

音訳支援システム DaisyRings の 開発と音訳コミュニティでの実証— 読むことが困難な人にも楽しめるオーディ オブック作成—

布目 光生^{†1} 黒田 由加^{†1} 芦川 平^{†1} 森田 眞弘^{†1}

^{†1} (株) 東芝

読むことが不自由な方でも楽しめる朗読図書には、古くからの「録音図書」やオーディオブックをはじめとして、近年の読み上げ音声付きの電子書籍、また専用の図書フォーマットである DAISY 形式の図書などがある。しかしながら、こうした図書の多くは、ボランティアによる朗読や編集にゆだねられており、1冊の図書を訳するためには多くの時間が必要とされてきた。一方、近年は音声合成技術の進展により、人の声と区別がつかないような高品質な合成音声や、それらを安価に提供する Web サイトなどの自動読み上げ手段なども一般的になってきたが、読みやアクセントの誤りを完全になくすることは難しく、全自動での合成読み上げでは内容や理解の誤りを避けることができず、あくまで自己責任の利用に任されていた。そこで、我々は自社の音声合成技術をベースに、正確な読み上げ音声付きの図書／ドキュメントを簡単に作成できる音訳システムを開発した。日本語特有の漢字の読みの曖昧性や、アクセントの違いで意味が異なってしまう仮名の読みを指定したり、より理解しやすい朗読のためポーズ指定を可能とすることが特徴である。本システムを Web アプリケーションとして構築し、音訳ボランティア団体や公共図書館・点字図書館をはじめ障がい者支援施設のスタッフ、大学の司書課程の実習などでご利用いただき、実践上の課題抽出と改良を行った。本稿ではこの音訳システム開発の取り組みについて述べる。

1. はじめに

1.1 音訳とは

電子書籍の普及に伴い、従来の紙文書では難しかった電子書籍ならではの活用が広がっている。ネットワーク経由によるコンテンツ配信や管理はもちろん、柔軟な表示形式として文字サイズやフォントの変更、さらには音声合成による読み上げ音声を楽しむことも可能となってきた。

一方、これまでも視覚障がい者など文字を読むことが難しい人のために、「録音図書」として、点字図書館・公共図書館では、対面朗読やカセットテープ、CDなどに朗読音声収録されたものが存在する。さらに近年は、こうした「録音図書」も、音声情報のみでなくアクセシビリティを考慮した DAISY (Digital Accessible Information SYstem) などの形式で提供されている場合も多い。

慣習的に、こうした紙媒体をはじめとした書籍や文書を読み上げて音声にすることが、音訳と呼ばれている。一般的には、音訳者が朗読することで読み上げ音

声を作成するが、声や発話スタイルの違いのほか、コンテンツそのものをどう聞き手が分かりやすいように解釈するか（図や表などの説明も含む）という違いも音訳者によって差がある。

1.2 音訳を取り巻く環境と DAISY 図書

従来「録音図書」に携わっておられたボランティアの方々の多くは、「読み手」として活躍されてきた。そのため、従来から音訳に携わってこられた方は、必ずしも近年の DAISY 図書などを作成するのに必要な、テキストのタグ付けやオーサリング作業など、PCを必要とする作業には馴染みがない場合も多い。また、1に示すように、現在の音訳作業は依然としてボランティア主体で行われている実情がある。そこでは、読み上げ音声を録音する「読み手」担当と、PCを操作して編集・オーサリング作業を行う「編集」担当に分かれていることが多く、分担しながら調整をしつつ作業を進めていく必要がある。

こうした作業では、通常、たとえば点字図書館や盲人図書館から、書籍の DAISY 化依頼を受けてから、DAISY 形式図書を納品するまでに、早い場合でも

DaisyRings は、(株) 東芝の登録商標です。

2～3カ月、長い場合には1年ほどが必要といわれている。そのため、従来の朗読による音訳作業では、ニーズやリクエストに対して、コンテンツの供給がまったく追いつかない、または、エンドユーザがほしいコンテンツと提供する側のコンテンツの選定にギャップがあるなどの課題が生じていた。

たとえば、識字障がいの児童・生徒を持つ保護者や教員にとっては、学年ごとに全教科のDAISY教科書や教材が必要だが、全国のボランティア団体を合わせても、改訂のタイミングに追従して、教科や出版社のバリエーションに応じたDAISYコンテンツを揃えることはできていない。保護者の中には、こうした状況を少しでも改善しようと保護者自らが、地域の音訳ボランティアに参加し、DAISY教科書の拡充に自主的に貢献しようとする事例などもある。しかし、そこで優先的に作業を指定された教科書が、自分の子どもが使用しているものとはまったく関係ない教科や出版社のもので、せっかくの作業が、今DAISYを必要とする自分の子どもに結びつかなかった、などのエピソードもある。

また、近年DAISY化が求められる書籍も、ジャンルや内容、対象ユーザが多様化している。従来は、「読み手」が朗読しやすかった絵本や物語、小説、雑誌記事などの文書が一般的であった。しかし、昨今では、義務教育のための教材だけではなく、高校や大学などで用いる参考書や参考文献、さらにはIT技術書、ビジネス書、医療系や、法律の専門書、ときには外国語学習の教材なども含む場合があり、朗読者にも専門性や特定分野の知識が要求される状況となってきた。

