視聴覚情報の補償を考慮したプレゼンテーション動画生成アプリの開発

前土佐 勇仁[1,a)]　三枝 亮[2,b)]

概要：本研究では，プレゼンテーションスライドから視聴覚情報が補償された動画を生成するアプリケーションを提案する．生成される動画では，CG エージェントが音声ナレーションや動作表現を使用してプレゼンテーションスライドを説明し，音声と同期して字幕が表示される．このような動画生成は，既存のシステムでも見られる手法である．しかし，本システムは利用者層によって字幕や音声ナレーションの有無や内容の変更を可能とすることで，より広い利用者層に向けた視聴覚情報の補償を実現した．本システムを次脳機能障害者のグループ訓練の手引きに関するプレゼンテーション動画の生成に適用し，従事者に対して DVD として配布した．

キーワード：福祉情報工学, 動画生成, 視聴覚保障

Development of Presentation video Generation Application Considering Compensation of Audiovisual Information

Abstract: This paper proposes an application that generates a video with compensated audiovisual information from presentation slides. In the generated video, a CG agent explains the presentation slides using audio narration and motion expressions, and subtitles are displayed in synchronization with the audio. This type of video generation can be seen in existing systems. However, this system enables the user to change the presence and content of subtitles and voice narration depending on the user group, thereby compensating audiovisual information to a wider range of user groups. The system was applied to the generation of a video presentation of a guide for group training of people with hypencephaly, which was distributed as a DVD to the people engaged in the training.

Keywords: Welfare information technology, Video generation, Audiovisual Compensation

1. はじめに

昨今，障害者ピアサポートの取り組みが各自治体で進められている．障害者ピアサポートとは，自ら障害や疾病の経験を持ち，その経験を生かしながら，他の障害や疾病のある障害者のための支援を行うものである [1]．厚生労働省，この障害者ピアサポーター育成事業を補助金の供出を通じて推進している [2]．ピアサポート研修にあたっては，テキスト作成の他，講師による対面やウェブ会議システムなどを用いたリアルタイム形式の講演が含まれている．しかし，研修者の自己研磨を促すためには，テキストや講演内容をオンデマンド化し講演日前後に提供することが求められる．

講演内容をオンデマンド化する方法には，講演会場の録画する方法と講演内容を事前に録画する方法が考えられる．事前録画を手助けするツールには，プレゼンテーション資料から自動で字幕付きの動画を作成するシステム（スライド動画生成システム）がある．スライド動画生成システム自体は既に多くの製品例があり [3]，利用者の説明業務の負担軽減に貢献しており，ピアサポート研修においても，同様のスライド動画生成システムの利用が想定される．加えて，動画内のカメラワークを自動生成する研究 [4] [5] な

[1] 電気通信大学大学院
Chofugaoka 1-5-1, Chofu, Tokyo 182-8585, Japan
[2] 神奈川工科大学大学院
Shimoogino 1030, Atsugi, Kanagawa 243-0203, Japan
[a)] yuto.maetosa@syblab.org
[b)] ryo.saegusa@syblab.org

© 1959 Information Processing Society of Japan

ど，動画表現の拡張に関する研究も盛んに行われている．

さて，病気や交通事故などが原因で脳機能の一部に障害が残る病気の 1 つに，高次脳機能障害がある．高次脳機能障害の患者には，視覚情報や聴覚情報に課題がある当事者も含まれている．一方で，既存のスライド動画化システムでは，このような当事者の利用を想定した動画の生成が難しい．したがって，視聴覚情報に課題がある当事者に対してピアサポーターへの参画を促すためには，オンデマンド教材の作成においても視聴覚情報保障を考慮した情報提供支援ツールの開発が要求される．

本研究では，プレゼンテーション資料から情報補償動画を生成するアプリケーションを提案する．既存製品と比べ，本システムは情報補償が施されたビデオ教材を簡便に作成できる．これにより，障害当事者に対して，音声字幕付き動画という形で情報提供の機会提供が実現される．

2. 動画生成システム

本システムの概要図を Fig.1 に示す．主に，.NET 環境の GUI アプリケーションと Mono 環境の動画生成アプリケーションの 2 つで構成されている．User は GUI アプリケーションでは，文章の修正を行った後にナレーション音声や字幕内容を含んだ表ファイル，スライド画像，設定値を保存したファイルを作成する．動画生成アプリケーションでは，GUI アプリケーションで作成したデータを利用して録画環境を構築し，ウィンドウ内で操作することで録画を開始する．

