

# LLMを用いた研究論文の会話形式ポッドキャストへの変換システム PaperWave の開発

矢作 優知<sup>†,††</sup> 中條麟太郎<sup>†</sup> 原田 悠我<sup>†††</sup> 韓 燦教<sup>†</sup> 杉山 昂平<sup>†</sup>  
苗村 健<sup>†</sup>

<sup>†</sup> 東京大学大学院 情報学環・学際情報学府 〒113-0033 東京都文京区本郷 7-3-1

<sup>††</sup> 日本学術振興会特別研究員 DC

<sup>†††</sup> Educe Technologies

E-mail: <sup>†</sup> {yahagi,chujo,hanc,naemura}@nae-lab.org, sugiyama\_kohei@iii.u-tokyo.ac.jp, <sup>†††</sup> yugaharada@gmail.com

**あらまし** 音声を聴くことで、他のタスクをしながら知識に関与することが可能になる。本研究では大規模言語モデル (LLM) を活用することで、研究論文のような人間の制作者が扱いきれないニッチなコンテンツを音声で提供する可能性を探る。テキストの読み上げではなく、2人の話者が会話により論文を紹介する形式を採り、「ながら聞き」しやすいコンテンツを目指した。本発表ではシステムの実装などを報告する。

**キーワード** ポッドキャスト, 研究論文, 大規模言語モデル (LLM), 自伝的デザイン, フィールド調査

## Developing PaperWave: A System for Adapting Research Papers into Conversational Podcasts with LLMs

Yuchi YAHAGI<sup>†,††</sup>, Rintaro CHUJO<sup>†</sup>, Yuga HARADA<sup>†††</sup>, Changyo HAN<sup>†</sup>, Kohei SUGIYAMA<sup>†</sup>, and  
Takeshi NAEMURA<sup>†</sup>

<sup>†</sup> Interfaculty Initiative in Information Studies / Graduate School of Interdisciplinary Information Studies, The  
University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

<sup>††</sup> Doctoral Course Research Fellow of Japan Society for the Promotion of Science

<sup>†††</sup> Educe Technologies

E-mail: <sup>†</sup> {yahagi,chujo,hanc,naemura}@nae-lab.org, sugiyama\_kohei@iii.u-tokyo.ac.jp, <sup>†††</sup> yugaharada@gmail.com

**Abstract** Listening to audio content allows people to engage with knowledge while performing other tasks. In this study, we explore the potential of using large language models (LLMs) to provide niche content, such as research papers, in audio form—content that human creators find challenging to adapt. Instead of simply reading the text aloud, we have adopted a format where two speakers conversationally introduce the paper, aiming to create content that is easy to listen to while multitasking. In this presentation, we will report on the implementation of the system and other related aspects.

**Key words** Podcast, Research paper, Large language models, Autobiographical design, Field study

### 1. はじめに

音声コンテンツは、知識に関与する活動を〈モバイル読書〉(mobile reading)として行うことを可能にする [1]。一般的な〈定位置型読書〉(stationary reading)である文字を視覚的に読む読書とは異なり、〈モバイル読書〉は他の作業と並行して知識に関与すること、すなわち、ながら聞きを可能にする [1]。モバイル端末やストリーミングサービスの普及により、オーディオブックやポッドキャストなどのオーディオコンテンツの利用

は、知識に関与する方法の1つとして受け入れられるようになってきた。特に会話形式のコンテンツは聴取者の内容理解を促すことができ、その効果はマルチメディア学習の研究で示されている (e.g. [2], [3])。そこで本研究では、会話形式ポッドキャストに注目する。これは、さまざまなデバイスからオンデマンドでアクセス可能な複数話者の対話音声が収録されたオーディオコンテンツを指す。ポッドキャストの厳格な定義とは異なるが [4], [5]、気軽に聴取できる会話形式コンテンツという点での類似性から、本研究ではこのように呼称する。



図 1 (A) PaperWave のエピソードページ。ユーザーはポッドキャストを選択して再生できる。(B) 8月18日に第1著者(A1)が草刈りをしながら PaperWeaver [6] のポッドキャストを聞いている。(C) 8月29日に A1 が犬の散歩をしながら MedPodGPT [7] 聞いている。

しかし人間の制作者がすべてのニッチな文書 (e.g. 学術論文や行政の報告書) を会話形式ポッドキャストに変換することは現実的ではない。従って、提供されるコンテンツは多くのリスナーがいるトピックか、制作者が関心を持つトピックに限られる。この課題を解決し、ニッチな文書に関与する方法を多様化するには、コンピュータを用いた自動変換が不可欠である。Text-to-Speech (TTS) 技術だけでは、オーディオブックのような文章をそのまま読み上げたコンテンツ [8] しか生成できない。大規模言語モデル (LLM) を用いると、テキスト文書から会話の台本を生成することが可能になり、この課題を克服できる。

LLM を用いた文書-音声変換システムは商用サービスとしていくつかリリースされている。2023 年には Wondercraft と recast が [9], [10], 2024 年 9 月には Google の NotebookLM の Audio Overview がリリースされ [11], [12] オープンソースのプロジェクトも活発に開発されている (e.g. [13], [14])。研究としては、Saadya と Davis の Wondercraft を使用した医学教育の事例報告があるが [15], [16], 技術以外の面でのデザインに関する詳細な知見は依然として不足している。このような活発な開発状況を踏まえると、ニッチなコンテンツに関与する方法の選択肢を追加するために、文書-音声変換システムのデザインを模索することは時宜を得ているといえる。

