

新生児蘇生講習における効率的なデブリーフィングを支援する書き下しシステムの提案

西本 騰¹ 安藤 潤人¹ 野間 春生¹

受付日 2023年9月8日, 採録日 2024年2月16日

概要: 新生児蘇生法 (以下 NCPR : Neonatal Cardio-Pulmonary Resuscitation) 講習において, 講師はシナリオ実習やデブリーフィングを実施する. シナリオ実習中, 講師と生徒は会話を通じて情報共有を行っており, このときの講師・生徒の発言内容からは, 講師の指示や生徒の判断ミスなどの情報が得られる. これは, シナリオ実習後にデブリーフィングを行う際の重要な情報源となる. そこで, 本研究では, デブリーフィングの際の状況把握を支援するため, NCPR 講習中の発話内容を自動的に書き下し, 話者特定を行うシステムの開発を行った. 開発したシステムは, 事前学習なしでの話者特定, および専門用語の含まれた発話内容の書き下しを行う. また, 認識結果から字幕ファイルを作成し, 動画と合わせて YouTube にアップロードする. これにより, 動画中の全会話を確認し, 特定の会話の開始時点に直接移動できるジャンプ機能を有する動画コンテンツを自動生成する. この機能により, 動画の中から特定のシーンを効率的に探すことが可能になる. システムの効果検証として, NCPR 講習開催資格を有する講師に対してアンケートを用いた有効性評価を行い, 結果を分析した. その結果, 書き下し精度は, 実習にて活用できるレベルであり, 自動生成した動画コンテンツはデブリーフィングにおけるケースレビュー等に対して有効であることを示した.

キーワード: 新生児蘇生法, 書き下し, デブリーフィング

Development of a Transcription System to Support Efficient Debriefing in Neonatal Cardio-Pulmonary Resuscitation Workshop

NOBORU NISHIMOTO¹ MITSUHIITO ANDO¹ HARUO NOMA¹

Received: September 8, 2023, Accepted: February 16, 2024

Abstract: In Neonatal Cardio-Pulmonary Resuscitation (NCPR) workshop, the instructor conducts scenario training and debriefing. During scenario training, the instructor and students exchange insights. This exchange includes information such as the instructor's instructions and student's errors in decision. This is an essential source of information for later debriefing. Therefore, in this study, we aim to assist in understanding the situation. We have developed a system to identify speakers by transcribing their speech during scenario training. The system identifies speakers without prior learning and transcribes the speech content containing technical terms. This system converts the recognized speech into subtitle files and synchronizes them with videos on YouTube. This process automatically generates video content equipped with a jump function that allows viewers to review all conversations within the video and directly navigate to the start of specific dialogues. The feature enables efficient location of scenes within the video. We conducted a questionnaire survey of instructors qualified to hold NCPR courses to verify the system's effectiveness and analyzed the results. The results showed that the accuracy of the transcriptions was within an acceptable range for practical use and our proposed method was effective for debriefing.

Keywords: Neonatal Cardio-Pulmonary Resuscitation, transcription, debriefing

1. はじめに

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

¹ 立命館大学大阪いばらきキャンパス
Osaka Ibaraki Campus, Ritsumeikan University, Ibaraki, Osaka
567-8570, Japan

新生児蘇生法 (NCPR: Neonatal Cardio-Pulmonary Resuscitation) 普及事業を開始した [1]。NCPR は、出生直後の新生児が呼吸循環を適切に開始できない際に、呼吸循環を回復させるために行う人工呼吸や胸骨圧迫などの救命措置である。NCPR 講習は、医師や看護師がこれらの救命措置を習得するためのトレーニングプログラムであり、主に座学と実習、デブリーフィングから成り立っている。座学では基本的な知識を説明し、実習では新生児を模したシミュレータを使用した実習を行う。実習後のデブリーフィングでは、講師が生徒の知識・手技の間違いを指摘するなどのフィードバックを行う。デブリーフィングの概念は、もともとは軍事分野で用いられており、兵士が任務終了後に上官へ進行内容を報告し、上官からのフィードバックを受ける場面を指していた。現在、この概念は教育分野にも応用されている。医療教育におけるデブリーフィングは、臨床やシミュレーションベースのトレーニング後に行われる反省・評価のセッションである。この過程では、参加者が直面した状況や実施した行動について詳しく話し合い、判断や手技の誤りや成功を共有することで、より効果的な学びや意識的な実践への洞察を深める。

先行研究では、デブリーフィングを支援するために、筆者らは画像認識結果および、撮影した動画からデブリーフィング教材を自動生成する A-VAD (Advanced-Video Assisted Debrief) システムを開発した [2]。検証実験の結果、デブリーフィングに掛かる時間の短縮、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。一方で、A-VAD システムによる画像認識を用いても、生成されたビデオコンテンツから、生徒および講師の詳しい状況や判断の流れ (生徒がどのような判断を行っているか、指示を行っているかなど) を把握することはできないため、支援の余地が残っている。

本研究の目的は、NCPR 普及を情報技術によって支援することにある。そこで、本研究では、先行研究における残された課題である、生徒・講師の状況の把握を支援するために、生徒・講師の発言内容を書き下し、デブリーフィング時に閲覧可能なシステムを開発する。筆者らは、NCPR 講習特有の課題から、システムに求められる要求条件を設定し、その要求を満たす話者特定・書き下し機能を開発した。そして、効果検証のために行った実験、実験結果の考察について述べる。