1.3 音訳支援の既存技術

ところで、このDAISY形式の書籍作成には、読み上げ音声データのほか、章立てなどの階層化されたテキ

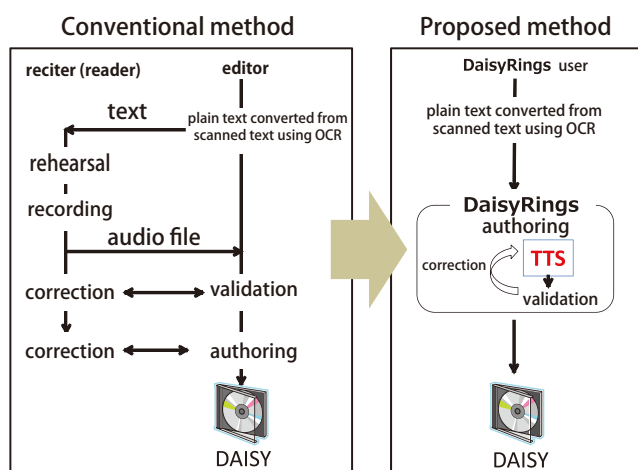


図1 音訳プロセス

スト情報、さらに音声とテキストを紐付けるための対応関係、目次情報などを作成する必要がある。この作業には、DAISYメタデータの仕様を理解して、タグを付与しながら構造化文書を作成することが必要になるため、XML（Extensible Markup Language）に関してある程度の知識やスキルが必要となる。HTML（Hyper Text Markup Language）の文書作成支援などと同様、DAISY化作業を支援するための統合環境やツール／プラグインなどが、無償のもの[1],[2]、有償のもの[3],[4],[5],[6]を問わず公開・販売されている。

また、表記と読みの対応付けが難しい数式の読み上げ[7],[8]、グラフと関連付けられたインタラクティブな学習コンテンツの提案[9]、音声ガイド提供のためのHTML5（track要素）の活用[10]、DAISY用HTMLタグの付与支援システム[11]などの試みもある。さらにコンテンツ再生環境にかかわる取り組み[12]などの報告もある。

2. 音訳作業における課題

「音訳システム」の開発にあたり、関係者へのニーズヒアリングと顧客要求の抽出を行った。初期プロトタイプを構築してのUI/UXの改良と並行して、川崎市盲人図書館（当時）をはじめ、日本点字図書館、千葉県立西部図書館、大阪府立図書館などの関係施設や、全国音訳ボランティアネットワークをはじめ、各地の音訳ボランティア団体など十数団体を訪問し、音訳作業フローや利用ツールをはじめとした具体的な内容から、作業者／利用者としての困りごとなどを伺った。

まずヒアリングの過程で、実際の音訳作業の多くがボランティアベースで成り立っている、という社会的・構造的な枠組み自体が大きな問題であることが分かった。これ自体もちろん大きな課題であるが、本稿では技術的な支援手段としての観点で、システム・アプリケーション提案による技術課題の解決・提案をスコープとする。

ユーザからのVoC（Voice of the Customer）は数百に上ったが、そこからの本質的な顧客要求・課題として、我々は次の4点を抽出した。

2.1 DAISY 提供に至るまでの長い納期

音訳作業の多くは図書館や支援施設に属するボランティア団体によって行われている。通常は、図1左に示すように、肉声で朗読しその音声データを収録する読み手と、DAISYフォーマットとして必要なメタデー

タ付与等のオーサリング作業を行う編集者グループとに分かれていることが多い。週に一度などの集まりで、図書館や施設からのリクエストに応じ音訳書籍の決定、作業の分担、全体のスケジュール調整を行うとともに、音訳の方針（ポリシー）について決定する。

実際の作業は、会合の時間以外にもグループや担当ごとに、施設の録音ブースやPC設備を用いたり、自宅での朗読や編集作業を実施することもある。ボランティアと呼べる範囲を越えるほどに、献身的に多くの時間を割いて、たとえば主婦業の合間を縫って深夜に作業をされている音訳者も少なくない。本の朗読をする読み手は、事前準備として、文書に出てくる用語の意味や読みの調査、さらに文字化されていない図表や写真の説明文作成、下読みなどが必要であり、読み誤りによるリトライなどを含めると、再生時間の3倍ほどの時間が朗読作業に必要といわれている。また、音声だけでなくテキストが同期したマルチメディアDAISYの場合は、本文のテキスト情報も必要なため、紙媒体からOCR（Optical Character Recognition）でスキャンしたり、直接編集者がタイピングすることでテキスト化を行う。最近では、一部出版社がアクセシビリティ用途に直接テキストデータが提供される場合もあるが、まだまだ一般的とはいえない。

あらかじめ決められた章単位などの一定の分量の朗読作業が終われば、データを受け渡して、別の者が朗読音声を確認（校閲）し、オーサリング担当が音声とテキストを同期させながらDAISY図書にビルドする。校閲も、その文書の性質によって、1度だけでなく、複数回行われることもある。修正要求が生じた場合には、前述の集まりやメールベースで調整したり、修正依頼票を起票して、どこがどういった理由のためにどう修正すべきか、などの情報をやりとりする。こうしたやりとりが発生することもあり、一般に図書1冊の音訳作業納期は、ほぼ1年後に設定されることが多い。

2.2 録音音声品質のばらつき

各地のボランティア団体に作成が委託されているため、当然のことながら朗読者の個人差があることは否めない。朗読者とデータ編集者などの自主的な役割分担があるが、善意のボランティアである以上、仮に聞きやすくない朗読があったとしても、指摘や要求を入れるにくい、という現状がある。たとえば、読み方そのものの上手・下手や、声の好み（男声・女声、年齢、明瞭性、滑舌の良し悪し）、音声の品質、呼吸ノイズの