そこで、本研究では LLM を用いた文書-音声変換システムが持つ可能性を、プロトタイピングと実践的な利用を通じて探索する。我々の研究は、2024 年 6 月 25 日に開始され、9 月 2 日に一旦の区切りを迎え、その後も続く PaperWave の集中的な開発と使用という実践に基づいている。この過程で、対象とする文書を我々自身が頻繁に関与する研究論文に設定した<sup>(注1)</sup>。そして自伝的デザイン (autobiographical design) のアプローチ [17], [18] を参考に研究を実施した。プロトタイプを自ら開発することで、

システムのより広範な部分を変更しながらデザインを探索したり、システムの利用ログを収集したりすることが可能となる。具体的には、1) ポッドキャストの長さを指定可能にする機能、2) フィールド調査の参加者の第 1 言語に合わせた日本語と韓国語のポッドキャストを生成する機能、3) 変換ログを記録しデータ分析の際に変換条件や出力結果を確認可能にする機能、などを実装して活用した。また、自伝的デザインの特徴は、システムの開発者自らがユーザーとなってシステムを使用し、使用と改善を繰り返しながらデザイン空間を探索することにある。これにより、適切な対話形式とポッドキャストの作成および再生インターフェースの両方のデザインに関する物語を記述し、生活に密着した視点からの洞察を得ることができる。

まとめると、本研究の目的は、LLM を使用した文書-音声変換システムが、人々がニッチなテキスト文書と関与する機会をどのように広げるかについて、より詳細な理解を提供することである。そして、システムをデザインする際の考慮事項を示すことで、LLM を活用したシステムのデザイン指針に関する知見を拡張する。本発表では特に、プロトタイプの開発に焦点を当て、自伝的デザインがどのようにデザインの探索に貢献したかを考察する。実践的な利用から見出されたデザイン上の考慮事項について詳しくは [19] を参照されたい。

## 2. 関連事例

本研究の目的は、LLM を応用したシステムのコンテンツ形式やインタフェースに関するデザイン上の考慮事項を示すことであり、LLM に関する技術的貢献をすることではない。そのため、本論文は PaperWave と他の文書-音声変換システムの技術的な差異には注目しない。それでも、PaperWave と関連事例の比較は、フィールド調査からの示唆をより正確に理解するために必要だろう。フィールド調査の文脈を提供するために、本章では機能に注目して PaperWave の位置付けを整理する。

文書から音声コンテンツを生成するシステムには様々な種類

(注1) : PaperWave は、テキストでの文書の読解を補うことを意図して開発している。原文の批判的な読解は、深い理解を得るために不可欠である。

がある。まず、生成される音声は独白形式であるか、会話形式であるかで分類できる。独白形式のコンテンツには、文書に含まれるテキストをそのまま読み上げたオーディオブック型のコンテンツがある [8], [20]。TTS 技術を用いるとこの形式の音声は文書から自動生成できる。研究論文をオーディオブック型の音声で「聴く」機能は、Taylor & Francis や ProQuest などのオンラインライブラリで提供されている [21], [22]。独白形式でも LLM を用いたアプローチはある。489.ai は医学系論文を LLM で要約し、出力された要約文を TTS で読み上げて音声コンテンツにしている [23]。489.ai では複数種類の音声（話者）を用いているが、話者同士の会話形式ではない。要約の原稿を人間のライターが作成し TTS で読み上げる論文紹介ポッドキャストもあるが、それらも医学分野の論文を対象としている [24], [25]。要約のアプローチは、より短時間で多くの論文に関与することを目指したものである。

会話形式のコンテンツは、複数の話者が対話する形式である。先駆的なプロダクトとしては、2023 年に公開された Wondercraft [16] と recast [10] がある。Wondercraft は音声コンテンツ制作ツールで、論文紹介ポッドキャスト (Research Paper to Podcast) 用のプロンプト例を持つ。Wondercraft は自動生成というよりも、自動的な処理の後に人間のクリエイターが編集を加える作業を支援することを重視している。PaperWave では各ポッドキャストをユーザーが毎回手間をかけて調整するよりも、手軽に聞けることを重視している。recast は記事をポッドキャストにするシステムとして紹介されており、研究論文に特化しているわけではない。生成される音声は 2, 3 分程度の長さであり、テキストを読む場合よりも何分短縮できるかを表示する UI があるなど、短時間での情報取得を志向している。PaperWave はながら聞きを志向しているため、長時間のポッドキャストも生成できるように実装している。