2. 関連研究と課題

本章では、関連研究を例に上げながら、NCPR 講習におけるデブリーフィング、先行研究において解決できなかった課題について述べる。

2.1 NCPR 講習のデブリーフィング

NCPR 講習のシナリオ実習においては、講師は、講習に係わる情報 (在胎日数や母体の年齢、新生児の心拍数)などを主に口頭で提示し、生徒は、情報の共有 (聴診した心拍数の数値の共有など)、他の生徒に対する次の行動の指示などを口頭のコミュニケーションで行っている。また、シナリオ実習後のデブリーフィングにおいても、講師が生徒の手技の良かった点・悪かった点、知識の誤りや情報伝達の適切さなどを口頭にて説明し、講師と生徒は相互にコミュニケーションを行いながら、自身の知識や行動の修正を行う。そのため、シナリオ実習時の生徒・講師の発言内容には、生徒がいつ・どのような判断をしたのか、他の生徒とどのように情報共有をしたのか、講師の設定した状況 (新生児の在胎日数や、心拍数など) など、デブリーフィングにおいて重要な情報が詰まっている。このように、生徒・講師の発言内容は状況を理解するための重要な手がかりとなりうる。

NCPR におけるデブリーフィングの手法は、大別して、口頭のみ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つの手法が存在する。VAD には、VD と比較して、自己の客観視や学習に対するモチベーションの向上 [3]、手技技術の向上、自己省察の促進 [4]、蘇生ガイドラインの遵守率の改善 [5] などの利点が認められている。また、NCPR ガイドラインにおいても、デブリーフィング手法として、VAD が推奨されている。その一方で、VAD は広く実施はされておらず、多くの場合、紙のチェックシート (図 1) を用いたデブリーフィングが行われている。しかしながら、講習中にシミュレータ操作、生徒の観察などの複数のマルチタスクを行いながら、紙のチェックシートを記載することは講師にとっては負担が大きく、状況の記録が適切に行えていない現状がある。そのため、講師はシナリオ実習の際の記憶を頼りにデブリーフィングを行っている。

シナリオトレーニング ポイント表

見学者: _____
対象インストラクター: _____

シナリオトレーニングの指導法	ポイント	評価*
情報提示を適切に行う	適切なタイミングで情報を提示したか? 明確で適切な情報を提示したか?	
学習者の行動を適切に観察・評価する	学習者の行動を適切に観察したか? 学習者の行動を適切に評価したか?	
自己内省を促進する (VE)	Vent: 緊張を解く Examine your activity: 自己振り返り 学習者の緊張を解き、心理的な側面に配慮したか? チェックシートを使用し、個人による振り返りを促したか?	
議論を促進する (CT)	Case review: ケースレビュー システマチックに振り返りをリードしたか? 学習者に発言の機会を与え、促したか? 学習者の行動を批判的・建設的に指摘したか?	
次への行動を支援する (OR)	Talk each other: お互いに話し合う Organize next plan: 次の目標を立てる Retry: 再トライ 学習者のディスカッションを支援したか? 学習者の質問や要望に適切に対応したか? 目標をより具体的に構成し着すことができたか? 再トライ後に学習者の行動変更を確認したか? 学習者への動機づけが出来たか?	

* 評価は、努力が必要 △ 出来ている ○ 優秀 ○

メモ: _____

図 1 紙のチェックシート

Fig. 1 Paper-based checklist.

係上、生徒・講師の発話内容が理解できる結果となっていないなければならない。そのため、実際に講師や生徒の発言を書き下した結果から、状況を把握できることが望ましい。

4. 設計

前述の4つの要求条件を満たす機能の設計について述べる。開発した書き下し機能は、話者特定モデル、発言内容の書き下しモデル、医療用語の校正モデルの3つからなる(図3)。以下、その詳細について述べる。

4.1 開発環境・実行環境

開発言語は、Python (version3.9.0)、開発期間は約2ヶ月、開発規模は約500行である。話者特定には pyannote.audio [6]、書き下しには Whisper API [7]、医療用語の校正には ChatGPT [8] を使用している。実行環境は、Google Colaboratory、開発PCには、MacBook Air を使用した。なお、Google Colaboratory に関しては、有料サービス Colab Pro を利用し、プログラムの実行にはGPU (Tesla V100-SXM2-16GB) を使用した。Python 選定の理由は、求められる機能の実装に適したライブラリが提供されており、また、開発速度の向上ができるためである。Google Colaboratory 選定の理由は、物理的なGPUを積んだPCやワークステーションの導入と比較すると、低コストで環境構築ができること、将来的なプログラムの配布が容易なためである。

4.2 話者の特定・書き下し

本節では、事前学習なしの話者特定・書き下しを行うための具体的手法について述べる。なお、各モデルについて特別な事前学習は行っていない。これは、事前学習のためには大規模なデータセットが必要であるが、NCPR 講習用の音声データセットが公開されていないことが理由である。