有無、朗読話者のクセなど、多くのばらつき要因がある。また、録音環境も大きく異なる。図書館や支援施設によっては専用の録音室を備えたところもあるが、室数や利用できる時間帯は限られており、多くはボランティアメンバーの自宅での収録となる。その場合、部屋の環境や反響、周辺ノイズ、マイクの品質や録音装置の違いがある。また、長い作品の場合は、一定の声の調子で読み上げることは非常に困難である。どんなに長くても1日あたり2～3時間の収録が限界であり、より長い作品の場合は日を分けて朗読音声を収録するのが普通であるため、同じ人物の朗読であっても作品の途中で調子が変わってしまう場合がある。特に、校正の結果、日を改めて読み直しに対応する場合は、それがより顕著となる。

2.3 解釈のゆらぎ

いわゆる「翻訳」では、翻訳者の知識、経験、言語スキル、感性などが反映されるのと同様に、「音訳」にも、音訳者によって、内容の解釈や読み上げ結果が異なるものとなる。たとえば、書籍中のテキスト（文字列）の内容に応じた読み上げはもちろんであるが、図や表に何が書かれているかを音声のみで伝わるように説明することは非常に難しく、結果、理解しやすい作品となっているかどうかの個人差も大きい。もちろん典型的なものについては読み方のガイドラインが設定されていたり、事前に勉強会や講習会を通じて、音訳者は一定のスキルを有することが期待されている。また、音訳者の中には元教員経験者も多く、国語や社会、数学など、適材適所になっている例もある。しかしながら、すべての作品がそうとはいえず、音訳者による解釈の良し悪しや、分かりやすさ・分かりにくさの違いが生じてしまうことは否めない。

2.4 読み上げスタイルのバリエーション不足

一般に電子書籍であれば、ユーザが見やすいようにフォントサイズや文字間隔などのカスタマイズが可能である。この比較として収録音声では、声の大きさ、スピード、間の取り方、話者の性別や年齢など、聞き手が希望するような読み上げスタイルのうち、ごく一部しかカスタマイズすることができない。さらに、最終納品がサピエ^{☆1}に登録される場合には、原則、一作品一データの収録となる。そのため、ある作品

☆1 日本点字図書館がシステム管理を行っている点字図書や録音図書などの全国最大の書誌データベース

のDAISYコンテンツがすでに登録されている場合、たとえその品質に満足できなかったとしても、別のバージョンを作成し公開することは事実上難しい。

3. 音訳者のニーズに応える基本コンセプト

「音声合成技術」によれば、自然文のテキストさえ与えれば任意の読み上げができるため、人の朗読とは異なり、音声収録設備や体調、時間に依存せず、常に同じクオリティで読み上げることができる。そのため、短い納期での作業や、頻繁に改訂や修正が入る文書でも、オーバーヘッドなく簡単に作成することができる。もちろん、すべての図書や文書の音訳を音声合成に置き換える必要はなく、それぞれの特性を活かし、速報性の高いものは音声合成に、一方で情感豊かな読み上げが必要な物語や詩などを人による朗読で、といった切り分けを行うことで、それぞれのメリットを最大限に活かし、作業負担を軽減することが可能である。

3.1 設計思想

我々は先の4つの顧客要求・課題に対し、以下の方針をとった。

- 音訳作業におけるプロセス改善

根本的な課題である「長い納期」に対しては、作業フローの革新的な改善が必要である。高品質な音声合成技術の利用とその修正手段をWebアプリケーションとして提供し、それを軸とした編集作業のみでDAISYコンテンツ作成作業を完結させることで、データの受け渡しや調整に必要な余計なオーバーヘッドやボトルネックを解消し、ユーザはいつでもどこでも高品質な音声付き図書の作成を可能とするシステム構成とした。(図1右側)^{☆2}

- 音声合成による多様・迅速・安定的な読み上げ音声提供
「録音音声の品質」に対して、音声合成であれば、人と異なり、体調や録音環境に依存することなく、また再収録に必要な時間調整や、度重なる変更にも追従して何度でも繰り返し、読み上げ音声を出力することができる。また、当社の音声合成の「声」は、男女や年代の違い、また怒りや喜びといった感情の有無などの「音声キャラクタ」を選択して使い分けることができるため、従来はコストや手間から現実的ではなかった話者の変更や、複数人による読み分けを簡単に実現することができ、よ

り聞き手に理解しやすいコンテンツを作成することができる。

- ワープロ感覚の簡単な編集手段の提供

音声合成技術の読み上げ精度は完全ではなく、読み誤りやアクセント誤りを含む。こうした誤りに対し、PCを扱える人であれば簡単に修正できるインタフェースを提供することとした。「解釈のゆらぎ」に対しても、読み上げ文書を簡単に修正して「音声」として確認することができるので、録音のプロセスを文書推敲の時間に充てることができ、表や図の解釈を文書として客観的にレビューし、自由に校閲できる。

- 業界標準の制御記号によるきめ細かなコントロール

「読み上げスタイルのバリエーション」を確保するためには、編集者のスキルに応じて音声合成をコントロールできる幅に広がりを持たせることが必要である。本システムでは、音訳初心者でも編集可能な入力インタフェースに加え、音声合成の標準的な読み上げ制御記号も入力可能とすることで、スキルを持った編集者が、より細かい韻律調整やポーズなどの編集を可能とする方針とした。