2024 年にはより発展したシステムが登場する。Google は 2024 年 9 月 12 日に、入力した文書の概要を紹介する会話形式ポッドキャストを作成する *Audio Overview* 機能を NotebookLM に追加した [11], [12]。公開当初はユーザーは生成内容を一切調整ができなかったが、開発チームはすぐに調整の必要性に気づき [26]、投稿日 (2024 年 10 月 25 日) 時点では「AI ホストが焦点を当てるべきこと」を入力できるようになっている [12]。また、Google は Illuminate という研究論文をポッドキャストに変換する実験的なプロダクトも発表しているが、投稿日時点では招待されたユーザーのみが利用できる状態であり、一般公開はされていない [27], [28]。現状では、これらのシステムではポッドキャストの言語が英語に限られている。また、オープンソースプロジェクトもさまざま存在し、例えば pdf-to-podcast [14] は 2024 年 6 月 9 日に公開された。またそのフォークである PDF2Audio [13] は日本語でも使用されている [29]。

このように、文書-音声変換システムは現在活発に開発されている。システムの機能が多様であるだけでなく、音声に変換する目的もさまざまである。このような状況下で、文書-音声変換システムのデザインに関する議論を前進させるには、ユーザー体験を視野に入れて検討することが重要である。そして、多様

なユーザーが各自の目的に応じてシステムを利用すること、あるいは利用しないことを詳細に記述することが求められる。そこで、特定の課題解決を目指すのではなく、「文書を音声に変換することが、知識との関与や生活にどのような変化をもたらすのか」という問いをシステムを作りながら探索的に検討する。本研究は、様々なシステムが活発に開発されている時期に、いち早く実践的な利用を元にデザイン上の考慮事項について検討したものである。

### 3. 方 法

本研究のフィールド調査 (著者らによる自伝的デザインのプロセスを含む) には、5 人の著者 (A1-A5) と 6 人の研究協力者 (P1-P6) が参加した。本稿では両者をまとめて、研究参加者と呼ぶ。本論文では、2024 年 6 月 25 日から 9 月 2 日までの期間に行われた PaperWave の開発と使用を議論の対象としている。調査の段階は大きく分けて 3 段階あり、1) 6 月 25 日から 7 月 9 日までに ChatGPT を用いたプロンプトの探索、2) 7 月 1 日から 8 月 21 日までにコマンドラインインタフェースの使用、3) 8 月 22 日以降にウェブアプリケーションの使用を行った。著者らは調査の始めから参加した。研究協力者は 8 月前半からエントリーし 2, 3 段階目に参加した。9 月 1 日と 2 日にはデザインワークショップを開催し、PaperWave の使用経験の振り返りと PaperWave の改善案を考える活動を行った。各日の内容は同一で、研究参加者はどちらか一方に参加した。研究参加者は、調査期間中の経験を日記として Slack に記録した。その結果、合計で 1259 件のメッセージが記録され、分析に使用された。

本研究では、HCI 分野で成熟しつつある自伝的デザイン [17], [18], [30] の方法を参考に文書-音声変換システムのデザインを探索している。この手法では、システム開発者自身がユーザーとなり、デザインについて詳細かつ微妙なニュアンスを理解する [18]。著者らは、言語の壁、研究と家事・育児の両立、学際的な研究における未知の分野への関与の難しさなど、論文を読む上での課題に直面している。こうした課題に直面する中で、PaperWave のデザインを〈真の使用〉 (*genuine usage*) [17] を通して探求した。[31], [32] が示したように、このアプローチにより、集中的な使用を通じて日常生活と密接に関連するデザインを探索することが可能になる。さらに、最新の LLM を活用した PaperWave の事例は、このアプローチによる〈素早いティンカリング〉 (*fast tinkering*) によって促進された〈初期のイノベーション〉 (*early innovation*) の好例である [17]。この研究における 2 か月という使用期間は長期間とは言い難いが、集中的な使用を伴う一人称アプローチを積極的に取り入れ、さもなければ隠蔽されてしまうデザインプロセスと研究者の影響を記述する [17], [30], [33]。著者らのバイアスに対処するため、先行研究の助言に従い非ユーザー (A5, A6) と二次ユーザー (P1-P6) とともに PaperWave を使用した [17]。A1 と A2 が PaperWave の開発を行い、その間、著者らは毎週ミーティングを開いて改善点や問題点を話し合った。

## 4. PaperWave

### 4.1 デザイン指針

自伝的デザインを基に我々は PaperWave の暫定的なデザイン指針を得た。これらの指針は潜在的な利点と原則を概説する。

#### a) 会話コンテンツとその形式

[FM1] は、ユーザーが一部を聞き逃しても内容を把握できるようにする。[FM2] は理解と関与を高める。[FM3] には、コマースやプレゼント企画のアナウンスなど論文と無関係なコンテンツの除外が含まれる。

[FM1] インタビュー形式を採用し、ホストによる質問や言い換えを含める

[FM2] 概要と内容の構成の提示を番組冒頭で行う

[FM3] 論文に書かれた内容を正確に伝える

#### b) ポッドキャスト作成と再生インターフェース

[IF1] は〈モバイル読書〉を可能にする。[IF2] は、ユーザーがポッドキャストを時間の制約に合わせて調整できるようにする。[IF3] は、新鮮さを与えエピソードの切り替わりを明確にする。[IF4] は、さまざまなトピックへの関与を促進する。