4.2.1 pyannote.audio による話者の特定

話者の特定については、話者特定のための Python によるオープンソースフレームワークである pyannote.audio を用いた。選定の理由は、他の話者特定モデルと比較して精度が高く、無料で使用が可能だったためである。モデルについては公開されている事前学習モデル pyannote/speaker-diarization [9] を用いた。なお、NCPR 講習のシナリオ実習においては、講師1名と1~2人の生徒にて実施するのが一般的である。このため、話者の推定人数を示す事前パラメータを、min_speakers=2 (最小話者数2名)、max_speakers=3 (最大話者数3名) と設定し、精度向上に努めた。

4.2.2 Whisper API を用いた書き下し

書き下しについては、Open AI 社の提供する汎用音声認識モデルである Whisper API を用いた。選定の理由は、

他の音声認識APIと比較して、日本語に対応しており、書き下し精度が高く、雑音などのノイズに強いためである。モデルについては、実験段階のために精度を重視して、最も大きな large-V2 モデルを使用した。

4.3 医療用語の校正

本節では、前節の書き下し結果から、医療用語の誤変換を自動校正するための2つの手法について述べる。

4.3.1 誤り辞書を用いた既知のエラーに対する校正

Whisper API による書き下し結果には、再現性がみられた。そこで、本節では、誤り辞書による誤字脱字の構成手法、誤り辞書の作成方法について述べる。過去の講習動画を Whisper API にて認識を行い、誤字脱字・誤変換がみられた医療用語を抽出し、正誤表の辞書を作成した。なお、辞書の作成については、日本周産期・新生児医学会監修の NCPR インストラクター教本 [10] を参照し、作成した。また、誤り辞書の水増し手法として、Open AI 社の提供する ChatGPT を用いて、もともと100単語あった辞書の内容を三倍(300単語)に水増しした。以下、そのプロンプトについて示す(図4)。

4.3.2 ChatGPT を用いた未知のエラーに対する校正

前節での Whisper API を用いた書き下し文には、学習データの不足の原因と考えられる医療用語に対する誤変換・誤認識がみられ(例: K産物: 経産婦, 筋キンちゃん: 筋緊張), その種類は録音環境や滑舌などによって、多岐に分かれ、未知のエラーも多く存在する。そこで、本節では、ChatGPT を用いて、医療用語の誤変換・誤認識の修正を行った。なお、使用したモデルは GPT-4 である(図5)。

なお、生成系 AI は英語による命令によって精度が向上することが一般的に知られていること。日本語と比して使用するトークン数が少ないことから、プロンプトには主に英語を使用している。また、トークンの制限への対応および、出力の発散を防止するために、1会話を句点が入るまでと定義し、全体の会話を5会話ごとのセグメントに分割、これらのセグメントを順次入力することによって、会話文全体の校正を行った。

```

入力文:
whisperAPI error, correction. Suggest similar errors.
Format: error,correct
凶骨, 胸骨
出力結果:
鳳骨, 胸骨
熊骨, 胸骨
雄骨, 胸骨

```

図4 校正プロンプト

Fig. 4 Proofreading prompt.

4.4 教材の提示方法について

前節の書き下し結果を srt 形式の字幕ファイルとして保存し、動画とともにアップロード、YouTube の字幕機能を用いて、動画中の全会話を確認し、特定の会話の開始時点へ直接移動できるジャンプ機能を備えた VAD 教材を自動生成する。例として、「0:25 SPEAKER_01 情報では考えられるので、1つ目がアンビュバッグ」の部分をクリックすると、動画の0:25秒時点から動画を再生できる(図6)。この機能により、デブリーフィングの際に、動画内で生徒がどのような発言をしたのかなどの前後の文脈や発言を把握し、重要なシーンを素早く探し、アクセスが可

```
Scenario: JP med edu dialogue.
Role: experienced JP doctor.
  C1: Refer dict for typos.
  C2: Input in JP.
  C3: Maintain original text, correct typos based on dict.
  C4: input Format: 'Speaker dialogue'.
  C5: Process following input text.

Dict:
Format: 'Term: pronunciation', separated by ', '.
No pronunciation for Katakana terms.
Terms: Medical equipment,
Maternity and Newborn conditions,
Procedure terms, Katakana terms.

Medical equipment
心電図: しんでんず, 他
Maternity and Newborn conditions
帝王切開: ていおうせっかい, 他
Procedure terms
初期処置: しょきしょち, 他
Katakana terms
アブガースコア, 他

Input:
27歳のKさんぶさん、妊娠38週で赤ちゃん推定は3000gです。
もうお子さんが順調に進行して、もうすぐ生まれそうですという状況です。
あなたはBBキャッチで呼ばれました。
はい。
じゃあ、準備をお願いします。

Output:
27歳の経産婦さん、妊娠38週で赤ちゃん推定は3000gです。
もうお産が順調に進行して、もうすぐ生まれそうですという状況です。
あなたはベイビーキャッチで呼ばれました。
はい。
じゃあ、準備をお願いします。
```

図5 誤り辞書作成プロンプト

Fig. 5 Error dictionary creation prompt.