本アプリケーションは，スライドの縦横比が 4:3 で作成された PowerPoint ファイル形式に対応している．

2.1 生成動画

本節では，生成される動画を詳述する．Fig.2 は，本システムの説明スライドを動画化しているスナップショットである．生成動画の縦横比は 16:9 で，画質は FWXGA である．

本動画は，事前に出力した画像化したスライドを CG エージェントが音声説明する形式で構成されている．CG エージェントはスライド画像右側に配置され，適宜動作する．スライド画像の下側には CG エージェントが話した内容が横書きの字幕として表示され，左側には音声内容を補足する字幕が縦書きで表示される．補足字幕は 5 秒間出力される．ただし，5 秒経過する前に新しい補足字幕を出力する場合は，古い字幕を上書きして出力する．後述する GUI アプリケーションによって CG エージェントの会話内容や音声，動画内の字幕表示方式を変更できる．

2.2 GUI アプリケーション

GUI アプリケーションでは文章の編集の他，音声の情報やアクセシビリティの設定を行う．これら情報は音声や表

表 1 アクセシビリティ設定値.
Table 1 Accessibility setting values.

設定値	音声の文章	音声の字幕	補足字幕
none	括弧以外の文章	なし	括弧内の文章
visual	括弧以外の文章	括弧以外の文章	括弧内の文章
visual+	全ての文章	全ての文章	なし

ファイル，設定ファイルに保存され，動画生成の際に利用する．音声は VOICEVOX を使用する．操作可能情報は音声名，音声の速度，抑揚である．

アクセシビリティの設定情報を Tab.1 に示す．起動時, 設定値はノート内の一部の音声と字幕が表示される「visual」となっている．この場合，聴覚障害当事者に対する情報補償が考慮された動画が生成される．また，PowerPoint ファイルのノートに記載されている全ての文章をスライド下部の字幕，音声で表現する設定を利用した場合，視覚障害当事者に対する情報補償が考慮された動画が生成される．

「ノート抽出」のボタンを押すと，VOICEVOX による音声化を開始する．音声化が終了した時に作成されたデータは保存され，アプリケーションを再起動した場合には保存データから作業を再開できる．

2.2.1 文章の編集

本アプリケーション内の「文章一覧」で表示される表を Fig.3 に示す．PowerPoint ファイルを読み込んだ際，本アプリケーションは自動的に文章表を作成する．

番号は行番号に該当する．この番号は，音声作成時にファイル名として使用する．ページは PowerPoint スライドにおけるスライド番号に該当する．この番号は，スライドを画像化する際にファイル名として使用する．動作 ID は使用する動作の ID 番号である．動作は，該当番号の音声再生中に使用される．文章表を作成した時点で，通常動作として割り当てられた 1〜4 の動作 ID の中からランダムで 1 つ選択されるため，編集せずに動画生成を開始した場合でもエージェントは動作表現を含めたプレゼンテーション動画が作成される．ただし，音声の文章が 30 文字以上の文章の内，60%の確率で動作 ID が自動付与される．このほか，7 種類の感情表現，3 種類の指差し動作，肯定・否定の動作などを動作 ID の手動変更で実装できる．

記号は文章の区切り文字である．対象文字は「！」や「？」を含む句点，鍵括弧，括弧文字である．対象の区切り文字を検出した場合，前回の検出点から今回の検出点までを音声の文章，字幕用の文章とし，DataGrid に格納する．音声の文章はナレーション音声に使用する文章である．字幕用の文章はナレーション音声の字幕である．1 つの音声につき，ナレーション字幕は約 65 文字出力できる．PowerPoint ファイルを読み込んだ際，音声の文章と字幕用の文章は文章情報が共有される．補足説明は補足用に使用する字幕である．括弧内に記載されている文章を直前の

