しい。そこで、本研究では、5.1 節の画像認識の結果及び、5.2 節の発言内容を ChatGPT を用いて要約することにより、状況の理解の補助を目指した。以下にその詳細を示す。

7.2.1 実装

発言内容の書き起こしは、タイムスタンプ付きの字幕形式で出力される。そこで、GVI の認識結果のタイムスタンプを合わせることで、シナリオ実習中の発言と手技を時系列に追えるテキストデータを構築した。このデータを基に、発言内容の要約を行うことを試みた。

ChatGPT 使用にあたって、入力文章の長さに制約が存在するため、効果的な要約を生成する工夫が必要であった。入力トークンを抑え、要約の内容が発散することを防ぎ、要約の精度を向上させる目的で、テキストを特定のタイミングで複数のセグメントに分割し、ChatGPT の入力として使用した。具体的には、新生児の出生、生徒の聴診のタイミングで分割を行い、これらのセグメントを順次入力することによって、テキストデータの要約を行った (図 14)。

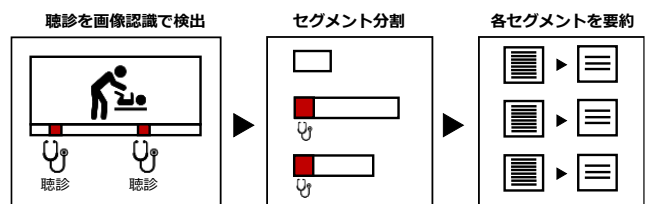


図 14：要約方法について

7.2.2 要約機能の検証実験

要約システムの精度検証を行うために、ユーザ評価を行った。具体的には、検証データとして、NCPR 講習におけるシナリオ実習を撮影した 5 分間の動画から、要約文を生成した。その後、要約結果と対応する動画をユーザに提示し、要約に関する 5 段階の選択肢応答によるアンケートを実施した。アンケートの評価項目は以下の 4 点とした。

- ・詳細性：要約が主要なポイントを網羅しているか
- ・冗長性：要約の内容が冗長かどうか
- ・正確性：要約が動画の内容を正確に反映しているか
- ・利用性：要約を通して内容が理解できたか

これらの項目を通じて、要約機能の性能を評価することを目指した。被験者は、非医療従事者 10 名、勤続年数 5 年以上の医療関係者 3 名である。なお、非医療従事者には、要約対象動画を閲覧する前に、NCPR の基本知識を習得する目的で、約 7 分間の説明動画を閲覧させた[20].

7.2.3 要約機能の検証実験結果

以下、要約プロンプトと出力結果（図 15,16）を示し、要約機能の検証実験の結果について述べる。

あなたはプロの新生児科の医師です。以下の内容は、新生児蘇生講習におけるシミュレータ実習の内容になります。あなたは、会話文と画像認識の結果（生徒の行動が示された文章）を受け取り、その内容を要約して下さい。

なお、要約に関しては、5 行程度にまとめ、生徒がどのような手技を行っているのか、判断を行ったのかについて、文意が損なわれないようにしながら、時系列順に要約を行い、後から見た時に状況を把握しやすいような内容にして下さい。

なお、与えられる内容は

- ・会話文

[00:00.00 ->11:00.00]A が起こっている

- ・画像認識の結果

(聴診)

といった形式になっています。

以下入力

(聴診)

[03:27 -> 03:29] そんなに泣いていない。

[03:29 -> 03:31] お擦ったりした。

[03:31 -> 03:33] 心を切ってますね。

[03:37 -> 03:39] ゆっくり。

[03:39 -> 03:41] 心拍 80 から 100 くらいですか？

[03:41 -> 03:45] 100 っていないので。

[03:46 -> 04:04] 今弱い呼吸がみられたので、

[00:00 -> 00:00] 筋緊張についてはちょっと分かりませんが、

[00:00 -> 00:00] これを認めたので気道開通。

[00:00 -> 00:00] 胎便除去もいましたし、

[00:00 -> 00:00] ホームをしてるのにこうなってるから。

[04:05 -> 04:13] 無理だから呼吸深迫を確認で、

[00:00 -> 00:00] SpO2 モニタとチェネズもお願いします。

(人工呼吸)