能になる。これにより、デブリーフィングの短時間で効率的な実施の支援を目指した。

YouTube 選定の理由は、字幕機能に対応しており、字幕のジャンプ機能を有すること、公式 API により投稿処理を自動化できることが理由である。話者特定に使用した動画は、講習の様子を把握しやすいように蘇生台に Go-Pro HERO 11 をマウントにて固定し、蘇生台を俯瞰する形式にて撮影を行った。なお、撮影時の解像度は 720p とした。以下、作成した教材(図6)、動画から話者特定、書き下し、校正処理、投稿処理までの一連の教材作成流れをシーケンス図(図7)にて示す。なお、NCPR 講習のシナリオ実習において一般的な講習時間である5分の講習動画1本(容量75MB)に対して、前述の実行環境において話者特定・書き下しを行った際の実行時間は1分程度であった。また、研究開発全体での Web サービスの使用料金は、Google Colaboratory, Whisper API, ChatGPT の総計で3,779円である。

5. システムの効果検証

本章では、システムの性能・効果検証のために行った評価実験の内容、その結果について述べる。

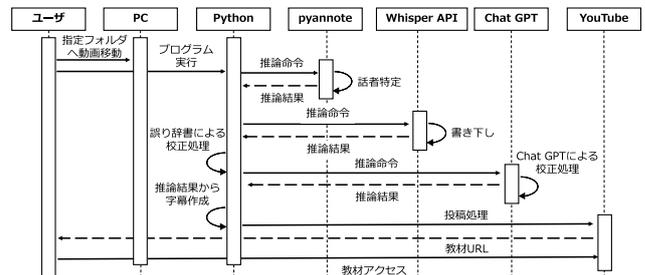


図7 シーケンス図

Fig. 7 Sequence diagram.

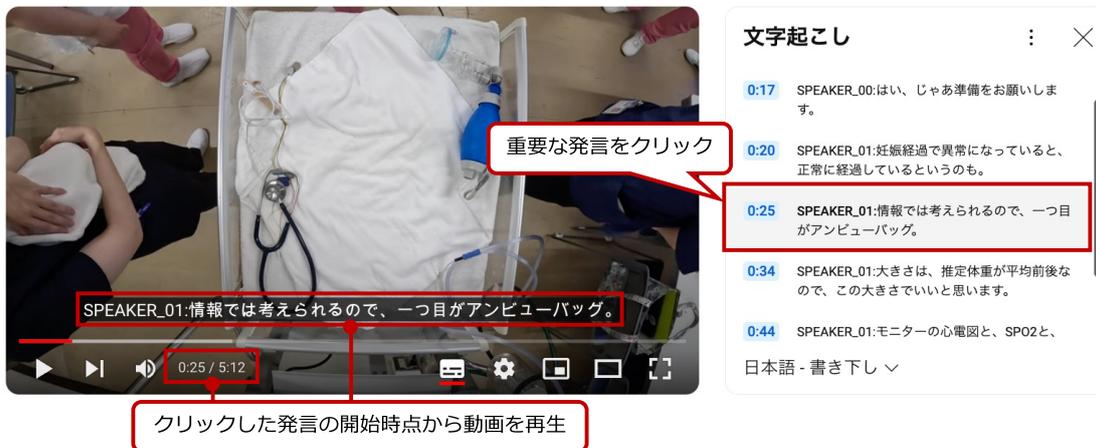


図6 システムにて自動生成した教材

Fig. 6 Materials automatically generated by the system.

5.1 評価実験1：認識精度の検証

書き下しシステムの精度を検証するために実験を行った。対象とする動画は、フレームレート 720p、音声のサンプリング周波数 48.0 kHz、ファイル容量 75 MB である。書き下しの対象とする動画は、話者 3 名（講師 1 名、生徒 2 名）の 5 分間の NCPR におけるシナリオ実習中の動画 3 本である。具体的な評価指標として、システムによる書き下し結果と、正解データとの間の話者一致率、発言内容の部分一致率（全会話中、どれだけの単語が正確に認識されたかを単語単位で測定した指標）、実行時間を比較し、検証を行った。

5.2 実験結果 1

実験の結果、話者の一致率は 58%、発言内容の部分一致率は 68% となった。これらの結果は、事前学習を行わずに得られたものである。なお、話者の一致率・発言内容の部分一致率の算出方法は以下のとおりである (1), (2)。

$$\text{話者の一致率} = \frac{\text{話者が正しく識別された発話数}}{\text{全会話数}} \quad (1)$$

$$\text{発言の部分一致率} = \frac{\text{正しく識別された単語の数}}{\text{テキストの単語総数}} \quad (2)$$

また、5 分間の動画データの書き下しに要した時間は 1 分であった。認識精度が低下する要因として、複数の話者が同時に喋る場面や単語の言い淀みや言い直しの部分が挙げられる。これらの場面で、認識精度が低下する傾向がみられた。

5.3 評価実験2：ユーザ評価

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）であり、かつ、勤続年数 10 年以上の医療関係者 9 名である。検証実験は感染症対策として、リモート・非対面にて行った。具体的には、システムによって生成した動画コンテンツを被験者に関覧させ、動画コンテンツを使用したデブリーフィングを想定して、アンケートへの回答をさせた。なお、認識精度については先の実験 1 の結果に準ずる。また、プライバシーの配慮のために、YouTube の公開設定は限定公開とし、動画コンテンツの共有を行った。

5.4 実験結果 2

評価実験の結果を以下で述べる。対象とした評価項目は、字幕の精度、各種機能に対する評価、システムを用いた際のデブリーフィングに対する効果である。

5.4.1 字幕の精度・理解について

五段階の選択式応答にて、字幕の内容が講師の認識と一致しているかを検証した。具体的には「字幕の内容は、あなたの認識と違いはありませんでしたか？」という質問に対して、最頻値は「やや異なる」が 9 名中 6 名となり、