[IF1] デスクトップとモバイルの両方に対応する

[IF2] ポッドキャストの長さを指定できる

[IF3] BGM を再生する

[IF4] 作成したエピソードを同僚と共有できる

### 4.2 システム概要: ユーザーシナリオの例

ユーザーが PaperWave を使用するシナリオを紹介することを通して、機能を概説する。説明に用いるシナリオは、ユーザーが大学への徒歩と電車での通学中に、PaperWave で論文を聞くというものである。身支度を始める前に、ユーザーは論文を PaperWave にアップロードし、歩きながらスマートフォンの PaperWave でその論文を聴く。そして、電車に乗ってから、別のユーザーが作成したポッドキャストを選択して聞く。この一連の流れの中で、〈収録〉機能、エピソード再生機能、〈チャンネル〉リスト機能を紹介する。

#### a) 身支度の前に〈収録〉する

ユーザーは論文を保管しているパソコンで、図 2 の画面から PDF ファイルをアップロードし、会話形式のポッドキャストを作成する。技術的には、複数の研究論文をアップロードして 1 つのポッドキャストを作成することも可能だが、複数本アップロードしたユーザーはいなかった。我々は、この変換処理をポッドキャスト制作のメタファーとして〈収録〉と呼んでいる。まず、ユーザーは録音する PDF ファイルをアップロードする (A)。次に、エピソードのタイトルを入力し (B)、エピソードの長さを設定する (C)。ユーザーは自宅と最寄り駅が徒歩 15 分程度の距離にあるため、15 分と指定した。このように、自身の都合に合わせて時間を指定できる。さらに、ポッドキャスト

で話者が話す言語 (D)、生成に用いる LLM モデル (E)、その他のオプション (F) を選択する。言語は元の論文の言語に関わらず、英語、日本語、韓国語を選択できる。最後に、ユーザーは「生成」ボタンをクリックして、〈収録〉を開始する。

#### b) 駅までの道のりでポッドキャストを再生する

ユーザーは〈収録〉中に、身支度を完了させる。駅まで徒歩で向かう間、ユーザーはデバイスをスマートフォンに切り替えて、図 1 (A) のモバイル用画面から〈収録〉したポッドキャストを聴く。様々なデバイスからアクセスできることで、ユーザーは家事や移動などの日常のタスクをしながら、さまざまな場所で論文をながら聞きできる。

#### c) 〈チャンネル〉リストで他の人のポッドキャストを聴く

時間指定して〈収録〉したポッドキャストは、ちょうど駅に到着する頃に聴き終わる。その後、ユーザーは図 3 の〈チャンネル〉ページから、他の研究室メンバーが作成したポッドキャストを探す。この機能により、ユーザーは似た興味を持つ他者が作成したコンテンツを聴くことができ、幅広いトピックに気軽に触れることができる。ユーザーは前日に研究室で話題になっていた論文を見つけ、電車の中でそれを聴く。以上が、フィールド調査での経験を参考に作成した PaperWave の使用例である。

### 4.3 バックエンドスクリプト (PaperWave CLI) の実装

バックエンドスクリプト (PaperWave CLI)<sup>(注2)</sup> は、研究論文を会話形式のポッドキャストに変換する役割を担っている。このスクリプトは、コマンドラインまたはウェブアプリケーションから実行できる。実装では、LLM として OpenAI の GPT を使用した。図 2 に示すように、モデルはユーザーが選択できる。アシスタントは OpenAI の Assistants API を使用して実装されている。アシスタントは JSON Schema で指定された構造化出力を生成するように指示されている。アシスタントは PDF コンテンツの取得のために File Search を利用している。また、TTS も OpenAI の API を使用した。実装には TypeScript を用いた。

図 4 のように PaperWave CLI は 3 つの主要なアシスタントで構成されている。〈構成作家〉は、チャプターの構成と取り扱う話題のリストを生成することで、台本をチャプターに分割し、長時間コンテンツの生成にも対応する。話題のリストによって内容の重複を防ぎ、エピソード全体の一貫性を確保する。同時に、〈情報抽出器〉は論文のタイトルと著者を取得する。これらのアシスタントからの出力は〈脚本家〉に渡され、構成に基づいてチャプターごとにホストとゲストスピーカーの対話を生成する。完成した台本は、TTS を使用して音声に変換され、BGM とミックスされて最終的な音声ファイルが作成される。

自伝的デザインの過程でポッドキャストの長さと言語を指定可能にする必要性が浮かび上がり、機能が実装された。ポッドキャストの長さは会話のターン数によって制御する。8 月 3 日までに作成した日本語のポッドキャストに含まれるターン 1557 個から、1 ターンの平均長さ (13.033141 秒) を求めた。ユーザーが指定した長さをこの平均値で除することにより、必要なターン数を計算する。全体のターン数を〈構成作家〉が各チャ

