

名詞句群短時間把握のための音声重畳再生手法の提案

池田 悠星¹ 土岐田 力輝² 齊藤 孝樹² 呉 健朗¹ 宮田 章裕^{1,a)}

概要：

視覚障害者が名詞句群を把握する一般的な手法として、スクリーンリーダを用いた音声再生が多く用いられている。しかし、従来手法では1つずつ連続的に音声出力される名詞句を聞き取る必要があり、名詞句群全体の把握に時間がかかる。たとえば、レストランのメニューを音声読み上げで把握したい場合、視覚障害者は全メニューが読み上げられるまで長時間待ち続けるか、メニューの全体を把握することを諦めないとはいえない。この問題を解決するために本稿では、視覚障害者が名詞句群全体を短時間で把握できるようにすることを研究課題とする。そして、この課題を達成するために音声時間差重畳再生手法を提案する。具体的には、名詞句群に含まれるある名詞句の音声再生されている最中に、次の名詞句の音声を部分的に重ねて再生する手法である。今後の研究では、実験参加者に時間差重畳音声を聞き取らせる対照実験を行い、定量評価ではリッカート尺度を用いたアンケートを実施し、定性評価では音声の把握のしやすさに関する自由記述を行うことによって本提案手法の有用性を確認する予定である。

1. はじめに

晴眼者は視覚的な手段を用いて効率的に情報を取得することが可能である。一方で、視覚障害者はその障害の程度に応じて、視覚による情報取得が困難な場合がある。そのため、視覚障害者は外界の状況や文字情報を把握するための手段として、主に聴覚や触覚を用いている。このような違いにより、視覚障害者は晴眼者よりも情報取得の過程で多くの制約や困難を抱えている。ゆえに、視覚障害者と晴眼者との間には情報格差が生じている。この格差は単に情報へのアクセスの難しさだけでなく、得た情報を処理する過程においても表れている。

例えば、情報へのアクセスにおいては、視覚障害者が公共交通機関を利用する際に時刻表や案内板を視覚的に読むことが困難であるため、目的の電車やバスを逃してしまうことがある。また、情報を処理する過程においては、視覚障害者は膨大な情報を処理する際にスキミング（飛ばし読み／斜め読み）と呼ばれる重要な部分やキーワードを拾って素早く概要を把握する非線形的な読み方が難しく、文書を一文ずつ線形的に読まざるを得ないため、晴眼者と比較して情報理解や処理が困難である。特に、レストランのメニューなどの名詞句群全体を把握する場面では、名詞句を1つずつ順番に聞き取る必要があるため、効率的な情報取

得が困難である。ある視覚障害者は、レストランでの商品注文時に、他者にメニューを読んでもらう手間を避けるため、毎回同じ商品を注文すると述べていた。このような状況は視覚障害者の食事内容の選択肢を狭めるとともに、晴眼者との情報格差の拡大につながる。

そこで我々は、名詞句群を短時間で把握するための音声時間差重畳再生手法を提案してきた [1]。この手法は、名詞句の音声を再生する際に、次の名詞句の音声を部分的に重ねて再生するというものである。本稿では、本提案手法の概要と、その提案手法の有効性を評価するための実験計画について述べる。

2. 関連研究

視覚障害者と晴眼者との間に生じる情報格差を解消するために、様々な研究が行われている。

Machull ら [2] は、晴眼者と視覚障害者とのスキミング方法を比較し、非視覚的スキミングを支援する6つのデザインインプリケーションを提案した。以下にその6つのデザインをまとめる。

- 文書内の重要な部分（章、節、色文字や太字など）を一貫した方法で提示する
- ユーザが文書内のどこにいるか（章、節や段落）を把握できるようにする
- 必要なところに直接飛んだり戻ったりできる非線形での探索をサポートする

¹ 日本大学文理学部

² 日本大学大学院総合基礎科学研究科

a) miyata.akihiro@acm.org

- 色文字や太文字などの大事な部分を音声で読んだときにわかるようにする
- ユーザが文書を読む速度を任意に調整できるようにする
- それぞれのフィードバック方法（視覚・聴覚・触覚など）の得意・不得意を考慮する

Guerreiro ら [3] は、視覚障害者が複数の情報の中から特定の情報を効率的に選択できるようにすることを目的として、同時再生された複数の音声情報の中から特定の音声を聞き取るかを調査した。この研究で用いられた音声は、10～11 秒程度のニュースの短文である。結果として、視覚障害者が特定の音声を容易に把握できるのは、音源が 2～3 個の場合であることが明らかになった。