「やや同じ」が 2 名、「全く同じ」が 1 名となった (図 8)。このことから、字幕の精度は講師の認識とは必ずしも一致していないことが分かる。

一方で、「誤変換・誤記が全体の理解に影響を及ぼすか」という質問に関しては、9 名中 7 名が「あまり影響はない」2 名が「少し影響がある」と回答している (図 9)。このことから、字幕の正確性が必ずしも視聴者の内容理解に影響を与えない可能性が考えられた。

また、「発言内容をどの程度理解できたか?」という質問に対しては、9 名中 7 名が「かなり理解できた」、2 名が「少し理解できた」と回答しており (図 10)、先の考察を支持する。さらに、自由記述の回答にて、「(字幕の精度は完全とは言えないが) 音、画像、言葉が合わさっているので、すべてを見ながら (状況を) 理解することができた」とのコメントを得た。このことから、動画と合わせて字幕を閲覧することにより、内容の理解が補完されたと考えられた。

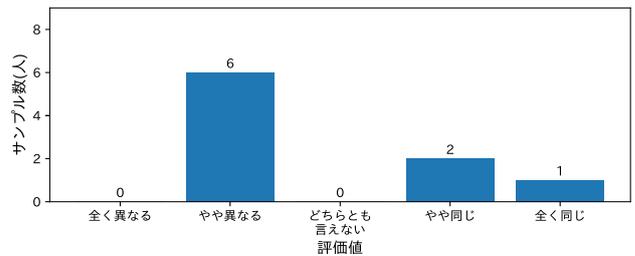


図 8 字幕の認識精度

Fig. 8 Subtitle recognition accuracy.

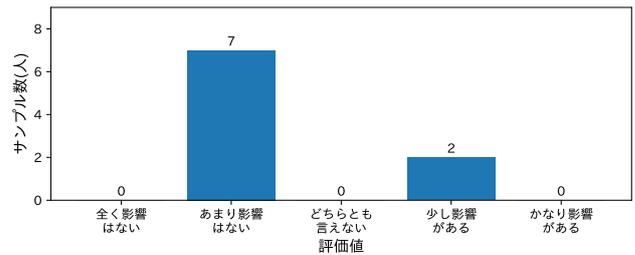


図 9 誤変換・誤記の影響

Fig. 9 Impact of misconversions & typos.

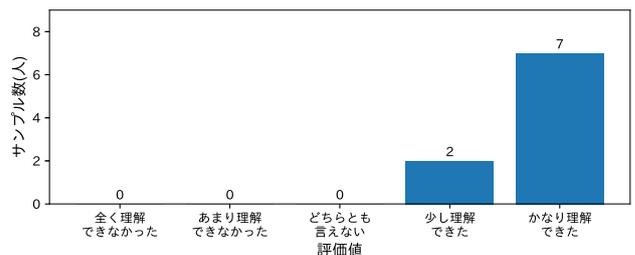


図 10 発言内容の理解

Fig. 10 Understanding of spoken content.

字幕の開始タイミングについては、9名中9名が「ちょうどいい」と回答した(図11)。このことから、字幕の提示タイミングと音声と適切に同期されており、講師の認識と相違がないことが分かる。

5.4.2 誤変換・誤記の分析、および校正手法の検討

誤変換・誤記については、自由記述の回答にて指摘を得たものを一部抜粋する(表1)。誤変換・誤記の傾向として、読みは合っているが、変換が不正確なものがみられた。これらは、誤り辞書を用いた校正によって、修正可能と考えられる。誤り辞書の拡充により、認識精度の向上が期待できる。ただし、誤変換・誤記がみられた単語について、機械的な校正処理を行った場合、本来正しい表現を不適切に修正してしまう可能性がある。しかしながら、この場合でも実運用上での影響は薄いと予想される。たとえば、本来「服」である変換を「拭く」として誤変換してしまうケースを想定した場合、生まれたばかりの新生児は服を着ておらず、新生児が服を着用するシチュエーションは稀であるため、発話の中で「服」発生する確率は低く、誤変換の発生確率も低いと考えられる。また、現状の認識精度でも、視聴者の内容理解に対する悪影響が薄いことは前述のとおりである。

一方で、さらなる精度向上のためには、文脈情報を理解したうえでの校正処理が必要であるが、本提案では、このために2種類の校正手法を用いている(表2)。誤り辞書による校正により、本来「服」である変換を「拭く」と誤変換してしまった場合でも、その後のChatGPTによる文脈を加味した校正処理を行うことで修正が可能である。

5.4.3 各種機能に対する評価

発話者の提示機能による効果を検証するために、選択式応答にて評価を行った(図12)。「発話者の提示機能は、

講習において生徒の理解を助けるか?」という質問に対して、9名中6名が「かなり助けになる」、3名が「少し助けになる」との回答を得た。すべての回答者が発話者の提示機能に対して肯定的な評価をしている。