(注2) : <https://github.com/nae-lab/paperwave-cli>



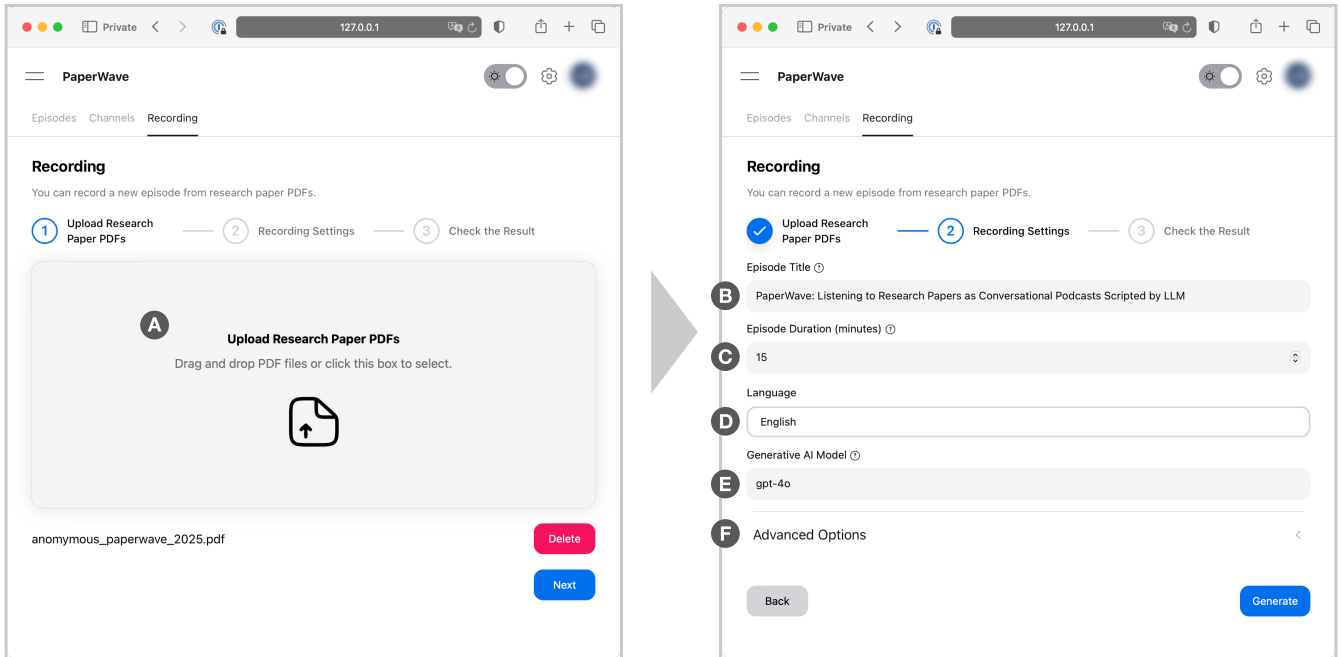


図2 〈収録〉ページ. 次の項目を設定可能である：(A) 変換したい PDF ファイル (B) タイトル (C) 再生時間 (D) 話者の言語 (E) スクリプトを生成する LLM モデル (F) 詳細オプション (エピソードの説明, キーワード, カバー画像の URL)

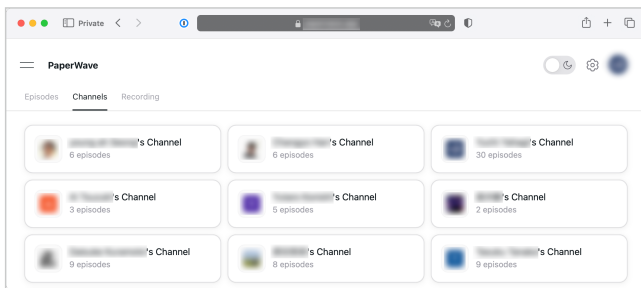


図3 他のユーザーのポッドキャストを一覧できる〈チャンネル〉ページ

プターに割り振り、〈脚本家〉は各チャプターの台本を指定ターン数で作成する。言語の指定は、LLM に与えるプロンプトの言語を、選択された言語のものに切り替えることで実装した。日本語と英語のプロンプトは A1 が、韓国語のプロンプトは韓国語が第 1 言語である A4 が作成した。

#### 4.3.1 ウェブアプリケーション

ウェブアプリケーション<sup>(注3)</sup>は、React フレームワークである NextJS と、UI ライブラリ NextUI で実装されている。バックエンドの PaperWave CLI は Docker を使用して Google Compute Engine 上にホストされている。ユーザーが〈収録〉をリクエストすると、オプションは Firestore を介して通信され、バックエンドスクリプトがポッドキャスト音声ファイルを生成する。生成されたポッドキャストは Cloud Storage for Firebase に保存され、ユーザーはウェブアプリケーションからアクセスできる。

(注3) : <https://github.com/nae-lab/paperwave-web>

## 5. 結 果

[19] で報告した結果について、ここでも概要を示す。

### 5.1 PaperWave は論文の〈モバイル読書〉を可能にする

参加者は PaperWave をさまざまな状況で使用した。移動中には、ほぼすべての参加者がポッドキャストを聴き、電車やバスなどの公共交通機関を利用したり、歩いたり、運転したりした。また、家事や育児、身支度、事務作業、ビデオ編集をしながらもポッドキャストを聴いた。休憩時間も聴取の機会となり、待ち時間や就寝前、食事中、マッサージチェアに座っているとき、仕事の休憩中に聴いた。PaperWave は、文章を読む〈定位置型読書〉が難しい状況で研究論文に関与することを促進していた。一方で参加者は課題も感じており、他の作業をしながら聴くと内容の理解が難しいと感じることがあった。

