

EchoAssist: サッカー侵入を用いた認知負荷推定による聞き逃し補完システムの提案

望月 崇広* 渡邊 恵太†

概要. 対面の授業や講演会などにおいて、登壇者の話を聞きながらメモを書いたり、関連する情報を探したりすることがある。この場合、人は話を聞きつつ手元の PC で作業をするというマルチタスクを行う。しかし、メモや情報検索に集中することで話を聞き逃すことがある。本研究では、眼球運動のひとつであるサッカー侵入の頻度から人の作業負荷を推定し、手元の作業に集中しているときに話された内容を要約して通知するシステム EchoAssist を提案する。このシステムにより、マルチタスク時の聞き逃しを補完できるようにし、手元の作業に集中しても聞き逃した部分だけを集中的に見返すことができる。

1 はじめに

授業や講演、ミーティングなど人の話を聞きながら重要と感じた部分や気付きなどをメモとして残すことは、多くの知的活動の基本である。特に、授業や学会では、発表中に参加者同士がオンラインでやり取りを行ったり、後から振り返るために感想やアイデアをテキストで残すことが推奨されている。ノートテイキングの重要性は理解を深め、内容の記憶を助ける点にある。しかし、講義のスピードが速すぎると内容についていけなくなることがあり、その結果として重要な情報を聞き逃すことがある [2][4]。また、人間は常に集中し続けることが難しく、注意力が途切れると聞き逃しが発生しやすくなる。さらに、自分の興味に意識が偏りすぎると、他の情報を無意識に排除してしまう選択的集中が生じることもある [1]。一度聞き逃すと、どの部分を聞き逃したかに気づけない問題も生じる [7]。その上、聞き逃した部分に追いつこうとすることで、現在の話の内容に置いていかれることにつながる。

これまでの研究では、「聞き逃し」を補助する様々な技術的手法が検討されてきた。坂東ら [8] は、オンライン授業において聞き逃しを防ぐため、静止画とテキストで授業内容を確認できるサポートツールを試作した。また、森脇らは、日常生活のマルチタスクによる音声の聞き逃しを補助するため、聞き逃した音を耳を塞ぐジェスチャーで聞き返すことができる聴覚拡張ヒアラブルデバイスの開発した [10]。この研究では重要な音を意識的に聞き分けるために「カクテルパーティー効果」 [1] に着目した。さらに、川村らは視聴時間を短縮するため、講義などの要約動

画を生成する「FastPerson」を提案し、見逃しを防ぎながら学習効率を向上させることに成功した [5]。これらの研究では、「聞き逃した情報の再取得」や「情報取得速度の向上」を主な目的としてきたが、「どこを聞き逃したか」を判別する機能については十分に検討されていない。

そこで本研究では、利用者が自身の「聞き逃し」を迅速に把握し、補完できるシステム「EchoAssist」を提案する。本システムは、講義中の発話内容について聞き逃した可能性の高い箇所を提示することで、ユーザーが聞き逃した内容をすばやく追いつけることを目指す。

2 EchoAssist

EchoAssist は話者の発話内容について音声認識を用いてテキスト化して保持し、ユーザーのメンタルワークロード (以下 MWL) の推定を組み合わせることで、どの時間帯の内容を聞き逃した可能性があるかを推定する。さらに、聞き逃し箇所を確認するために、新たに聞き逃しが発生することをできるだけ防ぐため、聞き逃した箇所の文章を要約してユーザーに通知する。これにより、ユーザーがメモを取るなど他の作業に集中してしまったとしても、システムがそれを検知し、ユーザーの作業が終わったタイミングでその時間帯の話者の会話内容を要約提示する。このため、ユーザーは安心してメモを取る作業ができるという利用体験の実現を目指す。

2.1 メンタルワークロードと聞き逃し推定

メンタルワークロード (以下 MWL) とは作業者にとってどの程度の作業負荷が望ましいかを検討するために用いられてきた概念である [11]。一般には NASA-TLX [9] などを用いて主観的評価を行う。サッカー侵入とは、眼球運動の一種である固視微動のうち、注視点から一瞬だけ逸脱し元の場所に戻