ジャンプ機能による効果を検証するために、選択式応答にて評価を行った。「ジャンプ機能は、生徒の理解を助けるか?」という質問に対して、9名中5名が「かなり助けになる」、3名が「少し助けになる」、1名が「どちらとも言えない」との回答を得た。また、ジャンプ機能に対する否定的な評価は認められなかった。

自動生成した動画コンテンツによる効果を検証するために、選択式応答にて評価を行った。「作成した動画コンテンツは、生徒の理解を助けるか?」という質問に対して、9名中6名が「かなり助けになる」、2名が「少し助けになる」、1名が「どちらとも言えない」との回答を得た。否定的な評価は認められなかった。

5.4.4 デブリーフィングの改善

インストラクター教本における評価項目[10]を参考にし、アンケート項目を作成し、本システムにて自動生成した動画コンテンツを用いることによって、デブリーフィングにおける各実施項目に改善がみられるかについて評価実験を行った。以下その詳細について示す(図13)。

表2 誤りの修正課程 (太字にて、校正が行われた部分について示す)

Table 2 Correction process (highlighted in bold: sections where revisions were made).

元文章
赤ちゃんに 服 を着せてあげて下さい。 この状況から考えられる症例は 3種類 あります。
元文章を誤り辞書によって校正
赤ちゃんに 拭く を着せてあげて下さい。 この状況から考えられる症例は 3週 類あります。
誤り辞書による誤変換をChatGPTによって校正
赤ちゃんに 服 を着せてあげて下さい。 この状況から考えられる症例は 3種類 あります。

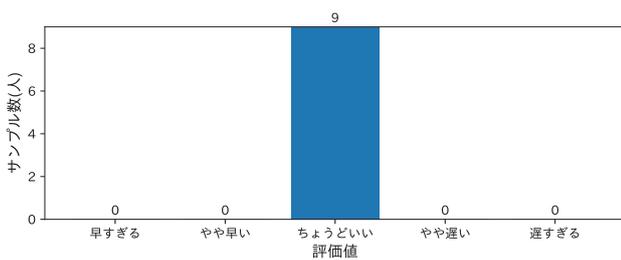


図11 字幕のタイミング

Fig. 11 Subtitle timing.

表1 誤変換・誤記の例

Table 1 Examples of misconversions & typos.

誤変換・誤記	正しい記述
種(しゅ)	週(しゅう)
本(ほん)	保温(ほおん)
服(ふく)	拭く(ふく)

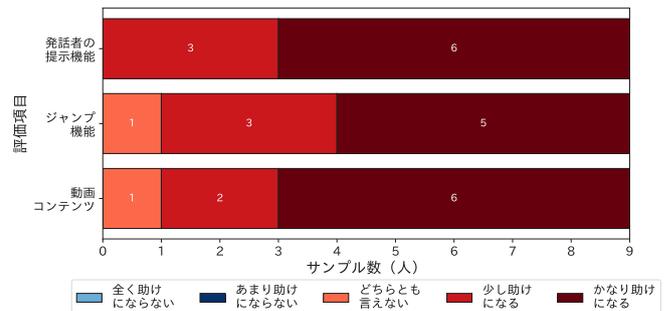


図12 各種機能に対する評価

Fig. 12 Evaluation of various functions.

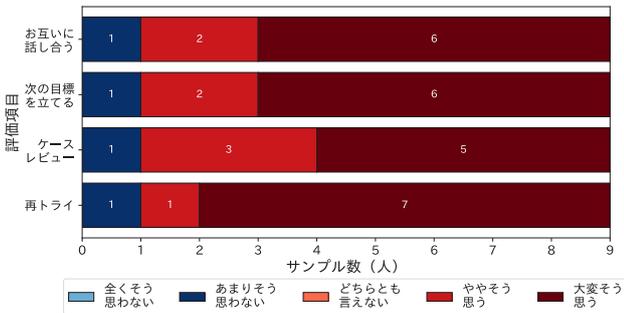


図 13 デブリーフィングの改善
Fig. 13 Debriefing improvements.

お互いに話し合う、次の目標を立てるについて、9名中6名が「大変そう思う」、2名が「ややそう思う」、1名が「あまり思わない」との回答を得た。ケースレビューについては、9名中5名が「大変そう思う」、3名が「ややそう思う」、1名が「あまりそう思わない」との回答を得た。再トライについては、9名中7名が「大変そう思う」、1名が「ややそう思う」、1名が「あまり思わない」との回答を得た。

デブリーフィングにおける各実施項目について、9名中8名の被験者が肯定的な評価をしたことから、本システムがデブリーフィングにおいて有効であると考えられた。なお、否定的な評価をした1名の被験者からは、「シナリオ中に会話がない場合もあるため、会話の内容だけではなく、実際の状況や行っていることを提示することも重要である」とのコメントを得た。

5.4.5 追加機能・改善案について

選択式のアンケートにて、システムの追加機能・改善案についての意見を得た(表3)。

「認識精度の向上」については、9名中9名が指摘している。このことから、5.1節にて、現状の字幕精度でも発言内容をかなり理解できることが分かっているが、さらなる認識向上が必要であることが分かる。

「発話者の役割の提示」については、具体的には、講師・生徒の役割を提示することである。自由記述の回答から、この機能の実装により、生徒自身が「良くできていたこと・自分に足りていなかった」ことをより確実に振り返ることができるのではないかと指摘がみられた。このことから、今後の検討項目として、役割の提示機能の実装のために、あらかじめ生徒・講師の声質を学習しておき、発話内容と役割を一致させるなどの手法が考えられる。

「発言内容の評価機能」については、具体的には、生徒の発言内容を分析し、生徒の間で情報共有がされているか・否かを評価する、生徒が医学知識に基づいて正しく状況を判断できているかどうかを評価する機能である。この機能の実装により、生徒が自身の行動を客観視しやすい可能性があるとの指摘がみられた。

表 3 追加機能・改善の要望

Table 3 Requests for additional features & improvements.