### 5.2 PaperWave は知識への関与の仕方を変える

参加者の第 1 言語でポッドキャストを作成することは、特に英語が得意でない人にとって、研究論文への関与のハードルを下げた。〈モバイル読書〉が可能になったことで、論文を読むためにまとまった時間を活用しないといけないという心理的負担が軽減された。また、他者から共有された論文をインプットする手段として、PaperWave が特に有効であると感じていた。これにより、従来よりも広範なトピックの論文に関与できたと話す参加者もいた。

参加者が論文をポッドキャストとして聴いたとき、文書を視覚的に読む場合とは異なる部分に注意が向いた。テキストでは読み飛ばしていた情報が、聞くことで重要性を理解することがあり、論文をよりバランスの取れた方法で読むことができたと感じていた。マルチタスクが可能になり、予算申請書の執筆を



図4 PaperWave CLIの実装。PDFの入力から出力のポッドキャストが得られるまでの処理過程を示している。枠線内がLLMに与えた指示で、灰色の網掛け内が出力のJSONである。

しながら、申請書で引用したい情報を含みそうな論文を流し、対象の論文を探し出すという使い方をした参加者もいた。対照的に、明確な目的なしにポッドキャストを聴くことで、新しい研究アイデアが生まれることもあった。ポッドキャストの形式は、参加者に著者との個人的なつながりを感じさせ、著者に対する親近感を生むこともあった。

PaperWaveのポッドキャストでは、参考文献はほとんど言及

されず、主に論文の本論に焦点が当てられた。この特徴により、参加者は単一の論文を大まかに理解することができたが、文脈情報の欠如を懸念する参加者もいた。また、視覚情報の欠如は音声コンテンツの本質的な限界であり、特にデザインのような視覚要素が重要な分野では、内容を把握するのが難しいと感じることがあった。また、特に自分の専門分野外のコンテンツを聴取する際、不正確さに対する懸念が高まった。

## 6. 考 察

PaperWave は〈モバイル読書〉を促進し、研究参加者がさまざまな状況で研究論文に関与することを可能にしていた。本研究で見られた文書から音声への変換の影響は、オーディオブックの事例と符合している [1], [34]。Tattersall-Wallin と Nolin は、オーディオブックへの関与を余暇活動として調査したものの、オーディオブックが勤務時間中にも聞かれていることを示唆しているが [1]、本研究はその逆を示している。すなわち、職業活動に関連する論文への関与が私生活に埋め込まれる可能性を示唆した。さらに、参加者の活動はしばしばポッドキャストの聴取と相互に影響していた。例えば、P2 はポッドキャストを聴きながら予算申請書を書くことができ、独特なマルチタスクの方法を示した。別の例として、A3 が「ARECA: A Design Speculation on Everyday Products Having Minds」 [35] を聴きながら散歩していたところ、自動販売機の横を通り過ぎ、自動販売機が心を持つことの影響を想像した (7月29日)。A3 は論文のコンセプトを自動販売機のデザインにその場で応用したのである。この例は、物理的な世界に存在しながら学術論文に関与することを可能にする〈モバイル読書〉の利点を強調している。これらの結果は、文書の読解支援システムがコンテンツのみに焦点を当てるのとは異なり (e.g. [6], [36]~[38])、周囲の環境との相互作用を考慮して文書-音声変換システムをデザインする重要性を示している。この結論は、実験室実験ではなく日常生活における実践を調査するアプローチの必要性も裏付けている。

テキストを音声に変換することは、学術論文への関与のハードルを下げ、内容の理解の仕方を変える可能性を示した。これは、会話形式のコンテンツが理解を助けることを示す先行研究と一致している [2], [3]。また、論文の概要を提供することが馴染みのない分野の論文へのアクセスを容易にするという点は、スキミング支援システムの利点と一致している [36]。しかし、PaperWave は必ずしも時間短縮を達成しない点は異なる。ながら聞きのためのデザインを行ったことで、冗長性を含む会話形式のコンテンツを生成するシステムになった。必ずしも効率化を達成しない、要約ではなく冗長性を増す変換という機能がもたらす微妙な (nuanced) 有用性は、我々が PaperWave を使用する中で徐々に浮かび上がった特徴である。

ハードルが下がることや理解の仕方の変化には、言語の違い、誰が話者であるか、ポッドキャストで強調する部分など、様々な会話の形式から影響を受けることが示唆された。今回のフィールド調査ではすべての参加者が共通のシステムを利用していた。自伝的デザインの過程で繰り返しシステムが修正されたとはいえ、1通りの形式に収束する探索しか行えていないことは本研究の限界であり、会話の形式についてはまだ未開拓の広大なデザイン空間が広がっている。LLM のプロンプトをユーザーが指定できるようにして、よりカスタマイズ可能にすることは限界を乗り越える 1つの方向性である。

本研究では、一人称アプローチを採用し、論文への関与についての反省を促進した。結果で述べたように、論文の本論に焦点を当てることは、文書-音声変換システムの有用性の評価を左