[04:27 -> 04:30] 人工呼吸していきます。

[04:35 -> 04:37] ありがとうございます。

[04:43 -> 04:47] 呼吸が入ってるか、

[00:00 -> 00:00] 他の確認方法をお勧めしましょうか。

[04:49 -> 04:52] 心音の聴診をお願いします。

図 15：プロンプトおよび入力文

- 1.赤ちゃんの泣き声が弱いことを確認。
- 2.赤ちゃんの体を擦って刺激を与える。
- 3.心拍数が 100 を下回っていることを確認。
- 4.弱い呼吸と筋緊張の低下が認められ、気道を確保して胎便も除去。呼吸の確認とサチュレーションモニタ、チェネズの使用が要請され、人工呼吸を開始。
- 5.呼吸の確認後、心音の聴診が行われる。

図 16：出力結果

詳細性（図 17）については、大多数の被験者（13 名中 11 名）が肯定的な評価をしており、3 名中 2 名の医療従事者が「やや網羅している」と評価している。これにより、要

約機能がコンテンツの主要なポイントを網羅する能力が高いことが分かる。一方で、網羅性に関して不十分と感じた被験者もいたため、改善の余地があることが分かる。

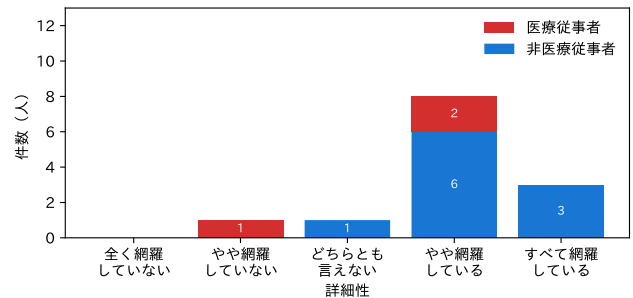


図 17：要約結果の詳細性

冗長性（図 18）については、大多数の被験者（13 名中 11 名）と、肯定的な評価をしており、その内 3 名中 2 名の医療従事者も肯定的な評価をしている。また、否定的な評価は、非医療従事者・医療従事者双方から認められなかった。これにより、要約内容の文章量が適切であることが分かる。

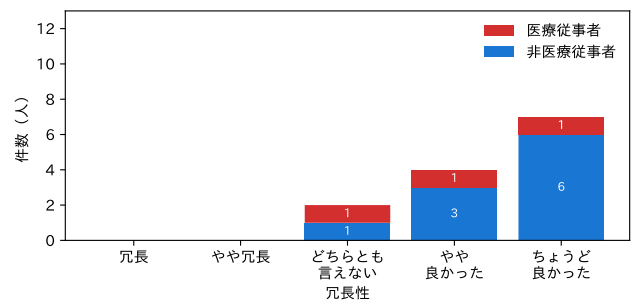


図 18：要約結果の冗長性

正確性（図 19）については、非医療従事者からは否定的な評価は認められなかった。医療従事者は 3 名中 2 名が「やや網羅している」と回答しており、比較的高い評価が得られていることが分かる。一方で、医療従事者 1 名から「やや異なる」という評価も確認された。

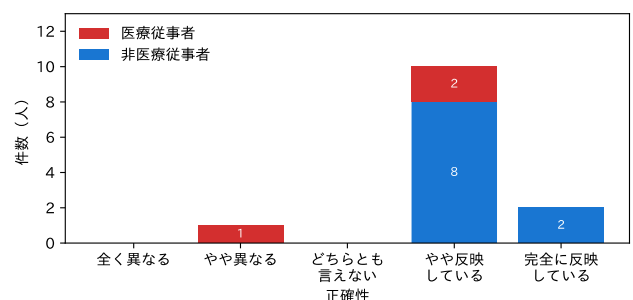


図 19：要約結果の正確性

利用性（図 20）については、大多数の非医療従事者（10 名中 9 名）が肯定的な評価をしている。一方で、3 名中 3 名の医療従事者が「どちらとも言えない」と評価していることから、要約の利用性について、課題があることが分かる。

