The Bulletin of Seisa Dohto University
No.7, 215－222, 2026

研究ノート

Marp とクラウド TTS を用いた発話障害者向け授業・プレゼン読み上げシステムの提案

佐　藤　雅　一[1]・倉　橋　　農[2]・越　智　　徹[3]

概要

本研究では，Markdown ベースのスライド生成ツール Marp とクラウド型音声合成（Text-to-Speech, TTS）を用いた，発話障害者向け授業・プレゼンテーション読み上げ支援システムを提案する。筆者自身の障害及び授業実践を背景とし，発話を完全に自動化するのではなく，話者が主体的に進行を制御しながら必要な部分のみを逐次的に読み上げる点に特徴がある。ベンダニュートラルで拡張可能な構成とすることで，高品質かつ低コストな発話支援を実現し，教育・研究現場への応用可能性を示す。

キーワード：Marp，クラウド，TTS，読み上げ，発話障害，プレゼンテーション

１．はじめに

1.1　背景と目的

筆者の一人である佐藤は，感染性心内膜炎により心臓手術を受けた後，後遺症として多発性脳梗塞を発症した。その結果，言語障害および失語症を併発し，現在も発話に困難を伴う状態にある（以下，発話障害者と呼ぶ）。このような状況において，筆者自身が担当する授業や発表の場面で，自身の代わりに発話を補助・代替する手段が必要となったことが，本研究の着想の背景である。そこで筆者らは，生成 AI による音声合成技術を用いて，発話障害者が授業やプレゼンテーションを円滑に実施するための発話支援システムの構築を試みた。本システムでは，プレゼンテーション用スライドの作成に Markdown 記法を用い，Markdown からスライドを生成できるツールである Marp を利用することを前提としている。Marp は，PDF，HTML，PowerPoint 形式などへの変換が容易であり，VS Code の拡張機能やコマンドラインツールとして提供されている点に特徴がある。一方で，プレゼンテーションにおいて本質的に重要なのは，レイアウトや装飾といった表層的な要素ではなく，「何を伝えるか」という内容そのものである。しかし，PowerPoint をはじめとする一般的なスライド作成ツールは，グラフ作成，アニメーション，デザイン調整など多機能であるがゆえに，利用者が内容以外の要素に過度に時間を費やしてしまうことが少なくない。筆者自身も，こうした機能に引き込まれ，レイアウトや細部のデザイン調整に多くの時間を割いてしまう経験を度々してきた。Marp は，スライド内容を

[1] 星槎道都大学経営学部
[2] 羽衣国際大学現代社会学部
[3] 大阪工業大学情報センター

テキストとして記述することにより，利用者がプレゼンテーションの内容に集中しやすい環境を提供する。一方で，テキスト記述に基づくツールであるがゆえに，プログラミング的な記述に不慣れな利用者にとっては敷居が高いという指摘も存在する。

本報告では，このような背景を踏まえ，Marp とクラウド型音声合成（Text-to-Speech, TTS）を組み合わせた，発話障害者向けの授業・プレゼンテーション読み上げ支援システムを提案する。本システムは，Marp により生成されたスライドを再現し，画面上のテキストを逐次的に読み上げることで，発話困難な状況においても円滑な授業・発表を可能にすることを目的としている。

1.2　目指すもの

本研究では，Markdown ベースのスライド生成ツールである Marp を用いた発声支援システムを提案する。本システムは，一定程度の PC 操作スキルを有する利用者が，授業や研究発表，プレゼンテーションの場面において，話者自身の発話を補助する目的で気軽に利用できることを想定して設計されている。

本システムの特徴は，プレゼンテーション用途に特化している点にある。既存の特定プラットフォームや専用フォーマットに依存せず，テキストベースで資料および発声制御を記述できることを重視している。これにより，スライド内容と発話内容を同一の Markdown 文書内で管理でき，修正や再利用が容易になる。また，拡張性の高い構成とすることで，将来的な機能追加や他システムとの連携も可能としている。

一方で, 本システムは自動再生型の音声提示や，事前に録音された音声を再生する仕組みではない。スライド中の選択された部分を逐次的に読み上げる方式を採用しており，話者が進行を制御しながら利用することを前提としている。また，本システムは視覚障害者向け支援を主目的としたものではなく，利用者が画面内容を視認できる状況を前提としている。発話以外の障害，あるいはあらゆる障害への対応を目的とした汎用的な支援技術ではない点も，本研究の対象範囲として明確にしておく。