© 1959 Information Processing Society of Japan

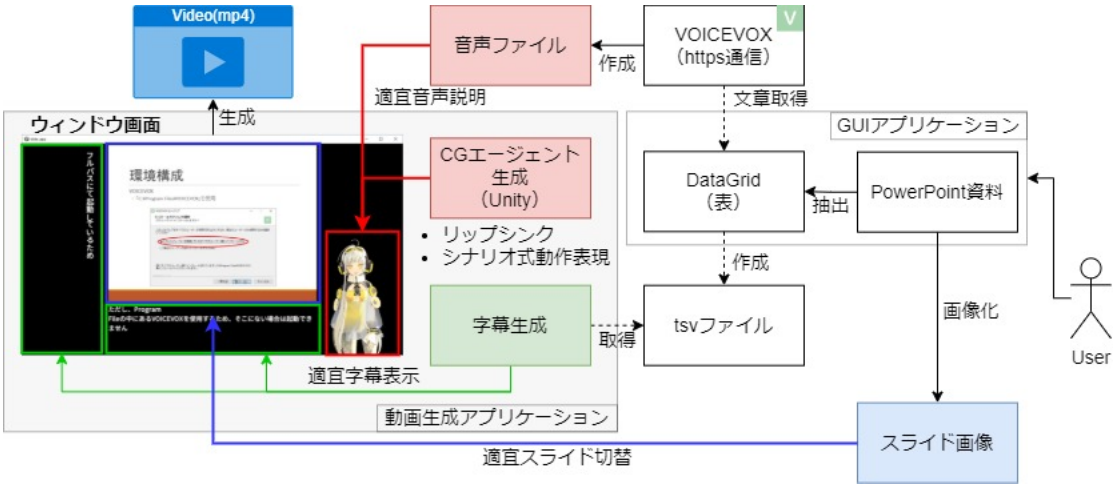


図 1 動画生成システム設計図.
Fig. 1 Application design diagram.

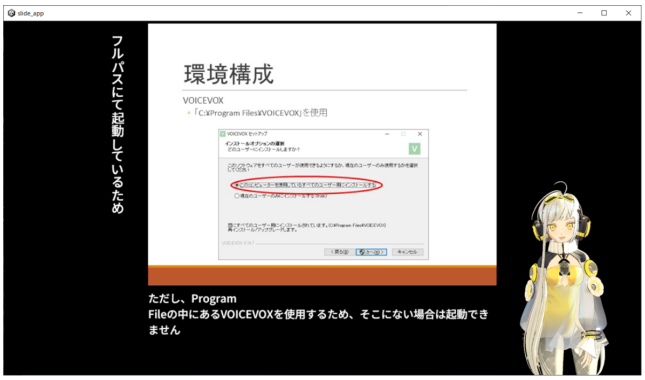


図 2 動画構成.
Fig. 2 Video composition.

音声・字幕用の文章に対する補足説明として利用する．

以上の情報は，PowerPoint ファイルのノート情報から抽出する．また，音声化操作の前には，使用音声情報の提示や「ここまでのご清聴，ありがとうございました.」という挨拶文が，お辞儀の動作 ID と共に挿入される．

2.3 動画生成アプリケーション

本アプリケーションは Unity で作成した．ウィンドウは，録画前に大きさを変更した場合でも縦横比が 16:9 になるように修正される．起動後，本アプリケーションのウィンドウを左クリックする事で mp4 動画形式で録画を開始する．録画終了時には Fig.4 のような録画終了ウィンドウの表示と効果音によって動画生成の完了を User に知らせる．CG エージェントは Ailis[6]，動作に使用しているモーションは Mixamo[7] を使用している．

通常，文章間の無音時間は 0.5[s] 確保している．ただし，句読点が区切り文字として検出した文章の場合は，対象の文章の後に 0.75[s] の無音時間が確保され，スライド画像が変わる際は 1[s] 確保される．

動画生成は画面録画による方式のため，動画生成時間は生成後の動画時間と同じである．

3. 生成動画の提供

本システムで生成した動画は情報補償 DVD として提供された．本章ではこれについて詳述する．

動画内容は，神奈川リハビリテーション病院で行っている高次脳機能障害のグループ訓練についてである．これは，訓練内容を地域で実践できる支援プログラムとして全国で高次脳機能障害に従事している支援者に広く周知する事が目的である．本動画は 5 章構成で，PowerPoint ファイルは合計 9 個ある．利用者層は高次脳機能障害支援に従事している家族会，支援拠点施設，地域支援者などが想定される．年齢層は 20〜30 歳代が主で，男女比はほぼ同じである．

元スライドと動画化後のスナップショットを Fig.5 に示す．ナレーション音声は VOICEVOX の春日部つむぎ [8]，CG エージェントは「Ailis」を使用した．また，PowerPoint ファイルのノートの内，括弧書きされた文章は補足字幕として表示し，それ以外の文章をナレーション音声・字幕として使用した．ただし，一部のナレーション音声・字幕は修正した．

