

クラウド版字起こしシステムの開発

はじめに

NHK 放送技術研究所が開発した音声認識技術を元に、番組制作支援のための字起こしシステムを開発しました。字起こしシステムは、現在番組制作者において必須のツールとなっており、今後も様々なシーンでの活用が想定されます。そこで、字起こしシステムのさらなる価値の向上を目的として、2022年度より字起こしシステムをクラウド化する取り組みを開始しました。この記事では、クラウド化するメリットや、クラウド向けに開発・改修した機能について紹介します。クラウド化の開発には AWS (Amazon Web Service) を使用しました。

字起こし支援システムの概要

字起こし支援システムとは、字起こししたい素材の音声を音声認識技術により自動に文字化し、その文字の誤りを修正するだけで字起こしが作成できるため、従来よりも作業時間を大幅に短縮できます。

字起こしシステムには、オフライン字起こしとリアルタイム字起こしの2種類があります。オフライン字起こしは収録済み素材の字起こしに利用し、mp4などの動画ファイルをサーバーにアップロードして字起こしをします。リアルタイム字起こしはライブ素材の字起こしに利用し、リアルタイムに入力される音声を逐次字起こしします。どちらの機能も字起こしした結果を複数人で同時に確認・修正ができます。そのため長時間の映像素材でも協力して作業することが可能で、効率よく字起こしを制作することができます。また、字起こし結果を修正した箇所はオレンジ色で強調表示されるため、修正作業の重複を避けることができます。修正前と修正後のイメージを図1、図2に示します。



図1 字起こし結果テキスト修正前



図2 字起こし結果テキスト修正後

クラウド化のユーザーニーズとメリット

現在、システムに必要なサーバーなどの機器をユーザー環境に設置する形態でサービスの提供を行っていますが、利用ユーザーから以下のようなご意見・ご要望をいただくことがあります。

- ・サーバーなどの機器を設置するスペースが少ない
- ・運用の開始や終了を柔軟に行いたい
- ・メンテナンスをユーザー環境で行う場合があり、作業の調整に時間がかかる

これらはシステムをクラウド化することで、上記の要望の改善につながると考えられ、利便性の向上が期待できます。またクラウド化はユーザーだけではなく、システムを運用する側にも以下のようなメリットがあります。

- ・サーバー調達の手間が少なくなる
- ・サーバーのハードウェア障害への対応が不要
- ・リモートによるメンテナンスが可能

クラウド化への取り組み

クラウド化するにあたり、以下3つの機能を追加・改修しました。

1. IPエンコーダーによるライブ素材の伝送
クラウド環境にある字起こしシステムにライブ素材を伝送するため、リアルタイム字起こしの入力をIPエンコーダーに対応するように改修しました。ストリーミングプロトコルにはSRT (Secure Reliable Transport) という方式を採用し、高速かつ安定した伝送を実現しました。

2. 簡便かつ頑健なユーザー認証機能

クラウド環境におけるセキュリティを確保するため、個人や団体を特定する認証機能を採用しました。これは業務での利用を想定して特定の場所からのアクセスには認証を不要とし、外部からのアクセスには頑健な認証を設けるなど、簡便かつ頑健なセキュリティ環境を構築するためです。

また認証機能には、AWS の標準認証機構である Cognito (コグニート) を使用することで、字起こしシステム自体に新たな機能を加えることなくユーザー認証機能の追加を実現しています。

3. 修正インターフェースの改修

クラウド化するにあたり、字起こしシステムの修正インターフェースを改修しました。主な改修項目は以下の通りです。

- ・ 表示位置や再生の有無に応じて最適な画面表示を行う自動スクロール機能
- ・ クラウド環境での伝送遅延を考慮した再生データの先読み機構や、シームレス再生のための TS 再生機構の改善
- ・ 映像のない放送素材を考慮した、テキストのみの表示モード (図 3) を新設



図 3 テキストモード

また、以下の検証を行って、クラウド版字起こしの動作に最適な EC2 (Elastic Compute Cloud) サーバーを選定しました。

- ・ リアルタイム字起こし4系統とオフライン字起こしを同時に使用し、問題なく字起こしされるか
- ・ 1時間分のリアルタイム字起こし結果を複数人で修正して、修正結果が問題なく他の人の修正

画面に反映されるか

この場合のクラウド利用料はオンプレサーバーを3年間運用した場合とほぼ同等で、クラウド利用によるコストのデメリットは無いことを確認しています。

クラウド版字起こしシステムの利用イメージを図 4 に示します。

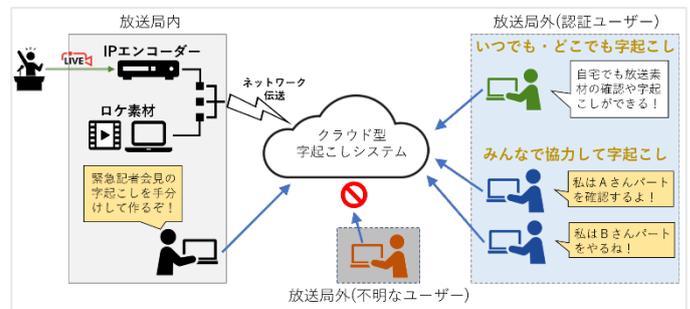


図 4 クラウド版字起こしシステムの利用イメージ

まとめ

2022 年度の取り組みは、クラウド環境においてもオンプレミス環境の字起こしシステムと同等の動作や操作性を実現できるかが大きな課題でした。リアルタイム字起こしにおいては、ライブ素材をクラウド環境に伝送するための IP エンコーダーの導入や、伝送遅延が発生しても確認用の音声をシームレスに再生可能にする方式の開発など、クラウド環境に特化した課題に多く対応する必要がありました。しかし様々な取り組みの結果、4つの独立したライブ素材のリアルタイム字起こしを有しつつ、オンプレミス環境と同等な試作システムを構築することができました。本成果を 2023 年の技研公開に展示し、様々な意見を収集する機会も設けています。

試作システムから実用システムに向けて、引き続き、取り組みを進めます。