2．Marp

Marp（Markdown Presentation Ecosystem）は，Markdown 形式で記述されたテキスト文書を，スライド資料として解釈・変換するためのプレゼンテーション生成環境である。Markdown の見出しや箇条書きといった構造的記法を，スライドのページ構成や内容要素として対応付けることで，文書構造と発表構造を一貫して記述できる点に特徴がある。

Marp では，Markdown 文書内の区切り記号によってスライド単位が定義され，記述内容は HTML・PDF・PowerPoint 形式などに変換可能である。また，スライド全体や個別ページに対して，テーマ設定や表示制御といったメタ情報を付与できるため，内容と表現の分離が可能となる。このような特性により，Marp はプレゼンテーション資料を「視覚的成果物」としてではなく，「構造化された文書」として扱うことを可能にし，内容の再利用性や機械処理適性を高める。特に，執筆・発表・共有を同一の記述形式で行える点は，教育・研究環境における資料作成および分析に適している。

プレゼンテーション資料の作成には，PowerPoint や LaTeX Beamer をはじめとする多様な環境が利用されてきた。PowerPoint は GUI ベースの操作により直感的な資料作成を可能とし，教育現場や企業において広く普及している。一方で，スライドの構造は視覚的配置に依存しやすく，内容構造が明示的なテキストとして残りにくいという課題が指摘されている。また，資料の再利用や自動解析を行う場合には，内部構造の把握が困難である。

LaTeX Beamer は，TeX 文書としてスライドを記述する方式を採用しており，数式表現や体裁の一貫性に優れる。スライド構造がソースコード

として明示される点で，内容の再現性や版管理に適している。しかし，記法が比較的複雑であり，特に初学者や非技術系利用者にとっては学習コストが高い。また，文書構造とスライド構造が必ずしも直感的に一致しない場合がある。これらに対し，Marp は Markdown という簡潔な記法を用いてスライド資料を記述する点に特徴がある。Markdown の見出しや箇条書きといった文書構造が，そのままスライド構造として解釈されるため，文書執筆とプレゼンテーション作成の間の乖離が小さい。また，プレーンテキストによる記述は，版管理や機械処理との親和性が高く，資料の分析や再利用を容易にする。

以上の比較から，PowerPoint は表現の自由度と操作性に優れ，LaTeX Beamer は厳密な組版と再現性に強みを持つのに対し，Marp は構造化された記述と簡潔な表現を両立する環境として位置付けられる。特に，内容構造を明示的に保持したままプレゼンテーション資料を生成できる点は，教育・研究用途における分析対象としての価値を有すると考えられる。

3．先行するツールや事例

3.1　Microsoft PowerPoint, Edge

Microsoft PowerPoint や Microsoft Edge には，標準機能あるいは拡張機能としてテキスト読み上げ（Text-toSpeech, TTS）機能が提供されている。PowerPoint をはじめとする Microsoft Office 製品群では，スライド内容やノートを音声で読み上げることが可能であるが，発話の自然さや抑揚といった点で品質は十分とは言えず，大学講義や学会発表といった実利用に耐える水準には達していない。

Microsoft Edge においても，ブラウザ上で表示されたテキストを読み上げる機能が提供されているが，利用環境は Edge ブラウザに強く依存している。また，読み上げエンジン自体も Microsoft の提供する音声合成基盤に囲い込まれており，他社製エンジンへの切り替えや柔軟な拡張は困難である。

Marp と Edge を組み合わせた利用も考えられるが，Marp のプレゼンタービューでは提示画面のプレビューが SVG として描画されるため，ブラウザの TTS 機能では内容を直接読み上げることができないという技術的制約がある。このため，視覚的なスライド表示と発話内容の連携という点では十分な解決策とはなっていない。

なお，Windows や macOS には OS 標準機能としてナレーションやスクリーンリーダが搭載されているが，これらは主にアクセシビリティ用途を想定したものであり，発話品質や操作性の面で授業やプレゼンテーション用途には適さない。

3.2　自動読み上げ，スライド作成など複合的ツール

従来から，PowerPoint と連携して自動的にスライドを読み上げるツールとして PowerTalk が存在するが，対応環境が古く，主に Windows および PowerPoint に依存した構成である点から，現在の利用環境には適合しにくい。また，HeartyPresenter は身体障害者向けの講演支援システムとして開発されており，視線入力などの支援技術を用いて PowerPoint を操作し，あらかじめ用意したテキストを音声合成で読み上げる機能を備えている。このように，発話困難者を主対象としたシステムでは，操作支援と音声提示を一体化した設計が多く見られる。