DVD 配布にあたり，アンケートを同封している．今後は動画内容に関するアンケートの分析を行う予定である．

4. まとめ

本研究では視聴覚情報を補償されたスライド動画生成システムを開発し，高次脳機能障害のグループ訓練動画の DVD を作成し，配布した．スライド動画に対し視聴覚情報補償を施すことにより，同じ障害を持つ者同士の相互的なサポートを促す機会となり得る．また，DVD として広く配布する事で，障害支援従事者が視聴覚情報補償に関する技術を感じる機会や，障害当事者に対しても障害支援に関する研究内容を視聴する機会を送出する事ができる．

© 1959 Information Processing Society of Japan

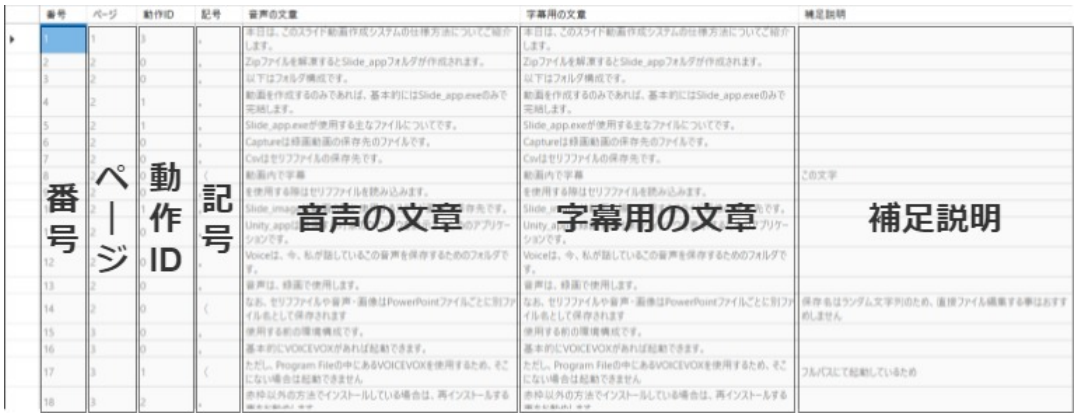


図 3 GUI アプリケーションで表示される文章表.
Fig. 3 Tables output to GUI applications.



図 4 動画生成完了ウィンドウ.
Fig. 4 Video generation completion window.



図 5 動画化するスライド (上) と生成動画のスナップショット (下).
Fig. 5 Top: Example of slides to be used, Bottom: Snapshot of created video.

本システムが利用可能な音声は VOICEVOX に準拠しているため，設定上は男性の音源が利用可能である．しかし，動画生成アプリケーションでは，利用可能な CG エージェントが 1 種類に制限され，なおかつ女性のモデルのみの実装であった．動画の利用者層を増やすため，今後は男性に限らず，年配者や中性的のような特徴のある CG エージェントの導入が考えられる．また，実際に病院施設関係者に本システムをご利用いただいた際，GUI アプリケーションの操作性の難度が問題となった．そのため，使いやすいアプリケーションとなるように UI デザインや処理方法を見直す必要がある．動画の提供方法も，DVD のみならず，紙面やホームページなど，多くの広告手段で活用できる動画生成システムへとしていきたい．

謝辞 本研究では，神奈川工科大学の小川喜道名誉教授，神奈川リハビリテーション病院の瀧澤学室長，神奈川県福祉子どもみらい局の池本達哉主任より多大なご助言，ご支援いただきました．厚く御礼申し上げます．

参考文献

[1] 後藤 博: 障害者の新たな活力を引き出すピアサポート〜持続的な活力向上のヒントに〜, LIFE DESIGN REPORT, https://www.dlri.co.jp/report/ld/176085.html (2021)
[2] 厚生労働省: 令和 5 年度障害者ピアサポート研修事業に係る指導者養成研修事業委託費の公募について, https://www.mhlw.go.jp/stf/newpage_30872.html (2023)
[3] 株式会社 4colors: *PIP-Maker*, https://www.pip-maker.com/
[4] Y. Yokoyama, K.Nagao: *VR2ML: A Universal Recording and Machine Learning System for Improving Virtual Reality Experiences*, 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp.756-757 (2020)
[5] 服部 広大, 長尾 確: *VR* 講義システムにおける *2D* 講義動画の自動生成, 第 83 回全国大会講演論文集, 2021, 1, pp.131-132 (2021)
[6] IZUMO: *IZUMO - Live as you want to be*, https://izumo.com/
[7] Adobe: *Mixamo*, https://www.mixamo.com/#/
[8] VOICEVOX: 春日部つむぎ, https://voicevox.hiroshiba.jp/

© 1959 Information Processing Society of Japan