右する要素であった。A3 と A5 は有用性の評価が特に鮮明に別れた 2 人の参加者だった。2 人はワークショップ終了後、論文執筆段階でさらに自伝的な省察を深めた。企業に勤務し、趣味あるいは同僚との 1on1 でのアドバイス提供のために論文を読む A3 は、論文に書かれた知見の部分的な抽出と日常生活への応用にも価値があると考えている。A3 は論文を分割可能な部品で構成された知識と捉え、関与によって得られるインスピレーションを優先している。この態度により、変換による潜在的な不正確さも受け入れている。一方、社会学の背景をもち生涯学習の研究に取り組む大学の研究員である A5 は、研究論文の IMRaD 構造を愛しており、PaperWave の品質が向上しても、元の文書のままで読むことを好むだろうと考えている。この差異は、文書-音声変換システムに対する異なる態度と、その背景にある個人と研究論文の関係性を示している。このように、自伝的デザインを通じて研究論文との新しい関与の形を探求することで、知識にどのように関与したいかについての各研究者の個人的な態度が浮き彫りになった。この過程には、実験的に論文の文脈に焦点化したポッドキャストを生成する指示を与えて試用することも含まれていた。システムの開発をしながら集中的に使用する中で、参加者たちは自分自身と研究論文との関係性を省察することができた。

## 7. 結 論

本研究は、LLM (大規模言語モデル) が研究論文との関わり方をどのように変化させる可能性があるかを示した。文書を会話形式のポッドキャストに変換することで、PaperWave は知識に関与するためのより柔軟で敷居の低い手段を提供する。調査結果は、PaperWave が効果的に利用できる多様な文脈を示し、ユーザーと環境との相互作用を考慮する必要性を示唆した。

しかし、この研究にはいくつかの限界がある。まず、本研究の結果は今回の調査の文脈に依存しており、一般化できない。また、Slack メッセージを送信する日記の記録方法では、PaperWave が使用されなかった事例を完全に捉えられなかった可能性がある。研究参加者は非使用報告を意図的に記録していたが、より詳細な検討のためには定期的な日記記録が必要である。さらに、すべての参加者が同じプロンプトから生成されたポッドキャストを聴取したため、パーソナライズには限界であり、さらなる探求が必要である。それでも、萌芽的な技術がもたらす変化を迅速に調査し、文書、音声コンテンツ、ユーザーの知識への関与の微妙な関係を示したことは、文書-音声システムのデザインを検討する上で有用な知見を提供できると、我々は信じている。

参加者からは、使用中に直面した課題に対処するための提案も行われた。そこには、PaperWave に論文のためのコンテンツ推薦やマルチモーダル拡張などの読解支援技術を統合する提案も含まれていた [6], [38]。HCI 分野のこれらの研究と文書-音声変換システムを接続することは、今後の課題である。