こうした基本コンセプトを踏まえ、具体的な機能を選定・設計していく上で、システム設計におけるポリシーとして機能要求を以下の通りに定義した。

- ジャンルや文書を問わずさまざまなコンテンツ読み上げニーズに応えられること（**入力文書の汎用性**）
- 簡単な操作で手早く思い通りのコンテンツを作成できること（**編集作業の限定性**）
- 音訳ボランティア団体にとって、求められるPCスキルの負担を軽減できること（**初期導入の容易性**）
- 従来の音訳コミュニティの方々や対面朗読者などの方々にも、既存手段と便利に使い分けいただけるようなコンテンツ作成支援手段であること（**既存の作業ルーチンとの親和性**）

3.2 DaisyRingsを用いた新しいワークフローの提案

これらの機能要求に対し、我々のとったアプローチは以下の通りである。

入力文書の汎用性：

まず、本システムでは入力として本応用に特化した独自のフォーマットではなく、汎用だが記述力が限定されるプレーンテキストを受理するものとした。

現状、音訳作業では紙媒体からのテキストの書き起しが必要な場合が多いため、レイアウト情報などを考

^{☆2} TTS: text-to-speech

慮せず、単純に各行の単位で音声化するというアプローチをとる。

これによって、作業者の文書作成支援環境によらず、たとえば書籍だけでなく、図書目録やメール、手紙、案内通知など、テキスト情報さえあれば、ニーズに応じた音訳が可能となる。

音声化にあたっては、当社の音声合成技術[13]を適用する。音声合成技術は、基本機能として、漢字仮名混じり文を入力し、その読み上げ音声出力する。これを活用することにより、編集者は、手持ちのテキスト文書があれば、それを手早く音声ファイルとして入手できる。人の朗読で必要だった録音環境（マイクやPCなどの機材のほか、防音室や防音対策用の環境も含む）は不要となり、それらの時間的／コスト的な制約が不要となる。

また、編集者が一人で作業できるので「読み手」のコンディションやスケジュールに捉われることがなく、音声を自動で合成するので常に一定の品質の音声出力できる。それらの作業は、編集者が並行して進めることが可能である。

一方で、音声合成で音質面が許容できたとしても、テキスト処理に由来する読み誤りやアクセント誤りが生じることは免れない。また、音声合成の調整手段として、読みやアクセント、さらにピッチやボリュームなどをSSML（Speech Synthesis Markup Language）のタグとして指定することにより、きめ細かな出力が可能である。しかし、一般の音訳編集者にとって音声合成特有の記法の習得は容易ではなく、これを強いると音訳作業のハードルがさらに高くなってしまう。

編集作業の限定性：

そこで、我々は編集できる作業を限定することにより、作業を単純化し、作業範囲を明確化する方針をとった。具体的には、入力テキストに対し、ルビベースによる編集GUIを用い、原則以下の編集のみを可能とする。

- 入力テキストの文字列修正
- 入力テキストへのポーズ情報（長・短・削除）の挿入
- 入力テキストに対する「ルビ」による記入
 - 読みの付与（漢字、かな）
 - 読みとアクセント情報の付与（カタカナによる表音文字列および記号）
- 入力行単位へにDAISY形式の章立て構造（論理階層）付与（選択）
- 入力行単位で音声合成の話者（キャラクタ）を選択指定（選択）

その他、基本操作として、音声再生・停止（Enterキー）、行移動（カーソル）など、効率的なキー操作によって入力テキストの音声を確認しながら、編集を進めていくことを特徴とする。

初期導入の容易性：

編集者によっては、新しいソフトのインストール作業や各自のPC環境に依存するような設定作業は極力避けたいという要望がある。また、システム提供者側にとっても、編集者から寄せられた要望を反映し、システム自体の機能拡張や音声合成エンジン・話者辞書の更新を迅速に行えることが望ましい。

既存の作業ルーチンとの親和性：

音訳作業のプロセスには、いくつかの特徴的なステップがある。

- (1) 書籍情報のテキスト化
- (2) テキストの読み上げ音声データ生成
- (3) 読み上げ音声データとテキスト情報の同期情報生成
- (4) 所望の書籍フォーマットへのまとめ上げ・変換

本提案システムでは、上記のステップのうち、(2)と(3)の作業を主な支援対象とし、(4)の一部としてDAISY形式への出力をサポートするという位置づけとした。

すなわち、本システムでは前処理として、たとえば紙媒体の書籍であればテキスト化作業が必要である。テキスト化には、スキャナでスキャンしたテキストへのOCRの適用、文字認識誤りの修正[14],[15]、あるいは目視によるテキスト書き起こし作業が必要である。一方、書籍が電子媒体の場合も、読み上げ対象となるテキスト情報をレイアウト情報を考慮して正しい読み順で取り出しておく必要がある。

本システムではこうした前処理を対象外とし、その後のテキストデータを入力として扱う。この結果、編集者は入力文書からテキスト以外のレイアウト情報やメタデータを活用できないが、汎用性と修正の容易性が確保できる。また、出力はDAISY形式とすることで、そのまま再生や閲覧が可能であるほか、編集者がさらにほかのアプリケーションやツールからインポートすることにより、所望の形式に加工したり変換できる。