タッチスクリーンデバイス上における読書支援に関する研究では、特に効率的なスキミングや再訪問を可能にするためのシステム設計が進められている [4][5]。これらの研究では、視覚障害者が視覚的な手がかりを用いず、情報を素早く取得できるように、インターフェースやナビゲーションの最適化が図られている。具体的には、タッチスクリーンデバイス上において、指の動き（ピンチイン・アウト）を活用して要約の長さを調整する機能や、音声と触覚フィードバックを用いた非線形ナビゲーションを導入している。また、文書を章や段落などの単位で整理し、情報の階層を明確化することで、ユーザが特定の箇所へ移動できるようにする仕組みも取り入れられている。一方で、Khan[6] らは、対象ユーザを視覚障害者に限定せず、状況的障害（通勤中、歩行中、目の疲れなど）のある人を対象としたオーディオスキミングアプリケーションを実装している。彼らは「視線を減らした」スキミングの概念を導き出し、その実現のため設計ガイドラインを作成した。そして、このガイドラインを具体化する形で、オーディオスキミングアプリケーションを設計し、視界が制限される状況でも効率的に情報を取得できる仕組みを提案した。

3. 研究課題

日常生活において、レストランのメニューのような名詞句群が提示される場面が存在する。視覚障害者がこのような名詞句群を把握する手法として、スクリーンリーダーを用いた音声再生が多く用いられている。一般的なスクリーンリーダーでは、音声を 1 つずつ連続的に出力するようになっている（以降従来手法と呼ぶ）。しかし、視覚障害者がそれらの名詞句群を把握したい場合、従来手法では 1 つずつ連続的に音声出力される名詞句を聞き取る必要があり、名詞句群全体の把握に時間がかかる。そのため、視覚障害者は晴眼者と比べて、複数の情報を効率的に把握することが困難である。これに関することが Machull ら [2] の研究で触

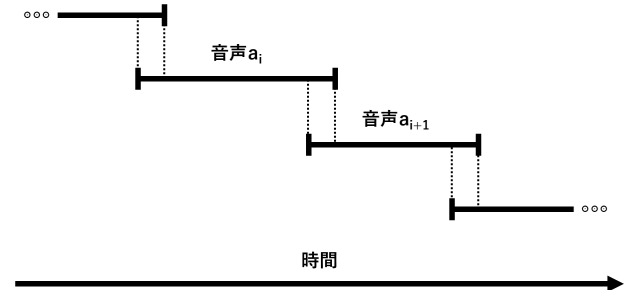


図 1 音声時間差重畳再生のイメージ図

れられており、視覚障害者が文書を聴覚的に参照する際に、視覚的な手がかりが失われてしまうため、効率的なスキミングが難しいと指摘されている。その結果、視覚障害者は効率の悪い探索戦略や、認知的負荷の高い方法に頼らざるを得ないことが示されている。この点に関して、Guerreiro ら [3] は、同時再生された音声群の中から特定の音声を聞き取るというタスクにおいては一定の成果を得られているが、音声群全体を効率的に把握することまでは達成されていない。また、タッチスクリーンデバイス上における読書支援に関する研究 [4][5][6] においても同様に、文書内の情報すべてを効率的に把握することは達成されていない。特に、視覚障害者が名詞句群全体を短時間で把握するという課題に対しては、まだ十分な解決策が提示されていない。

このような視覚障害者と晴眼者との情報格差を解消するために、視覚障害者が名詞句群全体を短時間で把握できるようにすることは重要である。よって、本稿では視覚障害者が名詞句群全体を短時間で把握できるようにすることを研究課題とする。

4. 提案手法

視覚障害者への情報取得支援を行ううえで、特殊なデバイスを用いることは視覚障害者にとって負担となる可能性があるため、より普遍的なデバイスやプラットフォームを採用することが望ましいとされている [7]。そのため、視覚障害者が日常的に携帯しているスマートフォンで出力可能な音声を用いて情報提示することが妥当であると考えられる。

これを踏まえて我々は、視覚障害者が音声を用いて名詞句群全体を短時間で把握できる手法として、音声時間差重畳再生手法を提案してきた [1]。この手法は、名詞句群に含まれる各名詞句の音声再生されている最中に、次の名詞句音声を部分的に重ねて再生するというものである。具体的には、音声群 $A = (a_1, a_2, \dots, a_n)$ が与えられたとき、音声群 A に含まれる各音声 a_i が再生終了する前の特定のタイミングで、次の音声 a_{i+1} の再生を開始し、2 つの音声を部分的に重畳させる。この操作を $i = 1, 2, \dots, n-1$ に対して繰り返し適用することによって、時間差重畳音声を