**謝辞** 本研究は JSPS 科研費 JP22KJ1010 と中山未来ファクトリーの助成を受けた。フィールド調査で示唆に富むフィードバックを提供して下さった研究協力者の皆様に感謝する。

## 文 献

- [1] E. Tattersall Wallin and J. Nolin, "Time to read: Exploring the timespaces of subscription-based audiobooks," *New Media & Society*, vol.22, no.3, pp.470–488, March 2020.
- [2] M.T.H. Chi, S. Kang, and D.L. Yaghmourian, "Why Students Learn More From Dialogue- Than Monologue-Videos: Analyses of Peer Interactions," *Journal of the Learning Sciences*, vol.26, no.1, pp.10–50, Jan. 2017.
- [3] R.E. Mayer, S. Fennell, L. Farmer, and J. Campbell, "A Personalization Effect in Multimedia Learning: Students Learn Better When Words Are in Conversational Style Rather Than Formal Style," *Journal of Educational Psychology*, vol.96, no.2, pp.389–395, June 2004.
- [4] A.J. Bottomley, "Podcasting: A Decade in the Life of a "New" Audio Medium: Introduction," *Journal of Radio & Audio Media*, vol.22, no.2, pp.164–169, July 2015.
- [5] C. Farivar, "New Food for iPods: Audio by Subscription," Oct. 2004. <https://www.nytimes.com/2004/10/28/technology/new-food-for-ipods-audio-by-subscription.html>
- [6] Y. Lee, H.B. Kang, M. Latzke, J. Kim, J. Bragg, J.C. Chang, and P. Siangliulue, "PaperWeaver: Enriching Topical Paper Alerts by Contextualizing Recommended Papers with User-collected Papers," *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp.1–19, CHI '24, Association for Computing Machinery, New York, NY, USA, May 2024.
- [7] S. Jia, S. Bit, E. Searls, L.A. Claus, P. Fan, V.H. Jasodanand, M.V. Lauber, D. Veerapaneni, W.M. Wang, R. Au, and V.B. Kolachalama, "MedPodGPT: A multilingual audio-augmented large language model for medical research and education," July 2024.
- [8] E. Tattersall Wallin, "Reading by listening: Conceptualising audiobook practices in the age of streaming subscription services," *Journal of Documentation*, vol.77, no.2, pp.432–448, Jan. 2021.
- [9] Wondercraft, Inc., "Find out who we are," [n. d.]. <https://www.wondercraft.ai/company/>
- [10] Fika Media, "This is recast". <https://www.letsrecast.ai/>
- [11] B. Wang, "NotebookLM now lets you listen to a conversation about your sources," Sept. 2024. <https://blog.google/technology/ai/notebooklm-audio-overviews/>
- [12] Google, "NotebookLM," 2024. <https://notebooklm.google/>
- [13] MIT Laboratory for Atomistic and Molecular Mechanics, "Lammmit/PDF2Audio," 2024. <https://github.com/lamm-mit/PDF2Audio>
- [14] S. Fitzpatrick, "Knowsuchagency/pdf-to-podcast," Oct. 2024. <https://github.com/knowsuchagency/pdf-to-podcast>
- [15] A. Saadya and C.R. Davis, "Revolutionizing Plastic Surgery Education: Leveraging AI for an Innovative Podcast Learning Platform.," *Plastic and Reconstructive Surgery*, vol.154, no.4, pp.847e–848e, March 2021.
- [16] Wondercraft, Inc., "AI Powered Audio Studio - Wondercraft," [n. d.]. <https://www.wondercraft.ai/>
- [17] C. Neustaedter and P. Sengers, "Autobiographical design in HCI research: Designing and learning through use-it-yourself," *Proceedings of the Designing Interactive Systems Conference*, pp.514–523, DIS '12, Association for Computing Machinery, New York, NY, USA, June 2012.
- [18] C. Neustaedter and P. Sengers, "Autobiographical design: What you can learn from designing for yourself," *interactions*, vol.19, no.6, pp.28–33, Nov. 2012.
- [19] Y. Yahagi, R. Chujo, Y. Harada, C. Han, K. Sugiyama, and T. Naemura, "PaperWave: Listening to Research Papers as Conversational Podcasts Scripted by LLM," Oct. 2024. <http://arxiv.org/abs/2410.15023>
- [20] S.S. Gonçalves and P.N. Silva, "Audiobooks in Information Science: A systematic literature review," *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, vol.22, p.e024014, Aug. 2024.
- [21] Taylor & Francis Group, "Taylor & Francis Online". <https://www.tandfonline.com>
- [22] ProQuest LLC., "ProQuest". <https://www.proquest.com>
- [23] 489.ai, "489.ai | 医師向け新着論文 AI 要約配信". <https://489.ai/>
- [24] 日経 BP, "日経メディカル 聴く論文". <https://medical.nikkeibp.co.jp/inc/all/hotnews/podcastpaper/>
- [25] メディカルレビュー社, "Quick Med Insights -耳で読む論文- ● A podcast on Spotify for Podcasters". <https://podcasters.spotify.com/pod/show/p5e72bttgb8>
- [26] R. Rogers, "グーグルの「NotebookLM」を使った AI ポッドキャストが注目されている," Oct. 2024. <https://wired.jp/article/ai-podcast-google-notebooklm/>
- [27] Y. Matias, "Google Research at Google I/O 2024," May 2024. <http://research.google/blog/google-research-at-google-io-2024/>
- [28] Google, "Illuminate," 2024. <https://illuminate.google.com/>
- [29] kun432, "PDF2Audio デモ," Sept. 2024. <https://www.youtube.com/watch?v=g-QiK2xvHcA>
- [30] A. Desjardins, O. Tomico, A. Lucero, M.E. Cecchinato, and C. Neustaedter, "Introduction to the Special Issue on First-Person Methods in HCI," *ACM Trans. Comput.-Hum. Interact.*, vol.28, no.6, pp.37:1–37:12, Dec. 2021.
- [31] D.A. Mellis and L. Buechley, "Do-It-Yourself Cellphones: An Investigation into the Possibilities and Limits of High-Tech DIY," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp.1723–1732, CHI '14, Association for Computing Machinery, New York, NY, USA, April 2014.
- [32] A. Desjardins and R. Wakkary, "Living In A Prototype: A Reconfigured Space," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp.5274–5285, CHI '16, Association for Computing Machinery, New York, NY, USA, May 2016.
- [33] C. Ellis, T.E. Adams, and A.P. Bochner, "Autoethnography: An Overview," *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, vol.12, no.1, pp.●–●, 2010.
- [34] E. Tattersall Wallin, "Audiobook routines: Identifying everyday reading by listening practices amongst young adults," *Journal of Documentation*, vol.78, no.7, pp.266–281, Jan. 2022.
- [35] H. Cho, J. Lee, B. Ku, Y. Jeong, S. Yadgarova, and T.-J. Nam, "ARECA: A Design Speculation on Everyday Products Having Minds," *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pp.31–44, DIS '23, Association for Computing Machinery, New York, NY, USA, July 2023.
- [36] R. Fok, H. Kambhamettu, L. Soldaini, J. Bragg, K. Lo, M. Hearst, A. Head, and D.S. Weld, "Scim: Intelligent Skimming Support for Scientific Papers," *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pp.476–490, ACM, Sydney NSW Australia, March 2023.
- [37] J.C. Chang, A.X. Zhang, J. Bragg, A. Head, K. Lo, D. Downey, and D.S. Weld, "CiteSee: Augmenting Citations in Scientific Papers with Persistent and Personalized Historical Context," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp.1–15, CHI '23, Association for Computing Machinery, New York, NY, USA, April 2023.
- [38] T.S. Kim, M. Latzke, J. Bragg, A.X. Zhang, and J.C. Chang, "Papeos: Augmenting Research Papers with Talk Videos," *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp.1–19, ACM, San Francisco CA USA, Oct. 2023.