追加機能・改善案	件数 (人)
認識精度の向上	9
発話者の役割の提示	5
発言内容の評価機能	4
生徒の行う手技の提示機能	3

6. 考察

本研究では、先行研究にて開発した A-VAD システムの追加機能として、話者特定・書き下し機能の実装を行い、講師・生徒の状況把握の補助を目指した。以下、設定した4つの要求条件を満たしたか、システムを用いたことによるデブリーフィングへの効果、NCPR 講習以外の領域への応用可能性について考察する。

6.1 システムは要求条件を満たしたか

本節では、3章にて設定した4つの要求条件(実用上許容できる処理時間、複数の話者に対応した話者特定機能、医療用語への対応、実用上許容できる書き下し精度)を開発したシステムが満たしたかについて述べる。

要求条件 1: 実用上許容できる処理時間

一連の字幕生成にかかる実行時間は、一般的なシナリオ講習5分の動画に対して、約1分程度であり、十分に高速であると言える。しかしながら、一連の動画コンテンツを作成するには、5分の動画に対して10分程度かかるため、実運用にて用いるには課題を残す。このことから、本要求条件については一部達成したものとみなす。

要求条件 2: 複数の話者に対応した話者特定機能

pyannote.audio を用いた話者特定を行い、字幕にて共有を行うことにより、複数話者に対応した話者特定機能を実装した。検証として、客観評価および、9名の講師に対して選択式応答による主観評価を行い、客観評価では話者の一致率58%、主観評価ではすべての回答者から肯定的評価を得た。58%の一致率にもかかわらず、主観評価で肯定的な評価を得た理由には、2つの理由があると考えられる。第一に、ユーザが背景情報を持っていること。第二に、字幕と映像から内容を補完できたことが理由である。具体的には、本実験の被験者は、勤続年数10年以上のNCPRインストラクターであり、NCPRに関して豊富な知識を有する。このため、彼らは動画の中で生じ得る会話や、発言者の特徴や傾向について理解していると考えられる。また、このような背景情報を理解している人間が、字幕と映像を合わせて閲覧していることから、内容が補完され、話者の誤りを脳内で修正できたためと考えられた。このことから、話者特定機能が状況の理解に寄与できたこと

が分かり、要求条件を満たす。

要求条件 3：医療用語への対応

誤り辞書を用いた校正、ChatGPTを用いた校正により、医療用語への対応を目指した。検証として、「誤変換・誤記が内容の理解に影響を及ぼすか？」という質問に対して、9名中7名の講師から「あまり影響はない」との回答を得た。このことから、医療用語の字幕精度が実用上許容できる範囲内であることが分かり、要求条件を満たす。

要求条件 4：実用上許容できる書き下し精度

検証実験の結果、話者の一致率は58%、発言内容の部分一致率は68%であることが分かっている。また、選択式応答の結果より、字幕の内容は講師の認識とやや異なることが分かっている。一方で、内容理解に関する質問に関しては、すべての講師が肯定的な評価をしていることから、字幕と映像と合わせて閲覧することにより、内容が補完され、文意を取ることができたと考えられた。これらのことから、要求条件を満たす。

なお、話者特定・書き下しについては、事前学習により精度向上が期待される。しかしながら、本研究では事前学習を実施していない。これには2つの理由が存在する。第一に、NCPR講習を包括するような音声データセットがインターネット上に公開されていないこと。第二に、事前学習済みモデルが現場での使用に適さないことが挙げられる。具体的には、事前学習済みモデルを使用する場合、そのモデル格納・実行するための計算資源が必要である。ローカル環境での実行には、モデル格納のストレージ、実行のためのGPUの確保が課題になり、クラウド環境での実行には、ストレージおよびGPUの使用に追加費用がかかる。このため、コストが増大してしまい、教育に多くの予算を投じることのできない医療現場での導入に適さないと判断した。

6.2 システムを用いたデブリーフィングへの効果

発話者の提示機能、ジャンプ機能、自動生成した動画コンテンツに対する選択式応答からは、すべての機能がおおむね肯定的に評価された。特に発話者の提示機能については、否定的な評価がみられず、この機能が特に効果的である可能性があると考えられた。ジャンプ機能と動画コンテンツに関しても肯定的な評価が多くみられ、これらの機能がデブリーフィングにおいて有効である可能性が示された。一方で、「どちらとも言えない」との回答もみられ、使用状況や個々のニーズによって、改善の余地があると考えられた。

ケースレビュー・お互いに話し合う・次の目標を立てるといった項目においても、肯定的な評価が多く、デブリーフィングにおける本システムの有効性が示されたと言える。また、否定的な評価をした回答者からのフィードバ