作成し、その音声を再生する。なお、図1では、音声 a_i と音声 a_{i+1} が時間差重畳している例を示している。

この提案手法により、視覚障害者が名詞句群全体を把握するのにかかる時間が、従来の音声再生手法よりも短縮されるという効果が期待される。

5. 実験計画

5.1 実験目的

本実験では、名詞句の音声同士を時間差で重畳させる秒数と、実験参加者が感じる時間差重畳音声の内容把握のしやすさとの関係を明らかにすることを目的とする。

5.2 評価指標（従属変数）

実験参加者は名詞句の音声同士を異なる秒数で時間差重畳させた複数の時間差重畳音声を聞き取り、その中に含まれていた個々の名詞句をどれだけ把握できたかを量的および定性的に測定する。定量評価は、実験参加者にリッカート尺度を用いたアンケートを回答させることによって測定し、そのアンケート結果を評価指標とする。また、定性評価は、実験参加者に音声聞き取りに関する自由記述を求めることによって測定し、実験参加者が感じた時間差重畳音声の内容把握のしやすさを評価指標とする。

5.3 実験条件

- 独立変数
 - ― 音声を重ねるタイミング：音声 a_i が再生終了する t 秒前に音声 a_{i+1} を再生する。 t は 0.0 から 0.5 まで 0.1 刻みで変化させる。
- 制御変数
 1. 時間差で重ねて再生する音声の数：10 個とする。
 2. 音声の種類：レストランのメニューとする。
 3. 各音声の再生時間：約 1.0 秒 ～ 約 2.0 秒とする。
 4. 各音声の話者：Apple 製品に搭載されている音声アシスタント Kyoko とする。
 5. 音声の再生速度：Kyoko の発話速度とする。
 6. 音の再生方式：モノラル再生とする。

5.4 実験手順

実験は下記の手順で行う。

Step 1 実験参加者は視覚障害者と同様の条件を再現するためにアイマスクを装着し、本提案手法の効果を厳密に測るためにヘッドホンを装着する。

Step 2 実験参加者は実験条件に沿った音声を 2 回聞く。

- 名詞句の音声同士を重畳させる秒数を 0.0 秒, 0.1 秒, 0.2 秒, 0.3 秒, 0.4 秒, 0.5 秒の計 6 種類として実験を行う。

- 実験参加者ごとに重畳秒数の順番を無作為に選ぶ。

Step 3 実験参加者はアイマスクとヘッドホンを外し、アンケートに回答する。

Step 4 Step 1 ～ Step 3 を 5 回繰り返す。

Step 5 実験参加者は事後アンケートに回答する。

Step 6 実験終了。

6. おわりに

本稿では、視覚障害者が名詞句群を短時間で把握できるようにするための、音声時間差重畳再生手法を提案した。具体的には、名詞句群の各要素を部分的に重畳させて再生することで、名詞句群全体の再生時間を短縮するという手法である。今後の研究では、視覚障害者が本提案手法を用いて名詞句群を短時間で把握できるか否かを明らかにする。

参考文献

- [1] 池田悠星, 土岐田力輝, 呉健朗, 宮田章裕: 名詞句群短時間把握のための音声重畳再生手法の基礎検討, ワークショップ 2024 (CN Workshop 2024) 論文集, Vol. 2024, pp. 81–82 (2024).
- [2] Machulla, T., Avila, M., Wozniak, P., Montag, D. and Schmidt, A.: Skim-reading strategies in sighted and visually-impaired individuals: a comparative study, *Proceedings of the 11th pervasive technologies related to assistive environments conference*, pp. 170–177 (2018).
- [3] Guerreiro, J.: Using simultaneous audio sources to speed-up blind people's web scanning, *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, pp. 1–2 (2013).
- [4] Ahmed, F., Soviak, A., Borodin, Y. and Ramakrishnan, I.: Non-visual skimming on touch-screen devices, *Proceedings of the 2013 international conference on Intelligent user interfaces*, pp. 435–444 (2013).
- [5] Li, Z., Jiang, Y., Liu, X., Zhao, Y., Yu, C. and Shi, Y.: Enhancing Revisitation in Touchscreen Reading for Visually Impaired People with Semantic Navigation Design, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 6, No. 3, pp. 1–22 (2022).
- [6] Khan, T. A., Yoon, D. and McGrenere, J.: Designing an eyes-reduced document skimming app for situational impairments, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14 (2020).
- [7] Gamage, B., Do, T.-T., Price, N. S. C., Lowery, A. and Marriott, K.: What do Blind and Low-Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study, *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 1–21 (2023).