4. 音訳システム「DaisyRings」の開発

4.1 各機能の特徴

以下、本提案システムであるDaisyRingsの概要について述べる。本システムでは、音訳作業者にとって必要な、

- ユーザ文書のアップロード

- 編集、読み・アクセント修正作業
- 確認作業・検証作業
- ビルドブック（出力）作業

を総合的に支援する環境を提供する。基本的に、ボランティア団体や著作権法第37条3項に該当する施設^{☆3}等でのグループユースを想定しており、作業者個別のログイン機能、作業者ごとのマイページ機能、編集機能、音声再生機能、ダウンロード機能などで構成される。

4.2 修正手段 GUI の具体例

次に、特徴的な各部の詳細について述べる。本システムの大きな特徴は、音声合成技術と連携した編集機能である。

まず、基本となるのは、図2に示すルビによる入力UIである。編集者は、入力文書中の任意の文字列^{☆4}を選択し、特定のキー操作（デフォルトでCtrl-space）により、選択文字列に対するルビを入力できる。

通常「ルビ」では、ひらがなやカタカナにより、語の読みや説明をつけることが目的であるが、提案システムではこの概念を拡張し、「漢字」「アクセント記号」もルビUIより入力可能としている。これらの文字種を入力することによって、読みとアクセントを編集者の理解度に応じて、段階的に修正できる。

- ひらがな：ルビ中の文字列を平板型アクセントで読み上げる。たとえば、「市場」を「しじょう」と読ませるのか「いちば」と読ませるのかの指定や、難読語や固有名詞などの読みを簡単に指定できる。

☆3 視覚障がい者その他視覚による表現の認識に障がいのある者の福祉に関する事業を行う者で政令で定めるもの。

☆4 文書例の一部は、青空文庫「ごん狐」より抜粋。

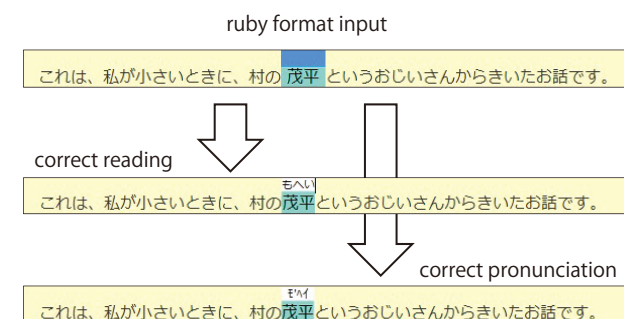


図2 ルビ形式入力

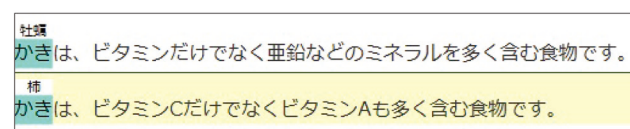


図3 ルビ漢字指定の例

- 漢字：図3に示す通り、小学校低学年などのコンテンツで頻出するひらがな表記などは、音声合成では正しいアクセントが付きにくい。こうした文字列にあえて漢字でルビを付与することによって、漢字のアクセントさえ分かれば、正しい漢字の読み・アクセントを指定できる。

- 半角カタカナと記号：音声合成での一般的なアクセント指定記号（' や%など）を習得していれば、前段よりもきめ細かな指定ができる。

また、音声合成読み上げでは、文脈によっては「間（ポーズ情報）」の入れ方が不適切になる場合[16]がある。

本UIでは、これを長／短の2種類用意し、図4に示す通り簡単なキー操作で入力文書中に挿入できるものとした。これによって、箇条書き番号の前後や、記号の前後、特定の重要語や初出語の場合などに、適切な間を簡単に挿入することができる。また、音声合成が自動的に挿入したポーズ情報を削除することも可能である。

また、最終的なDAISY出力では、作品タイトルなどの書誌情報や、はじめの粹や終わりの粹読みといった固定の注釈情報が必要となる場合がある。こうした定型的な読み上げは、音声合成の特徴を活かして、作品ごとに差し替えが必要な個所のみをUIから入力したり項目を選択することで、簡単にカスタマイズして再利用できる。

入力文書中の編集や確認が完了すれば、最後に「出力」操作を行うことで、編集者が修正した読み情報を反映した音声合成による音声ファイルと、テキストと同期情報（SMIL ファイル）、インデックス情報などが自動生成され、DAISY形式としてダウンロードが可能となる。

5. 音訳関係者による実証実験

5.1 実用化に向けた施策・改良

本システムの開発にあたっては、まず同様の基本コンセプトで最低限の編集ができるプロトタイプシステム

長めのポーズ挿入

その中山から、少しはなれた山の中に、「ごん狐」 という狐がいました。

短めのポーズ挿入

その中山から、少しはなれた山の中に、「ごん狐」 という狐がいました。

ポーズの削除

その中山から、少しはなれた山の中に、「ごん狐」 という狐がいました。

図4 ポーズ指定の例

ムを構築した。評価作業対象として「青空文庫」からいくつか作品を抜粋し、実際に音訳作業を行った。

5.1.1 プロトタイプシステムによる音訳作業期間

各作品と作業ステップ数、対象とした作品の文書行数を表1に示す。

本作業の結果、大まかな傾向として、作業対象行の長さや作業ステップ数に、一定の相関が見られた。なお、本傾向から概算される結果として、“必要作業ステップ数=1.89×作品行数”が得られた。

なお、ここでの作業ステップ数とは、システム上の各行に対し、作業者の操作により、画面上のいずれかのGUI要素にフォーカスインまたはフォーカスアウトされた回数で構成されている。そのため、この回数は、実際の編集操作とは多少の誤差があり、たとえば、編集のアクション意図がなくGUI要素を選択した場合のカウント増加や、テキストエリアなど、一度フォーカスしたのちに、複数の文字入力でも1回とカウントされるため、実操作よりも少なくカウントされている可能性を含む。