また、詳細性・冗長性に対して否定的な評価をした医療従事者から、自由記述の回答の中で、「この要約内容は医学上ありえない内容であり、誤りである」との医療従事者からの指摘がみられた。これは、ChatGPT が会話内容を要約

する過程で、文脈から自動的に内容を補完し、その結果として誤った情報を生成してしまう、いわゆる「ハルシネーション」が発生したためだと考えられた。

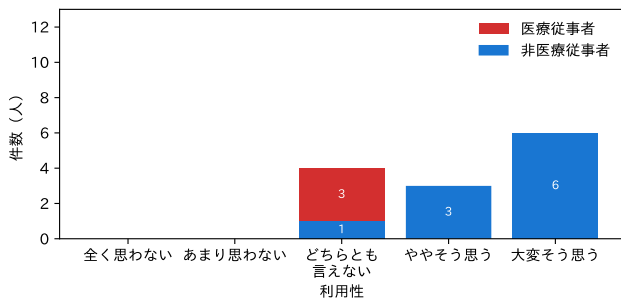


図 20：要約結果の利用性

ユーザ評価の結果、詳細性・冗長性・正確性について、肯定的な評価を受け、医療従事者と非医療従事者で大きな違いは認められなかった。一方で、利用性については、医療従事者全員が「どちらとも言えない」と評価している。この背景には、誤字脱字に起因する要約の誤りと、ChatGPT が持つ医療知識の限界により、要約の時に細かいニュアンスが失われていることが原因と推測された。

8. 考察

本研究では、「すべての分娩に新生児蘇生法を習得した医療スタッフが新生児の担当者として立ち会うことができる体制」の確立を目標に、シナリオ実習を支援する学習効果が高く、導入コストの低いシミュレータを開発、振り返りを支援するために振り返り教材自動生成システム、状況認識のさらなる支援のために音声書き下し、画像認識結果、音声認識結果から状況の要約機能の開発を試みた。以下の節にて、その詳細を述べる。

8.1 安価で効果的なシミュレータ

シナリオ実習に関しては、一般的に普及している新生児モデルを、電子部品・スマホアプリからなる模擬医療機器で拡張することによって、導入コストを抑えた。主題分析の結果から、シミュレータの再現機構によってリアリティが向上した結果、生徒の自己判断の想起、講師・生徒のモチベーションの向上できたと考えられた。また、講師のマルチタスクを支援することができた結果、講師のタスクの減少、シナリオの複雑化の寄与などの効果が見られ、学習効果の向上が考察された。

また、対応プラットフォームの拡大により、利用範囲を拡大。近接センサを用いて聴診判定を行うことで、低照度環境下での誤動作を防止した。さらに、ハブ端末開発により、遠隔講習での使用を可能とした。

8.2 VAD 教材自動生成システム

生徒の行動の自動認識に関しては、85%以上の精度で認識可能であり、アンケートの結果からも、その精度は NCPR の知識を有する講師の認識と遜色ないことが分かる。また、一連の処理はプログラム上で行われ、撮影した動画から教

材の自動生成が可能である。さらに、本システムを使用することにより、指摘したいシーンを探す時間を短縮、生徒に対して指摘をしやすくなるなどの効果が認められた。これは、システムにより手技の種類・タイミングを可視化されたためと考えられた。

8.3 講習内容の要約機能

8.3.1 音声認識

今現在の音声認識精度は、機械音声にて作成した人工音声を元に追加学習を行っている。このため、人間の持つ自然な発音のニュアンスを完全に捉えることはできていない。しかしながら、本アプローチにて、追加学習なしと比べ、CER, WER とともに精度の向上が見られている。今後の研究では、実際の人間の発言した音声データを学習データとすることによって、さらなる認識精度の向上が期待される。なお、一般用語の認識精度について若干の低下が見られたが、こちらについては、より大きなモデルの使用、デコーダ層の凍結し、最終層のみの学習することによって、一般用語の認識率を維持しながら、専門用語の認識率を向上させることが可能と考えられる。