近年では，より高度な自動化を志向したシステムも登場している。AimeTalk Virtual Presenter は，ユーザーが用意した顔写真から生成したアバターが，スピーカーノートを読み上げながら自動的にスライドを進行するソフトウェアである。一度設定すれば，プレゼンテーション全体を完全自動で実行できる点が特徴であり，発表動画の生成にも利用されている。さらに，このシステムに関連する研究として，大規模言語モデル（LLM）を用いてスライド内容から自動的に話者ノートを生成する試みが報告されており，プレゼンテーション準備の省力化という観点から注目されている。

加えて，近年提案されたPASS（Presentation Automation for Slide Generation and Speech）は，入力文書からスライド資料を自動生成し，各スライドに対応した発話スクリプトを生成した上で，音声合成による読み上げまでを一貫して行う包括的なプレゼンテーション自動化システムである。この研究では，コンテンツ生成と発表行為の完全自動化を特徴としているが，評価は主として生成結果の一貫性や関連性を自動的に測定するものであり，実際の聴衆を対象とした発表支援としての有効性評価は行われていない。

3.3　本研究の位置づけ

プレゼンテーションにおける音声提示や発話支援に関する研究は，これまで主としてアクセシビリティ支援や発表内容の自動化を中心に進められてきた。Peng らの研究［１］は，視覚障害者の視点から，スライド中の視覚情報と話者の口頭説明との不一致に着目し，プレゼンテーションの非視覚的アクセシビリティ向上を目的としたフィードバック手法を提案している。この研究は，スライドと発話の関係性を分析対象としている点で本研究と共通点を持つが，主眼は発表者の発話内容を改善するための評価・支援であり，発話そのものを補助・代替するシステムの設計を目的としたものではない。

一方，PromptTTS++ に代表される近年の音声合成研究［２］は，深層学習を用いた高品質な音声生成や話者特性の制御に焦点を当てており，TTS 技術そのものの高度化を目的としている。本研究においても音声合成技術は重要な構成要素であるが，TTS エンジンの性能向上自体を研究対象とするものではなく，既存の音声合成技術をどのようにプレゼンテーションという実践的文脈に組み込むかに主眼を置いている点が異なる。

教育分野における TTS 利用の有効性については，Wood ら［３］によるメタ分析があり，読み上げツールが学習支援として一定の効果を持つ可能性が示されている。ただし，これらの研究は主として学習者による受動的な利用を想定しており，話者自身がプレゼンテーション中に発話を補助するという利用形態については十分に検討されていない。また，神谷ら［４］は，スライド情報と発話音声との関係を分析することで，プレゼンテーション内容の理解や評価を支援する試みが報告されている。このような研究は，スライドと音声の対応関係を技術的に扱う点で本研究と共通するが，発表時に話者が利用する実時間の発声支援システムを対象としたものではない。

これらに対して本研究は，プレゼンテーションを完全に自動化するのではなく，話者が主体的に進行を制御しながら，必要な部分のみを逐次的に読み上げる発声支援を目的としている点に特徴がある。また，Marp によるテキストベースのスライド記述と連携することで，特定のプラットフォームやフォーマットに依存せず，発話内容とスライド内容を一体的に管理できる点も既存研究には見られない。本研究は，アクセシビリティ支援や自動プレゼン生成とは異なる立場から，教育・研究現場における実用的な発表支援の一形態を提示するものである。

４．Marp によるツールの現状

4.1　概要

本研究で構築している発声支援システムは，Markdown からスライドを生成する Marp を基盤とし，生成された HTML スライドに対して読み上げ機能を付加する構成を採っている。現状では，特定の統合開発環境（IDE）に依存した拡張機能としては実装しておらず，一般的なテキストエディタ上で Markdown 原稿を作成し，Marp により HTML スライドを生成する運用形態を採用している。

生成された HTML スライドには，発声支援を実現するための JavaScript コードを後処理として挿入しており，スライド中のテキストを取得して音声合成 API に送信することで，逐次的な読み上げを可能としている。音声合成エンジンとしては，現状では Google の提供する TTS サービス

図 1　システム概要

を利用しており，API 呼び出しは中継用の API Proxy を介して行われる。

この一連の処理の流れを図示すると，図 1 のような構成となる。

現状では，本システムは特定のエディタや IDE に依存しない形で運用されており，利用者は任意のテキストエディタを用いて Markdown 形式でスライド原稿を記述する。この設計により，OS や開発環境の違いによる制約を受けにくく，既存の Marp 利用環境に容易に組み込むことが可能となっている。