5.1.2 実作業結果と1冊あたりの想定作業期間

また、異なる評価として、本システムを実際に利用して、15ページ相当の児童書の音訳を実施した。さらに、そこから類推される作業見積りを、一般に公表されている作業コストと比較した結果を図5に示す^{☆5}。

^{☆5} 作業項目「DAISY-level」は、テキストにDAISYの章立て構造を付与する作業である。

表1 予備評価結果

作品名（著者）	文書長 (行数)	作業 ステップ数
ごん狐（新美南吉）	213	508
我が輩は猫である（夏目漱石）	290	699
新版 放浪記（林芙美子）	398	940
白雪姫（グリム兄弟）	379	84
おいてけ堀（田中貢太郎）	93	137
おおかみと七ひきのこどもやぎ （グリム兄弟）	156	149
雨ニモマケズ（宮沢賢治）	62	119
鉛筆のシン（香俱土三鳥）	26	26

ここでの前提条件は、音訳対象が児童書1冊に相当する150ページ相当の図書（小学校5年生程度）、入力が文字起こし済みプレーンテキスト、出力はDAISY形式データの生成であるとした。なお、従来手法の制約条件は[17]に準ずる。また、本提案手法では制約として、以下を仮定している。

- 人の朗読作業は、週2回、各2時間に制限
- その他の作業は、1日あたり2時間までに制限

ここでの作業準備とは、作業の分担や編集上の協議、データの結合作業などを含む。なお、上述の制約条件は、日常の作業時間を比較的多めに割いている作業者を想定した設定である。

この事例から類推される結果として、従来の手作業では44日ほど必要となっていた作業が、7日ほどに短縮できる見込みが得られた。この期間は実例から類推された仮定値であり、今後厳密な検証を行う必要がある。

なお、実運用を想定した場合には、さまざまな変動要因が考えられるが、一般的には朗読作業による制約（同時並列化ができない、また、通常1話者に制限されることによる作業のボトルネック化）などで、従来手法はより期間が長くなる可能性がある。一方、音訳システムでは、作業を一人のみで行う想定であるため、作業の並列化による期間短縮が期待できる。

5.2 実証実験の概要

基本コンセプトに基づき、実証実験システムをWebアプリケーションとして開発し、2013年の秋から段階的に無償公開を行った。実証実験終了の2015年末までの公開期間中、国内48団体、1,029人に利用された[18],[19]。ログイン回数で延べ17,021回、また、4,722のユーザ文書がアップロードされた。

5.3 図書館、支援施設による作成事例

利用頻度が高い団体としては、公共図書館や障がい者支援施設のほか、当初は予定していなかった大学の実習教材（図書館司書課程）としての利用があった。4,722のユーザ文書を概観すると、平均の文書長で289.5行、一人あたりの平均的なユーザで5.3文書のアップロードが行われていた（最も多いアカウントで323文書）。システムログ分析によれば、平均的に1つの文書に対して238回のアクションが行われていることが分

Working contents	Preparation	DAISY-level	Recording	Page metadata	Correction	Building book	Total
Conventional method	8 days	8 days	600 min (3 weeks) =21 days	1 day	5 days	1 day	44 days
Proposed method	1 day	6 days					7 days

図5 本システム利用時の想定音訳作業時間

かった。これらのアクションとは、文字列の編集、ルビの編集、ポーズの挿入、音声の再生、メニューの選択、DAISY出力操作などのイベントに相当する。

5.4 作成図書の傾向と分析

公共図書館や障がい者支援施設での実利用を概観するため、各団体のスタッフにより作成された音訳文書から、13文書をランダムにピックアップしてその文書傾向を見た。文書の特徴は表2に示した通りである。

これらの文書を目視でカテゴリごとに分類し、単位文書長（10行）あたりに必要だった作業時間を図6にまとめた。

まずフロー文書であるが、これは自治体（図書館）が音訳ボランティアに依頼して地域居住者に配布するような文書である。平均行数は、517.5行、録音時間は音声合成の通常速度で読み上げた場合、61.5分である。仮定として、DaisyRingsを用いずにこの文書を従来の音訳手順で作業した場合には、朗読者の読み上げに3時間必要となり、DAISY図書として完成させるには合計6時間半程度かかることになる。しかし、実際には構成や読み修正等のステップを経るため、依頼から通常30日程度（1カ月程度の納期）で行われていたものである。

次にストック文書であるが、これは固有名詞や用語等の正確さが求められ、少なくとも読み誤りがないことが期待される文書である。今回の対象には、図書目録や図書館などで視覚障がい者が受けられるサービス

の案内文書などが含まれていた。文書サイズの平均は119.67行、音声ファイルの平均的な長さは12.11分であった。目録やサービス案内の朗読については、正確性は求められるものの、文書長が比較的短いものであること、また必ずしも専門性を必要としないことから、図書館から音訳者に依頼して作成していたものを、図書館のスタッフ自らが本システムにより内製することで、やりとりなどのオーバーヘッドを削減し、効率的に作成しようという試みが行われていた。

アーカイブ文書は、図書館の蔵書や教科書など、誤りが許されず、しかも一度作成されると適切に保存され、長い期間多くの人に参照されることを想定した文書である。今回の対象では、平均154.5行程度であり、担当個所の一部をトライアルとして利用されているケースが多い。