8.3.2 要約

本研究では、音声認識と画像認識の結果を組み合わせ ChatGPT に入力することで、医師の持つ要約能力に近い結果を得ることができた。ChatGPT の利用に際して、要約精度の向上のための追加学習を行っていないにも関わらず、その初期状態でも十分に適用可能であったことが確認された。今後、さらなる会話データを学習させることによって、要約精度の向上が見込まれる。

一方で、音声認識のエラーおよび、ChatGPT が持つ医療知識の限界による要約の不備が確認された。これらの問題は、音声認識の精度向上、ChatGPT に医療知識をファインチューニングさせることで、解消されると考えられる。

現状は、視覚情報を画像認識の結果、聴覚情報を音声認識の結果から取得し、要約を行っているが、新生児のバイタル情報をシミュレータの操作履歴から入力することで、より高度な状況の把握が可能と考えられる。今後の研究にて、これらの機能を統合した要約機能の実装を目指す。

8.4 運用の観点からの議論

本研究は、現場の産院、病院の状況を鑑み、普及している低機能モデルに対して手を加えることなく、スマートフォン・電子部品を用いた模擬医療機器、およびネットワークサービスを使用した VAD 教材自動生成システム、要約機能からなる。置き換えが不可能な物理デバイスの使用を最小限に抑え、破損・劣化時に容易に交換可能なデバイスの使用により、維持・メンテナンス費用を低減させた。

シミュレータの利用により、シナリオ実習のリアリティ向上、マルチタスクの支援が可能になる。これにより、講師は、NCPR 講習運営の負担を軽減しながら、質の高い医療教育を提供でき、生徒は、より良い学習機会を得られる

と考えられる。なお、シミュレータの通信は数文字の平文を通信するのみで、携帯回線でも問題なく使用できる。

教材自動生成システムの利用により、講師は、デブリーフィング時に指摘したいシーンを迅速に探し出せるため、生徒への指摘が容易になる。これにより、NCPR 講習運営の負担を軽減できる。また、生徒は、自身の行動をビデオによって客観視することで、自己省察が可能になり、効果的なデブリーフィングを実施できる。ただし、現状のシステムでは、5 分間の動画から教材を生成に 10 分程度の時間を要する。実際の講習に用いるには、1 分程度の時間で教材を生成することが望ましく、実行速度に課題を残す。

要約機能の利用により、講師・生徒にとって、デブリーフィングにおける状況理解の効果が期待される。しかしながら、ChatGPT の医学知識の限界による要約内容の誤りが医療従事者から指摘されている。このため、現段階での使用においては、講師による要約内容のチェックと修正が必要であり、将来的には、さらなる追加学習が必要である。

これらの研究アプローチは、大きな予算を割くことの難しい国内外の医療教育において、有効な方法論となると考えられる。さらに、本システムの国際的な普及を通じて、世界の新生児の死亡率の低減が期待できる。

9. おわりに

本研究では、NCPR 講習の実施を情報技術によって支援し、NCPR の普及を助けることを目的とし、シナリオ実習を支援するために、安価で効果的なシミュレータを開発。デブリーフィングを支援するために振り返り教材自動生成システムを開発した。これにより、シナリオ実習におけるリアリティの向上、マルチタスクの支援、デブリーフィングにおいて指摘したいシーンを探す時間の短縮、生徒に対する指摘をしやすくなるなどの効果が認められた。

また、振り返りのさらなる支援のために、音声認識・画像認識の結果を利用して、シナリオ実習の流れを把握する要約システムを開発した。検証実験の結果、要約内容の詳細性・冗長性・正確性は、医療関係者から好意的に評価された。一方で、利用性については否定的な意見がみられた。これは、音声認識のエラーおよび、ChatGPT が持つ医療知識の限界による要約の誤りによるものと考えられた。これは医療知識の追加学習によって改善可能である。また、今後の研究にて、提案システム群を用いて、シナリオ実習からデブリーフィングを含む NCPR 講習を実施し、システム全体としての有効性評価を行うものとする。