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 明治大学大学院先端数理科学研究科

† 明治大学総合数理学部

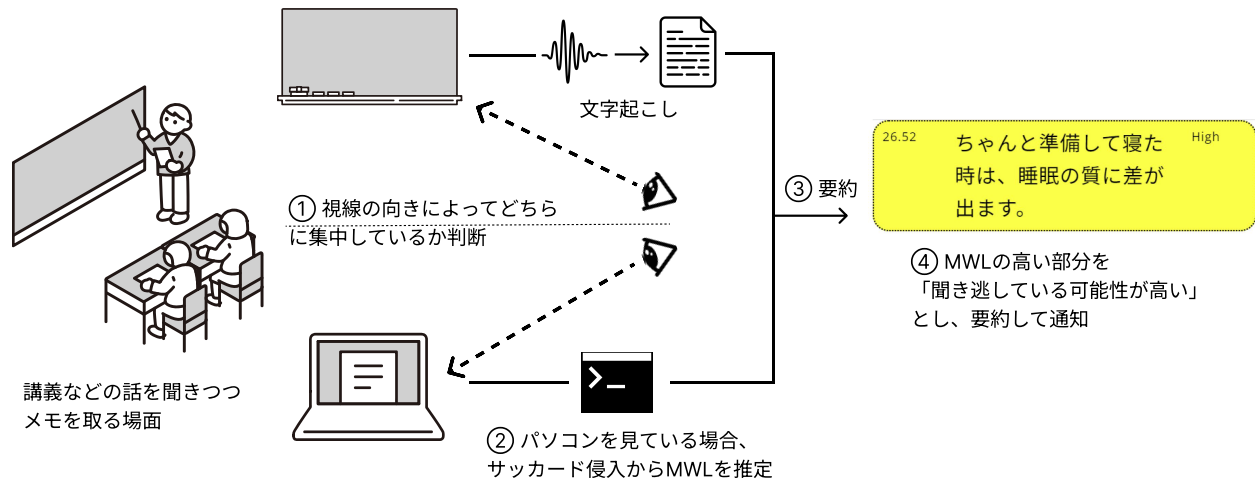


図 1. 提案システムの全体構成

る眼球運動のことである。MWLが高いと、このサッカード侵入の頻度が上がるとされている [6]。本システムでは、10秒おきにサッカード侵入の回数を計測し、その頻度から認知負荷レベルを推定する。手元を見ており、かつ認知負荷が高い場合は手元のタスクに集中していると判断し、聞き逃しをしている可能性があるとして推定する。

2.2 システム構成

ユーザーがノートなどを手元のPCで取っている間、EchoAssistはMWLを推定し、選択的集中が行われている可能性があればその時間の音声情報について要約して通知する。本システムは視線情報を処理してアプリケーションに送信する視線情報処理系と、話の内容を文字起こしし、タイムスタンプで管理した上でMWLの値を反映した処理を行う文字起こし処理系からなる。視線情報処理系は視線情報を取得するアイトラッカーと視線情報を処理するプログラムで構成した。手元のPC画面を見つめている間は文字入力などに取り組んでいる可能性があるとして、MWLの値を推定する(図1①, ②)。アイトラッカーはTobii社製Tobii Pro Fusionを用いた。文字起こし処理系は文字起こしデータについてタイムスタンプで区切って保管し、同時刻のMWLの値と照らし合わせる。MWLの値の高い時刻に書き起こされた内容を要約し(図1③)、通知する(図1④)。音声認識にはWeb Speech APIのSpeechRecognitionを用いた。ユーザーはMWL値が高く聞き逃したと推定された部分がタイムライン状に通知されるため、その部分を見るだけで情報に追いつくことができる。全体の構成について、図1に示す。

2.3 利用法と通知

ユーザーはパソコンにアイトラッカーを接続し、ソフトウェアを起動すると音声認識とMWLの推定が開始される。システムは視線データと音声データをリアルタイムで収集し、認知負荷を推定する。MWL値が上がり聞き逃しが推定されると、画面右上に聞き逃し対象の要約が通知表示される。なおこの通知をクリックすることで、音声認識された元の文字起こしや前後の文章を読むことが可能である。

3 議論

EchoAssistの動作や有用性を確認するために、聞き逃し箇所をハイライトするプロトタイプを利用し、8名の学生を対象に簡易的なユーザスタディを行った。EchoAssistを用いた介入群と通常の動画のみを視聴する対称群の双方について複数の評価指標を用いて調査を行った。参加者は一時的にMWLを上げるため、動画を見ながら断続的にN-backタスク[3]の2-backタスクを行った。この結果、N-backタスクの正答率は対称群が75%、システムを利用した介入群が91.67%だった。カイ二乗統計量は5.41、 $p=0.020$ であり、正答率は介入群について有意に高かった。主観評価(RSME)について、対称群の平均RSMEは84.0、介入群の平均RSMEは53.0、t統計量は14.46、 $P < 1.64 \times 10^{-38}$ であった。介入群は対称群に比べて有意に正答率が高く、また負担も少なく感じていた。EchoAssistを利用した介入群について、タスク中のMWL値が大きく下がる傾向が見られた。また、動画の内容に関する問題の正答率についてほとんど差が見られなかった。

参考文献

- [1] E. C. Cherry. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America*, 25(5):975–979, 1953.
- [2] H. Hembrooke and G. Gay. The laptop and the lecture: The effects of multitasking in learning environments. *Journal of computing in higher education*, 15:46–64, 2003.
- [3] iROCKBUNNY Lab. N-Back A JavaScript Implementation of the N-Back Task. <https://github.com/iRB-Lab/N-Back>.
- [4] R. Junco. In-class multitasking and academic performance. *Computers in Human Behavior*, 28(6):2236–2243, 2012.
- [5] K. Kawamura and J. Rekimoto. FastPerson: Enhancing Video-Based Learning through Video Summarization that Preserves Linguistic and Visual Contexts. In *Proceedings of the Augmented Humans International Conference 2024*, pp. 205–216, 2024.
- [6] S. Tokuda, G. Obinata, E. Palmer, and A. Charro. Estimation of mental workload using saccadic eye movements in a free-viewing task. In *2011 annual international conference of the IEEE engineering in medicine and biology society*, pp. 4523–4529. IEEE, 2011.
- [7] Q. Zhu and S. Ma. What Did I Miss? In *Adjunct Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 53–55, 2019.
- [8] 坂東宏和, 上西秀和, 山下真幸. 授業中に今までの内容を静止画とテキストで確認できるツールの提案と試作. 第 86 回全国大会講演論文集, 2024(1):399–400, 2024.
- [9] 三宅晋司, 神代雅晴. メンタルワークロードの主観的評価法 NASA-TLX と SWAT の紹介および簡便法の提案. *人間工学*, 29(6):399–408, 1993.
- [10] 森脇哲人, 長尾正太郎, 山本絵里香, 浦上ヤクリン, 金岡利知. 聴覚拡張ヒアラブルデバイス: 耳を塞いで聞き返すユーザインタフェースの開発. In *IEICE Conferences Archives*. The Institute of Electronics, Information and Communication Engineers, 2023.
- [11] 彦野賢, 篠原一光, 松井裕子. 繁忙感とメンタルワークロードとの関係に関する実験的検討. *人間工学*, 51(4):248–255, 2015.