クを考慮すると、講習の実施形式によっては、会話内容以外の情報提示が重要であることが示唆された。一方で、会話の不在が情報共有の欠如を示すのか、生徒間での高度なチームワークを示すのかの判断は具体的な状況分析に依存する。しかしながら、講師が動画と合わせて状況を閲覧することによって、状況分析は可能と考えられる。これらの結果より、生成した動画コンテンツが振り返りににおいて有効であることが分かる。

6.3 NCPR講習以外の領域への応用可能性

本プラクティスは、動画から事前学習なしに実用上許容できる精度で話者特定・書き下しを行うことが可能であり、各会話の起点となる会話の開始時点へのジャンプ機能を有する。これは、NCPR講習のみならず、発話による教育指導を行うケースにおいて応用が可能であると考えられる。たとえば、医療領域での応用として、NCPR講習とアルゴリズム上の類似点の多い、成人分野におけるCPR (Cardio-Pulmonary Resuscitation) 講習におけるデブリーフィングへの応用が考えられる。また、医療外の領域への応用として、デブリーフィング時に指導員と学習者が相互にフィードバックを行う自動車の安全運転指導などにも応用が可能と考えられる。類似の研究・製品として、株式会社京都科学の「ふりかえ朗」がある。このシステムは、シミュレータと連動して行動を記録し、行動の開始地点へのジャンプ機能を持つ。ただし、全機能を使用するには、同社のシミュレータとの連動が不可欠である。一方で、本システムはシミュレータに依存せず全機能が使用可能である。このため、本システムは拡張性に優れ、別フィールドにも応用が容易と考えられる。

7. おわりに

本研究の目的は、NCPR普及を情報技術によって支援することにある。そこで、本研究ではデブリーフィングをより円滑に進めるための書き下し機能を開発し、講習実施の支援を目指した。支援のために、筆者らは4つの要求条件（実用上許容できる処理時間、複数の話者に対応した話者特定機能、医療用語への対応、実用上許容できる書き下し精度）を設定した。そして、設定した4つの要求条件を満たす新機能を開発し、その効果を検証した。書き下し機能は、話者の分離モデル、発言内容の書き下しモデル、医療用語の校正モデルの3つからなる。医師9名に対するアンケートの結果から、書き下し結果を利用して作成した動画コンテンツは、デブリーフィングに対して有用であることが分かった。また、認識精度に関しては向上の要望が認められた。発話内容の書き下し、話者特定に関しては、インターネットに公開されているAPI、ソースコードを利用し、ユースケースに応じたカスタマイズを行うことにより、事前学習なしに実用に足る精度となることが分かって

いる。このことから、新生児医療領域のみならず、成人医療領域、医療分野以外の領域においても、本成果は応用可能と考えられる。

将来の展望として、処理時間のさらなる改善、書き下しの内容について自然言語処理を行い、状況の文脈的な判断、要約を行う機能の実装、および先行研究にて開発した画像認識機能と書き下し機能を統合することにより、視覚情報と聴覚情報を含めた状況判断、要約機能の実装を目指す。

謝辞 本研究の趣旨に賛同していただき、快く協力していただいた医療関係者の皆様に深く感謝いたします。

参考文献

- [1] 細野茂春：日本版救急蘇生ガイドライン 2020 に基づく第4版新生児蘇生法テキスト 2020, メディカルビュー社 (2021)
- [2] 西本 騰：NCPR 講習を支援するアプ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 (2023)
- [3] 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, 情報処理学会デジタルプラクティス (2022)
- [4] 田村美子：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)
- [5] Skare: Implementation and effectiveness of a video-based debriefing programme for neonatal resuscitation, *Acta Anaesthesiol Scand* 2018 (2018)
- [6] Bredin H: pyannote.audio: neural building blocks for speaker diarization, *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing* (2020)
- [7] Radford A: Robust Speech Recognition via Large-Scale Weak Supervision, *Proc. of the 40th International Conference on Machine Learning in Proceedings of Machine Learning Research*, 202: 28492-28518 Available from <https://proceedings.mlr.press/v202/radford23a.html> (2023)
- [8] Open AI: GPT-4 Technical Report (2023)
- [9] Bredin, H., and Laurent, A.: End-to-end speaker segmentation for overlap-aware resegmentation, *In Proc. Interspeech 2021*, Brno, Czech Republic (2021)
- [10] 細野茂春：日本版救急蘇生ガイドライン 2020 に基づく第5版インストラクターマニュアル 2020, メディカルビュー社 (2021)
- [11] 京都科学：デブリーフィングシステム“ふりかえ朗”, https://www.kyotokagaku.com/jp/products_introduction/mw42/ (2023年10月12日最終閲覧)



安藤 潤人 (非会員)

2020年筑波大学大学院システム情報工学研究科修了。2020年同大学研究員。2021年立命館大学助教。現在に至る。ロボティクス、感覚情報の研究に従事。博士 (工学)。



野間 春生 (非会員)

立命館大学情報理工学部教授。立命館先進研究アカデミーアソシエイトフェロー、日本VR学会理事。VR、HI、医療情報システムの研究に従事。博士 (工学)。



西本 騰 (学生会員)

2021年立命館大学情報理工学研究科修士課程修了。同大学大学院同研究科博士課程在学中。情報処理学会学生会員、次世代研究者挑戦的研究プログラム学生フェロー。