本成果は、患者の状態を判断して人工呼吸・胸骨圧迫を実施する必要性があるなど、NCPR 講習と類似点が多く、NCPR 講習と同様に、低機能モデルを使用した講習が一般的であり、講習運営に際し、講師が多大な努力を有する成人心肺蘇生 (Cardio-Pulmonary Resuscitation : CPR) 講習においても応用可能と考えられる。

謝辞

本研究開発の一部は、総務省戦略的情報通信研究開発事業 SCOPE(受付番号 181607012)及び、電気通信普及財団の助成を受けたものである。また、本研究の趣旨に賛同していただき、協力していただいた医療従事者、研究室の皆様にご丁寧に感謝の意を表す。

参考文献

- [1] 日本版緊急蘇生ガイドライン 2020 に基づく NCPR 新生児蘇生法インストラクターマニュアル第 5 版
- [2] 阿部 幸恵：医療におけるシミュレーション教育，日本集中治療医学会雑誌 23 巻 1 号，(2016)
- [3] 田村 美子：小児看護学におけるケアリングを育むシミュレーション教育，看護・保険科学研究誌，(2013)
- [4] 阿部 幸恵：臨床実践力を育てる！看護のためのシミュレーション教育，医学書院，(2013)
- [5] 西本 騰：新生児蘇生訓練のための新生児人形を用いたシミュレータの開発，情報処理学会，論文誌デジタルプラクティス (TDP)，vol.3，28-38，(2022)
- [6] 西本 騰：NCPR 講習を支援するデブリーフィング教材自動生成システムの開発と評価，情報処理学会，論文誌デジタルプラクティス (TDP)，vol.4，(2023)
- [7] Leardal : Sim baby，<https://laerdal.com/jp/products/simulation-training/obstetrics-paediatrics/simbaby/>，(2023 年 10 月 12 日最終閲覧)
- [8] KOKEN : 新生児モデルアドバンス LM-111，<https://www.koken-educational-medical-models.com/products/668>，(2023 年 10 月 12 日最終閲覧)
- [9] Y.Takebe : Development of Neonatal Airway Management Simulator for Evaluation of Tracheal Intubation, In Proceeding of the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (2021)
- [10] Leardal : Newborn AnneTM，<https://laerdal.com/jp/products/simulation-training/obstetrics-paediatrics/newborn-anne/>，(2023 年 10 月 12 日最終閲覧)
- [11] KYOTO KAGAKU : デブリーフィング&データ管理システム Skills Album”ふりかえ朗”，https://www.kyotokagaku.com/jp/products_introduction/mw42/，(2023 年 10 月 12 日最終閲覧)
- [12] 堀 理江：看護基礎教育における高性能シミュレータを用いた心肺蘇生演習の学びと課題，ヒューマンケア研究会誌 第 4 巻 1 号，(2012)
- [13] 松田 直正：高性能乳児医療トレーニングシミュレータ“シムベビー”を活用した教育の可能性の検討，千葉大学看護学部紀要 第 32 号，(2010)
- [14] 河合 裕子：急変時対応シミュレーション教育におけるビデオを用いた振り返りの有用性，日本シミュレーション医療教育学会雑誌 第 7 巻 2019 年，(2019)
- [15] Skare : Implementation and effectiveness of a video-based debriefing programme for neonatal resuscitation, Acta Anaesthesiol Scand 2018，(2018)
- [16] 阿部 幸恵：臨床実践力を育てる！看護のためのシミュレーション教育，医学書院，(2013)
- [17] Ouyang, L. : Training language models to follow instructions with human feedback, (2022), <https://arxiv.org/pdf/2203.02155.pdf>
- [18] Radford, A. : Robust Speech Recognition via Large-Scale Weak Supervision, (2021), <https://cdn.openai.com/papers/whisper.pdf>
- [19] Google : Text-to-Speech AI, <https://cloud.google.com/text-to-speech?hl=ja>，(2023 年 10 月 12 日最終閲覧)
- [20] 【NCPR】新生児蘇生のイメージ 2021/3 公開，<https://www.youtube.com/watch?v=wgoLauCd6JM&t=863s>，(2023 年 10 月 12 日最終閲覧)