文書タイプによるコスト傾向として、アーカイブ文書では、音声再生回数の削減がキーとなることから、連続再生や速度を早めた再生、過去の音訳済みデータを活用した支援手段が有効だと考えられる。

一方、ストック文書は、再生回数と編集回数がほぼ同じで、一文について1.5回程度のアクション（再生／編集）がされていることから、言語解析による読みやアクセント付け、ポーズの自動挿入などが不十分であることが無視できない。

また、フロー文書は、再生時間よりも短い時間でDAISYコンテンツを作成されていることから、現在の時点でもDaisyRingsを効果的に活用できていることが明らかとなった。

5.5 編集作業に対する有効性評価

また、従来の音訳作業の経験に応じて、DaisyRings上

表2 評価対象の文書タイプと平均作業時間

文書タイプ	文書数	平均行数	平均再生時間（分）
フロー文書	6	517.50	61.5
ストック文書	3	119.67	12.11
アーカイブ文書	4	154.50	18.5

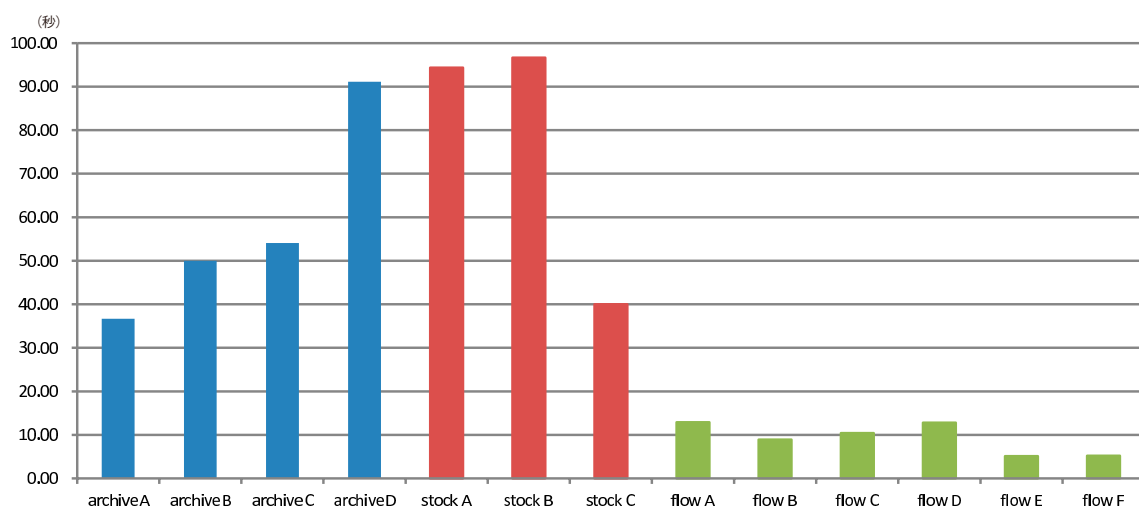


図6 文書別の10行あたりに必要な編集作業時間 (秒)

表3 同一ターゲット文書の音訳にかかる作業時間分析

	ルビ編集 (回)	ポーズ編集 (回)	作業時間 (秒)	再生時間 (秒)	作業時間／ 再生時間の 割合
経験者 (9名)	3.22	3.44	946.78	194.00	4.91
初心者 (10名)	6.70	22.00	1,487.10	199.10	7.50
合計 (19名)	5.05	13.21	1,231.16	196.68	6.27

でどのような作業の違いが生じるかを概観した。

被験者として、音訳経験者のグループ（9名）および、音訳作業の経験はないが一般的な文書作成などのPCのスキルを有している（10名）、合計19名を招集し、同じ小規模な音訳タスクを実施してもらい結果をみた。すべての被験者は、DaisyRingsを利用したことはなく、タスク実施前に30分ほどの簡単な使い方に関する講習を受けてもらった。

作業対象は、1,072文字からなる企業広報（プレスリリース）文書である。

結果を表3に記載した。この音訳作業では、読みの訂正、アクセントの訂正、ポーズ挿入などが主な作業となる。平均的に音訳経験者は再生時間の4.91倍で作業が完了しており、音訳初心者が7.5倍かかっているのに比べてより短い時間で編集作業を終えていることが分かった。特にルビ編集、ポーズ編集の回数にその差が現れており、経験者の場合には音訳として編集したり効果を加えるべき箇所とその落とし所が見えており、一方で初心者の場合は、丁寧に試行錯誤を加えた結果、編集回数が増加していることが分かった。いずれのユーザも本システムの扱いは初めてであることから、この差は音訳経験の違いに基づく見極めの差と考えられ、少なくとも経験者の知見の差からくる作業効率を妨げることなく、音訳作業の支援が可能となることが分かった。

6. おわりに

音声合成を活用した音訳支援システムであるDaisyRingsを開発した。本システムは、

- 音声合成の活用とWebブラウザベースのアプリケーションで音訳作業を支援。
- 全自動の音声合成で避けられない読み・アクセント誤りを、ルビベースの直感的な操作で簡単に修正可能。
- 業界標準のDAISY形式で出力することにより、そのまま再生・閲覧が可能なおほか、ほかのアプリ

