**⾳声テキスト変換の概要**

[アーティクル] • 2024/09/24

Azure AI Speech サービスには、⾼度な⾳声テキスト変換機能が⽤意されています。 この機能では、リアルタイムおよびバッチの両⽅の⽂字起こしをサポートしており、オーディオ ストリームをテキストに変換するための汎⽤性の⾼いソリューションを実現できます。

# コア機能

⾳声テキスト変換サービスで提供する主要な機能は次のとおりです。

リアルタイムの⽂字起こし: ライブ オーディオ⼊⼒に対する、中間結果を伴うインスタント⽂字起こしです。

ファスト トランスクリプション: 予測可能な待機時間がある状況での最速の同期出⼒です。

バッチ⽂字起こし: ⼤量の録⾳済みオーディオに対して効率的な処理を⾏います。 Custom Speech: 特定のドメインおよび条件に対して精度を強化したモデルです。

# リアルタイムの⾳声テキスト変換

リアルタイムの⾳声テキスト変換では、マイクまたはファイルからオーディオを認識すると、その⽂字起こしを⾏います。 これは、次のような即時の⽂字起こしを必要とする⽤途に最適です。

**ライブ会議の⽂字起こし、キャプション、または字幕**: アクセシビリティと記録保持のためのリアルタイムのオーディオ⽂字起こしです。

**ダイアライゼーション**: オーディオ内のさまざまなスピーカーを識別して区別します。

**発⾳評価**: 発⾳の精度に関する評価を⾏いフィードバックを提供します。

**コール センター エージェントの⽀援**: カスタマー サービス担当者を⽀援するリアルタイムの⽂字起こしを提供します。

**ディクテーション**: ⽂書化を⽬的として、発話された単語を⽂字テキストに⽂字起こしします。

**⾳声エージェント**: 対話型⾳声応答システムでユーザーのクエリとコマンドを⽂字起こしできるようにします。

リアルタイムの⾳声テキスト変換には、Speech SDK、Speech CLI、REST API を介して

アクセスできます。このため、さまざまなアプリケーションおよびワークフローへの統合が可能です。 リアルタイムの⾳声テキスト変換は、Speech SDK、Speech CLI、REST

API (ファスト トランスクリプション API など) を介して利⽤できます。

# ファスト トランスクリプション (プレビュー)

ファスト トランスクリプション API を使⽤すると、オーディオ ファイルが⽂字起こしされ、その結果が同期して返されます。これは、リアルタイム オーディオよりも⾼速です。 ファスト トランスクリプションは、オーディオ録⾳の⽂字起こしを予測可能な待機時間でできるだけ早く必要とする次のようなシナリオで使⽤されます。

**オーディオまたはビデオのクイック⽂字起こしおよび字幕**: ビデオまたはオーディオ ファイル全体の⽂字起こしを、⼀度にすばやく取得します。

**ビデオ翻訳**: 様々な⾔語のオーディオが含まれる場合でも、ビデオの新しい字幕をすぐに取得します。

７

**注意**

ファスト

トランスクリプション

API

は、⾳声テキスト変換

REST API

バージョン

2024-05-15-

preview

以降のみで使⽤できます。

ファスト トランスクリプションを使い始めるには、ファスト トランスクリプション

API (プレビュー) の使⽤に関する記事をご覧ください。

# バッチ⽂字起こし API

バッチ⽂字起こしは、ファイルに格納されている⼤量のオーディオを⽂字起こしできるように設計されています。 この⼿法は、オーディオを⾮同期的に処理し、次の場合に適しています。

**事前に録⾳されたオーディオの⽂字起こし、キャプション、または字幕**: 保存されたオーディオ コンテンツをテキストに変換します。

**コンタクト センターの通話後の分析**: 録⾳された通話を分析して貴重な分析情報を抽出します。

**ダイアライゼーション**: 録⾳されたオーディオ内のスピーカーを区別します。

バッチ⽂字起こしは、次の⽅法で使⽤できます。

⾳声テキスト変換 REST API: 柔軟性の⾼い RESTful 呼び出しを利⽤してバッチ処理を容易にします。 使⽤を開始するには、「バッチ⽂字起こしの使⽤⽅法」および

「バッチ⽂字起こしのサンプル」を参照してください。

Speech CLI: リアルタイムおよびバッチ⽂字起こしの両⽅をサポートしているので、⽂字起こしのタスクの管理が簡単になります。 バッチ⽂字起こしに関する⾳声 CLI ヘルプを表⽰するには、次のコマンドを実⾏します。

|  |
| --- |
| Azure CLI |
| spx help batch transcription |

**Custom Speech**

Custom Speech を使⽤すると、アプリケーションや製品に対する⾳声認識の正確性を評価して改善できます。 カスタム⾳声モデルは、リアルタイムの⾳声テキスト変換、