一方で，スライド生成や読み上げ機能の設定，音声合成エンジンの切り替えといった操作をより簡便に行うためのインタフェースが求められる場面も多い。そのため，今後の拡張として，Marp 利用者の多い VS Code 上で動作する拡張機能として実装することを計画している。VS Code 拡張とすることで，スライド生成操作と発声支援機能を統合し，利用者の操作負荷をさらに低減できると考えている。

4.2　利用者とシステムの役割分担

本ツールにおける利用者とシステムの役割分担を以下に整理する。

利用者が行う操作は，主に以下に限定される。

・Markdown 形式によるスライド内容の記述
・Marp の記法に基づいたスライド構造の編集
・Marp を用いたスライド生成操作一方，システム側で行う処理は以下の通りである。
・Marp により生成された HTML スライドへの JavaScript コードの挿入
・スライド中のテキスト情報を取得し，TTS API へ送信する処理
・読み上げ制御のためのユーザインタフェースの提供（例：発表中に即座に入力可能なアドリブ用テキスト入力欄）

本研究で新たに実装した要素は，HTML スライドに対する発声支援用 JavaScript の組み込み機構，および逐次読み上げとアドリブ入力を可能とするユーザインタフェースである。

4.3　API キーの扱い

音声合成 API を利用するにあたり，API キーの管理が課題となる。現状では，利用者が各自で API キーを取得し，設定する必要があるが，教育現場や複数人での利用を想定した場合，この運用は必ずしも適切とは言えない。

本研究では，API Proxy を介した構成を採用しているため，将来的には Proxy 側で API キーを集約管理する方式や，利用範囲を限定したキー配布の仕組みを導入することが可能である。これらの点については，今後の課題として検討を進める予定である。

5．まとめと今後の課題

5.1　利点と問題点

5.1.1　利点

本研究で提案する発声支援システムには，以下のような利点がある。

第一に，ベンダニュートラルな構成である点が挙げられる。本システムは，特定のプレゼンテーションプラットフォームや音声合成エンジンに強く依存しない設計となっており，音声合成 API についても独立した構成を採用している。このため，将来的に利用する音声合成サービスを他社製のものへ切り替えることが比較的容易である。

第二に，高い拡張可能性を有している点である。Markdown と HTML を基盤とした構成により，発声制御の追加やユーザインタフェースの改良などを段階的に行うことができる。特に，読み上げタイミングの制御やアドリブ入力支援など，発表現場のニーズに応じた機能拡張が可能である。

第三に，高品質な音声合成を比較的低コストで

利用できる点が挙げられる。クラウド型の音声合成 API を活用することで，専用機材や高価なソフトウェアを導入することなく，自然な発話品質を実現できる。この点は，教育現場や個人利用においても導入しやすい特徴である。

5.1.2　問題点

一方で，本システムにはいくつかの課題も存在する。

第一に，利用にあたって一定程度の技術的背景が求められる点である。Markdown 記述やスライド生成，API 利用といった操作に不慣れな利用者にとっては，初期導入のハードルが高い可能性がある。この点については，操作手順の簡略化や将来的な GUI 支援が課題となる。

第二に，ネットワーク接続が必須である点が挙げられる。音声合成 API を利用する構成上，オフライン環境では読み上げ機能を利用できない。そのため，ネットワーク環境が不安定な場所での利用には制約が生じる。

第三に，API キーの管理に関する問題である。現状では，利用者が各自で API キーを取得・設定する必要があり，教育現場や複数人での利用においては運用上の負担となる。API Proxy を用いたキーの集約管理など，より安全かつ簡便な運用方法については，今後の検討課題である。

5.2　課題と展望

本提案で提案した発話支援システムは，現時点では基礎的な機能を中心とした実装に留まっており，まだまだ課題があり，今後は次の 3 点の改善・拡張を予定している。

5.2.1　利用者インタフェース

まず利用者インタフェースの改善として，Visual Studio Code 上で動作する拡張機能としての実装が挙げられる。現状では，エディタとスライド生成，読み上げ機能の設定が分離した形で運用されているが，VS Code 拡張として統合することで，スライド作成から発声支援機能の利用までを一貫して行える環境を提供できると考えている。これにより，利用者の操作負荷を低減し，技術的背景を持たない利用者にとっても扱いやすいシステムとなることが期待される。

5.2.2　音声合成エンジン

次に，音声合成エンジンの多様化である。本研究では現状，クラウド型 TTS サービスの一つを利用しているが，ベンダニュートラルな設計を活かし，複数の TTS エンジンを選択可能とすることで，発話品質やコスト，利用条件に応じた柔軟な運用が可能となる。特に，教育現場や研究発表といった利用文脈に応じて，最適な音声合成エンジンを選択できる仕組みの検討が重要である。