ケーションやツールから作業ニーズに応じた再利用が可能。

といった特徴を持つ。

今後は、当社のクラウドサービスの1つ[20]として図書館関係者や音訳ボランティアの皆様を対象に、継続的／安定的にサービス提供を実現することで、音訳作業の効率化を図る。また、その結果、一人でも多くの音訳図書を必要とする人へ、タイムリーに正確な情報を伝えることの一助となれば幸いである。

参考文献

- 1) DAISY Consortium: DAISY Pipeline,
<http://www.daisy.org/project/pipeline> (2016年10月5日現在)
- 2) The Urakawa Project: A Toolkit for Accessible Multimedia,
<http://urakawa.sourceforge.net> (2016年10月5日現在)
- 3) Dolphin Computer Access Ltd.: Dolphin Publisher,
<http://www.yourdolphin.com/products.asp> (2016年10月5日現在)
- 4) Innovative Rehabilitation Technology: eClipseWriter,
<http://www.irti.net/eclipse/eClipseWriter/> (2016年10月5日現在)
- 5) Science Accessibility Net: ChattyInfty,
<http://www.sciaccess.net/jp/ChattyInfty/index.html> (2016年10月5日現在)
- 6) Shinano-kenshi Co., Ltd.: Producer,
<http://www.plextalk.com/jp/products/producer> (2016年10月5日現在)
- 7) 山口雄仁, 川根 深: 数式を含む文書の日本語読み上げ用試作システムについて, 信学技報, Vol.SP2001-74/WIT2001-28, pp.9-16 (2001).
- 8) 山口雄仁, 鈴木昌和: Daisy形式の日本語理数系教材が抱える諸問題, 信学技報, Vol.WIT2012-5, pp.23-27 (2012).
- 9) 岡本愛弓, 福島裕介, 矢入郁子: 音と触覚を用いた視覚障害児向け中学数学学習コンテンツの開発, 情報処理学会研究報告, Vol.2012-CE-117, pp.1-8 (2012).
- 10) 福田健太郎, 小林正明: HTML5標準案に基づく音声ガイドの提供の現状と課題～合成音声を活用した音声ガイドの普及に向けて～, 信学技報, Vol.WIT2011-86, pp.89-94 (2012).
- 11) 天野純子, 力宋幸男: ユーザフレンドリーなマルチメディア図書オーサリングシステム, 46(3), pp.715-727 (2005).
- 12) 都木 徹, 今井 篤, 清山信正, 世木寛之, 田高礼子, 田澤直幸, 岩鼻幸男: 話速変換技術・音声変換技術の放送および関連ビジネスへの応用, 情報処理学会研究報告, Vol.2012-SLP-93, No.6, pp.1-6 (2012).
- 13) 森田真弘, 田村正統, 布目光生: 多様な声や感情を豊かに表現できる音声合成技術, 東芝レビュー, Vol.68, pp.10-13 (2013).
- 14) 井床利生, 佐藤大輔, 畠山園子: アクセシブルな電子書籍の拡充へ向けてーテキストデイジー図書製作の効率化事例ー, 信学技報, Vol.WIT2011-90, pp.113-118 (2012).
- 15) 長妻令子, 福田健太郎, 柳沼良知, 広瀬洋子: クラウドソーシングを活用した効率良い字幕作成方法, 信学技報, Vol.WIT2012-25, pp.7-12 (2012).
- 16) 布目光生, 鈴木 優, 森田真弘: 電子書籍の論理構造に基づくポーズ情報の推定とSSML構造化, 情報処理学会研究報告デジタルドキュメント (DD), Vol.2011-DD-80(6), pp.1-7 (2011).
- 17) 文部科学省: 平成22年度「民間組織・支援技術を活用した特別支援教育研究事業」マルチメディアDAISY教科書アンケート結果抜粋別添1,

http://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afildfile/2011/09/09/1310526_3_1.pdf (2016年10月5日現在)

- 18) Fume, K., Kuroda, Y., Ashikawa, T., Mizuoka, Y. and Morita, M.: TTS-Based DAISY Content Creation System: Implementation and Evaluation of DaisyRings, Springer International Publishing, pp.69–76 (2014).
- 19) Fume, K., Kuroda, Y., Ashikawa, T., and Morita, M.: Practical Evaluation of DaisyRings: Text-to-Speech-Based DAISY Content Creation System, Springer International Publishing, pp.401–408 (2016).
- 20) TOSHIBA: Transliteration Editor DaisyRings,
<https://www.toshiba.co.jp/cl/pro/recaius/lineup/daisydings.html>
(2016年10月5日現在)

布目 光生 (正会員) kosei.fume@toshiba.co.jp

2001年北海道大学大学院工学研究科電子情報工学専攻修了。同年(株)東芝入社。知識・自然言語処理、音声応用処理、情報アクセシビリティに関する研究開発に従事。ACM会員。

黒田 由加 (正会員) yuka.kuroda@toshiba.co.jp

2002年慶應義塾大学大学院政策・メディア研究科修了。同年(株)東芝入社。機械翻訳・音声合成の研究開発に従事。

芦川 平 (非会員) taira.ashikawa@toshiba.co.jp

2003年九州大学大学院システム情報科学府知能システム学専攻修了。同年(株)東芝入社。知識メディア・音声認識技術の研究開発に従事。

森田 真弘 (非会員) masahiro.morita@toshiba.co.jp

1996年大阪大学大学院基礎工学研究科物理系専攻(生物工学分野)前期課程修了。同年(株)東芝入社。音声合成の研究開発に従事。日本音響学会会員。

採録決定：2016年10月5日

編集担当：住田一男 ((一社) 人工知能学会)