⾳声翻訳、バッチ⽂字起こしに使⽤できます。

 **ヒント**

**バッチ⽂字起こし API** で Custom Speech を使⽤するには、**ホストされたデプロイエンドポイント**は必要ありません。 **Custom Speech モデル**がバッチ⽂字起こしにのみ使⽤される場合は、リソースを節約できます。 詳細については、「**Speech**

**Services の価格** 」を参照してください。

⾳声認識では、Microsoft が所有するデータを使⽤してトレーニングされ、⼀般的に使

⽤される⾳声⾔語を反映する基本モデルとしてユニバーサル⾔語モデルが活⽤されます。⾯倒な設定はありません。 基本モデルは、さまざまな⼀般的なドメインを表す⽅⾔と発⾳で事前トレーニングされています。 ⾳声認識要求を⾏うと、既定では、サポートされている各⾔語の最新の基本モデルが使⽤されます。 この基本モデルは、ほとんどの⾳声認識シナリオで適切に動作します。

Custom Speech を使⽤すると、アプリケーション固有のニーズに合わせて⾳声認識モデルを調整できます。 これは、次の場合に特に役⽴ちます。

**分野固有のボキャブラリの認識を向上させる**: 該当するフィールドに関連するテキスト データを使⽤してモデルをトレーニングします。

**特定のオーディオ条件について精度を向上させる**: オーディオ データと参照⽂字起こしを使⽤してモデルを調整します。

Custom Speech の詳細については、Custom Speech の概要に関するドキュメントおよび「⾳声テキスト変換 REST API」を参照してください。

⾔語およびロケールごとのカスタマイズ オプションの詳細については、「⾳声サービスの⾔語と⾳声のサポート」を参照してください。

# 使⽤例

Azure AI ⾳声テキスト変換を活⽤する⽅法の実⽤的な例を以下にいくつか紹介します。

| **ユース ケース** | **シナリオ** | **解決策** |
| --- | --- | --- |
| **ライブ会議の文字起こしとキャプション** | 仮想イベント プラットフォームでは、リアルタイムのキャプションをウェビナーに提供する必要があります。 | Speech SDK を使用してリアルタイムの音声テキスト変換を統合し、イベント中にライブで表示されるキャプションに音声コンテンツを文字起こしします。 |
| **カスタマー サービスの強化** | コール センターでは、顧客の通話の文字起こしをリアルタイムで提供することで、エージェントを支援したいと考えています。 | Speech CLI を介したリアルタイムの音声テキスト変換を使用して通話を文字起こしすることで、エージェントが顧客のクエリをより適切に理解し応答できるようにします。 |
| **ビデオ字幕** | ビデオ ホスティング プラットフォームでは、ビデオの一連の字幕をすばやく生成する必要があります。 | ファスト トランスクリプションを使用して、ビデオ全体の一連の字幕をすばやく取得します。 |
| **教育ツール** | E ラーニング プラットフォームでは、ビデオ講義の文字起こしを提供することを目的としています。 | 音声テキスト変換 REST API を介したバッチ文字起こしを適用して事前に録画された講義ビデオを処理し、学生向けのテキスト トランスクリプトを生成します。 |
| **医療文書** | 医療機関は、患者の相談内容を文書化する必要があります。 | ディクテーションにリアルタイムの音声テキスト変換を使用し、医療専門家が自分のメモを読み上げると、それが即座に文字起こしされるようにします。 カスタム モデルを使用して、特定の医療用語の認識を強化します。 |
| **メディアとエンターテイメント** | メディア企業は、ビデオの大規模なアーカイブについて字幕を作成したいと考えています。 | バッチ文字起こしを使用してビデオ ファイルを一括処理し、ビデオごとに正確な字幕を生成します。 |
| **市場リサーチ** | 市場リサーチ会社は、オーディオ録音から、顧客フィードバックを分析する必要があります。 | バッチ文字起こしを採用してオーディオ フィードバックをテキストに変換し、分析と分析情報の抽出が容易に行えるようにします。 |

# 責任ある AI

AI システムには、テクノロジだけでなく、それを使⽤する⼈、それによって影響を受ける⼈、それがデプロイされる環境も含まれます。 「透過性のためのメモ」を読み、システムでの責任ある AI の使⽤とデプロイについて確認してください。

透過性に関するメモとユースケース特性と制限統合と責任ある使⽤

データ、プライバシー、セキュリティ

**関連するコンテンツ**

⾳声テキスト変換の概要バッチ⽂字起こしを作成する価格の詳細については、⾳声サービスの価格に関するページを参照してください。