5.2.3　ローカルフォールバック

最後に，ネットワーク環境への依存を低減するためのローカル環境へのフォールバック機構が挙げられる。現行システムでは，音声合成処理にクラウド API を用いるため，安定したネットワーク接続が必須となっている。今後は，ネットワーク障害や利用制限が生じた場合に備え，簡易的なローカル TTS エンジンへ自動的に切り替える仕組みを導入することで，可用性の向上を図ることができると考えている。

これらの拡張により，本システムは発話障害者に限らず，教育・研究現場におけるプレゼンテーション支援ツールとして，より実用的かつ柔軟な基盤へ発展させることが可能であると考えられる。

5.2.4　フィードバック

本研究に関して，研究者・教育関係者を含む小規模なミーティングにおいて概要説明を行い，意見交換を行った。以下は，その際に得られた主な指摘や示唆を整理したものである。

まず，音声合成 API の選定に関して，Gemini API には年齢制限（18 歳以上）が存在する可能性が指摘された。教育現場での利用を想定した場合，利用者の年齢条件や利用規約の制約について事前に十分な確認が必要であることが示唆された。

次に，本システムの想定利用者について，発話が困難な場面緘黙の利用者にとっても有用である可能性が指摘された。一方で，そのような利用形

態を想定する場合，既存の読み上げ支援サービスや福祉用途向けシステムとの比較検討をより丁寧に行う必要があるとの意見も得られた。

操作性に関する観点では，Markdown や VS Code といったツールに対して，一定数の利用者が「難しい」と感じる可能性がある点が指摘された。特に，情報系以外の教員や一般利用者を想定した場合，ツール選択や操作フローの分かりやすさが課題となることが示唆された。

また，音声合成 API の利用に伴い，API キーを利用者が各自で取得・管理する必要がある点についても，運用上の課題として指摘があった。教育機関や複数人での利用を想定する場合，API キー管理の負担を軽減する仕組みが求められる。

さらに，比較対象として Google Slides に標準搭載されている読み上げ機能を確認すべきであるとの指摘もあり，既存のプレゼンテーションツールが提供する音声機能との違いを明確にする必要性が示された。

最後に，スライド原稿の作成方法について，Markdown 記述を生成 AI に任せるという意見も出された。しかし，現時点ではスライド構造や発話内容を細かく調整する必要がある場面が多く，利用者自身が直接記述した方が効率的であるケースも少なくないと考えられる。この点については，今後の利用状況を踏まえて検討する余地がある。

付記

本原稿は情報処理学会 第 183 回コンピュータと教育研究会で発表した内容である。

参考文献

［1］Peng, Y.-H., Jang, J., Bigham, J. P., and Pavel, A.: Say It All: Feedback for Improving Non-Visual Presentation Accessibility, Proceedings of the ACM on HumanComputer Interaction, Vol. 5, No. CSCW1, Article 47, 2021. https://doi.org/10.1145/3449181

［2］Shimizu, R., et al.: PromptTTS++: Controlling Speaker Identity in Prompt-Based Text-to-Speech Using Natural Language Descriptions, arXiv preprint, arXiv: 2309. 08140, 2023. https://arxiv.org/abs/2309.08140

［3］Wood, S. G., Moxley, J. H., Tighe, E. L., and Wagner, R. K.: Does Use of Text-to-Speech and Related Read-Aloud Tools Aid Reading Comprehension? A Meta-Analysis, Journal of Learning Disabilities, Vol. 51, No. 1, pp. 73–84, 2018. https://doi.org/10.1177/0022219416688170

［4］神谷賢太郎，東中竜一郎，川瀬卓也，長尾確：プレゼンテーションにおけるスライド情報を用いた発話内容分析，第 83 回全国大会講演論文集，Vol. 2021，No. 1，pp. 215–216，2021.

Proposal of a lecture/presentation reading system for speech-impaired people using Marp and Cloud TTS

SATO Masaichi　　KURAHASI Minori　　OCHI Toru

Abstract

This study proposes a speech-reading support system for classes and presentations for individuals with speech impairments, utilizing the Markdown-based slide generation tool Marp and cloud-based text-to-speech (TTS) technology. Based on the author's own disability and classroom experience, the system's key feature is that it does not fully automate speech. Instead, it allows the speaker to proactively control the flow while sequentially reading only necessary sections. By adopting a vendor-neutral and extensible architecture, it achieves high-quality, low-cost speech support and demonstrates applicability in educational and research settings.