

リアルタイム文字起こしの要約を多角的に生成・配信する講演支援システム

青山 柊太郎* 大伏 正泰†

概要. 講演は、同期性と一元性という特性から、情報伝達における効率性と網羅性を提供する一方で、個々の聴衆のニーズへの対応や理解度のばらつきへの対処が難しい。本論文では、この課題を解決するため、講演内容をリアルタイムで多様な形式に要約し、聴衆のデバイスに配信するシステムを提案する。本システムは、講演の音声を実時間でテキスト化し、大規模言語モデル (LLM) を用いて粒度、専門性、対象読者層など多様な観点から要約を生成する。聴衆は自身の理解度やニーズに応じて最適な要約を選択し、必要に応じてそれらを切り替えながら参照できる。これにより、リアルタイムな情報把握と個別ニーズへの対応を両立させ、WISS における講演体験の質的向上を目指す。特に、質疑応答時の理解補助や議論の深化、発表内容の記録・共有促進への貢献が期待される。

1 序論

講演は情報伝達において効率的かつ網羅的な手段であるが、その一元性と同期性から、個々の聴衆のニーズへの対応や理解度のばらつきへの対処が難しいという課題を抱えている。例えば、ある聴衆にとっては専門的すぎる内容であったり、別の聴衆にとっては既に知っている情報であったりするなど、講演内容の理解度は聴衆によって大きく異なる。また、リアルタイムで進行する講演においては、聴衆が自身のペースで情報を確認したり、重要なポイントを記録したりすることが困難である。

これらの課題を解決するため、本論文では講演内容をリアルタイムで多様な形式に要約し、聴衆のデバイスに配信するシステムを提案する。本システムは、講演の音声を実時間でテキスト化し、大規模言語モデル (LLM) を用いて、粒度、専門性、対象読者層など、多様な観点から要約を生成する。聴衆は自身の理解度やニーズに応じて最適な要約を選択し、必要に応じてそれらを切り替えながら参照できる。これにより、リアルタイムでの情報把握を可能にする同期性に加え、個々のニーズに対応した非同期的な情報アクセスを両立させ、講演体験の質的向上を目指す。

近年の音声認識技術と LLM の進歩は、リアルタイムでの高精度な音声テキスト化と高度なテキスト処理を可能にしている。本研究はこれらの技術を組み合わせることで、講演というリアルタイム性の高い情報伝達においても、個々の聴衆に最適化された情報提供を実現することを目指す。特に、WISS のような学会においては、発表内容の理解を深め、質

疑応答を活性化し、議論を深化させる上で、本システムの活用が大きな効果をもたらすと期待される。さらに、講演内容の記録や事後共有の促進にも貢献できる。

2 システム概要

提案システムは、講演会場において発表者の音声を入力として受け取り、リアルタイムで多様な要約を生成し、聴衆のデバイスに配信する。

WISS 会場での利用を想定し、Wi-Fi 環境を利用した配信システムとする。聴衆は各自のコンピューターやスマートフォンから専用の Web アプリケーションにアクセスすることで、リアルタイムに更新される要約を参照できる。



図 1. 文字起こしと要約を俯瞰するダッシュボード画面

ユーザインタフェースは、WISS 参加者が提示された要約を最大限活用できる形を目指す。具体的には、以下のような機能を提供する。

- 多様な要約の表示: 粒度、専門性、対象読者層の異なる複数の要約を並べて表示し、聴衆が自由に選択できるようにする。
- 要約のダウンロード: 生成された要約をテキストファイルとしてダウンロードできるようにする。

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.

* コロンビア大学

† 茨城工業高等専門学校

2.1 要約生成手法

本システムの核となるのは、LLM を用いた多様式要約生成である。LLM は、大量のテキストデータから学習した知識に基づき、高精度かつ自然な要約を生成することができる。本システムでは、WISS での講演内容や聴衆を踏まえて、以下の観点から要約を制御する。

- **粒度:** 要約の長さや詳細さを制御する。箇条書き形式の簡潔な要約から、詳細な説明を含む長い要約まで、多様な粒度の要約を生成する。(図 1)
- **専門性:** 要約の専門用語の使用頻度や説明の深さを制御する。初心者向けの平易な要約から、専門家向けの高度な要約まで、対象読者の専門知識レベルに合わせた要約を生成する。
- **対象読者層:** 要約の内容や表現を、特定の読者層に最適化する。例えば、研究者向けの要約、実務家向けの要約などを生成できる。

3 システム実装

本システムでは、LLM としては、GPT-4o API、音声認識には、Azure AI Speech API を用いる。リアルタイム処理を実現するため、音声認識結果を逐次 LLM に入力し、要約を生成するパイプラインを構築する。

要約の配信には、Websocket を用いることで、低遅延なリアルタイム配信を実現する。クライアントサイドは JavaScript で実装し、Web ブラウザ上で動作するアプリケーションとして提供する。

4 WISS2024 における活用と期待される効果

WISS2024 において本システムを実際に運用することで、発表者と聴講者双方にとって多くのメリットが期待される。

4.1 発表者にとってのメリット

- **質疑応答の質向上:** 要約によって聴衆の理解が深まることで、より質の高い質疑応答が期待できる。
- **聴衆の理解度把握:** 聴衆が読んだ要約の範囲や種類を分析することで、どの部分を重要視

しているか、どういった説明が求められているか把握することができる。

4.2 聴講者にとってのメリット

- **リアルタイムな理解促進:** 講演内容を多様な形式で要約することで、自身の理解度やニーズに合わせた情報取得が可能になる。
- **質疑応答への参加促進:** 要約を参照することで、より深く内容を理解し、質疑応答に積極的に参加できるようになる。
- **情報収集の効率化:** 講演内容を要約として記録することで、後から重要な情報を効率的に復習することができる。

4.3 アクセシビリティの向上

また、アクセシビリティの向上が期待できるため、より多くの聴衆を包摂することが可能となる。

- **多様な学習ニーズへの対応:** 視覚・聴覚障害を持つ参加者を含め、様々な学習スタイルや障害に対応した情報提供が可能になる。
- **言語バリアの軽減:** 多言語での要約提供により、非母語話者の理解を促進し、国際的なアクセシビリティを向上させることができる。
- **時間的制約の緩和:** 要約を後から参照することで、情報処理に時間を要する参加者も自身のペースで内容を理解することが可能になる。

WISS2024 会期中は講演会場に本システムを導入し、参加者全員が利用できるようにする。さらに、会期終了後にアンケートを実施し、システムの評価と今後の改善点を収集する。

5 結論と今後の展望

本論文では、講演内容をリアルタイムで多様な形式に要約し、聴衆のデバイスに配信するシステムを提案した。本システムは、LLM を用いることで高精度な要約生成を実現し、聴衆の理解度やニーズに合わせた情報提供を可能にする。WISS2024 における本システムの活用は、講演体験の質的向上に大きく貢献すると期待される。