title: Custom Voice を作成する - Speech サービス titleSuffix: Azure Cognitive Services description: 実際のデータをアップロードする準備ができたら、Custom Voice ポータルに移動します。Custom Voice プロジェクトを作成するか、選択します。このプロジェクトでは、実際の音声トレーニングに使用するデータとして適切な言語またはロケールと性別プロパティを共有する必要があります。 services: cognitive-services author: erhopf manager: nitinme ms.service: cognitive-services ms.subservice: speech-service ms.topic: conceptual ms.date: 11/04/2019 ms.author: erhopf ms.openlocfilehash: 541448f08e4ce9961d34063dcc225bf89d969a73 ms.sourcegitcommit: f28ebb95ae9aaaff3f87d8388a09b41e0b3445b5 ms.translationtype: HT ms.contentlocale: ja-JP ms.lasthandoff: 03/30/2021 ms.locfileid: "101703373"

Custom Voice を作成する

Custom Voice 用のデータの準備に関するページでは、カスタム音声のトレーニングに使用できるさまざまなデータの種類と、さまざまな形式の要件について説明しました。実際のデータが準備できたら、Custom Voice ポータルに、または Custom Vision Training API を使用して、そのデータのアップロードを開始できます。ここでは、ポータルを使用したカスタム音声のトレーニング手順を説明します。

> [!NOTE] このページでは、「Custom Voice の概要」とカスタム音声用のデータの準備に関するページを読み、Custom Voice プロジェクトを作成していることを前提としています。

カスタマイズ用の言語に関するセクションで、カスタム音声用にサポートされている言語を確認してください。

データセットをアップロードする

実際のデータをアップロードする準備ができたら、Custom Voice ポータルに移動します。Custom Voice プロジェクトを作成するか、選択します。このプロジェクトでは、実際の音声トレーニングに使用するデータとして適切な言語またはロケールと性別プロパティを共有する必要があります。たとえば、英国アクセントの英語で音声を録音した場合は `en-GB` を選択します。

[データ] タブに移動し、**[データのアップロード]** をクリックします。ウィザードで、準備したものと一致する正しいデータの種類を選択します。

アップロードする各データセットでは、選択したデータの種類の要件が満たされている必要があります。アップロードする前に、データを正しく書式設定することが重要です。これにより、データが Custom Voice サービスによって確実に処理されます。Custom Voice 用のデータの準備に関するページに移動し、実際のデータが正しく書式設定されていることを確認します。

> [!NOTE] Free サブスクリプション (F0) ユーザーは、2 個のデータセットを同時にアップロードできます。Standard サブスクリプション (S0) ユーザーは、5 個のデータセットを同時にアップロードできます。制限に達した場合は、少なくとも 1 つのデータセットのインポートが終わるまで待機します。その後、やり直してください。

> [!NOTE] サブスクリプションあたりのインポートできるデータセットの最大数は、Free サブスクリプション (F0) ユーザーの場合は .zip ファイル 10 個、Standard サブスクリプション (S0) ユーザーの場合は 500 個です。

アップロード ボタンを押すと、データセットが自動的に検証されます。データ検証には、ファイル形式、サイズ、サンプリング レートを確認する、オーディオ ファイルの一連のチェックが含まれます。エラーが見つかった場合は、修正して、もう一度送信します。データのインポート要求が正常に開始されると、先ほどアップロードしたデータセットに対応するエントリがデータの表に表示されます。

次の表に、インポートされたデータセットの処理状態を示します。

State	意味
処理中	ご自分のデータセットは受信され、処理されています。
成功	ご自分のデータセットは検証が済み、音声モデルの作成に使用できるようになっています。
失敗	ファイルのエラー、データの問題、ネットワークの問題など、さまざまな理由により、処理中にご自分のデータセットが失敗しました。

検証が完了すると、ご自分の各データセットについて、一致した発話の合計数を **[Utterances](発話)** 列で確認できます。選択したデータの種類では長いオーディオのセグメント化が必要な場合、この列には、実際のトランスクリプトに基づいて、または音声文字起こしサービスを通じて、自動的にセグメント化された発話のみが反映されます。さらに、検証済みのデータセットをダウンロードして、正常にインポートされた発話とそのマッピング トランスクリプトの詳細な結果を確認できます。ヒント: 長いオーディオのセグメント化では、データ処理が完了するまでに 1 時間以上かかることがあります。

データ詳細ビューでは、各データセットの発音スコアとノイズ レベルをさらにチェックできます。発音スコアの範囲は 0 ～ 100 です。スコアが 70 未満の場合は、通常、音声のエラーまたはスクリプトの不一致を示しています。アクセントが強いと発音スコアが下がることがあり、生成されるデジタル音声に影響します。

高い信号雑音比 (SNR) は、オーディオのノイズが低いことを示します。一般に、専門スタジオでの録音によって、SNR が 50 以上に達するようにできます。SNR が 20 未満のオーディオでは、生成される音声に明らかなノイズが含まれる可能性があります。

発音スコアが低い場合や SNR が悪い場合は、発話を録音し直すことを検討してください。再録音できない場合は、それらの発話をデータセットから除

外してもかまいません。

> [!NOTE] カスタム ニューラル音声を使用している場合は、**[Voice Talent](ボイス タレント)** タブでボイス タレントを登録する必要があります。録音スクリプトを準備するときは、TTS 音声モデルを作成して合成音声を生成するために音声データを使用することについて、ボイス タレントの同意を得るため、以下の文を必ず含めてください。"I [state your first and last name] am aware that recordings of my voice will be used by [state the name of the company] to create and use a synthetic version of my voice." (私 [自分の姓名] は、私の音声の合成バージョンを作成して使用するために、私の音声が [会社名] によって使用されることを承知しています。) この文は、トレーニング データセット内の録音が、同意したのと同じ人物によって行われたかどうかを確認するために使用されます。データが処理される方法およびボイス タレントの確認が行われる方法の詳細については、こちらで確認してください。

ご自分のカスタム音声モデルを作成する

ご自分のデータセットの検証後、それを使用してご自分のカスタム音声モデルを作成できます。

1. **[テキスト読み上げ]** > **[Custom Voice]** > **プロジェクトの名前** > **[モデル]** に移動します。
2. **[Train model](モデルのトレーニング)** をクリックします。
3. 次に、このモデルを識別しやすい **名前** と **説明** を入力します。

   名前は慎重に選択します。ここで入力する名前が、SSML 入力の一部としての音声合成の要求時に、音声を指定するために使用する名前になります。アルファベット、数字、およびいくつかの区切り文字 (-、_、(、) など) だけを使用できます。音声モデルごとに、異なる名前を使用します。

   [Description](説明) フィールドの一般的な用途は、モデルの作成に使用されたデータセットの名前を記録することです。

4. **[Select training data](トレーニング データの選択)** ページから、トレーニングに使用する 1 つまたは複数のデータセットを選択します。送信前に、発話の数を確認します。"アダプティブ" トレーニング方法を使用する en-US と zh-CN の音声モデルについては、任意の数の発話から始めることができます。他のロケールでは、"統計的パラメトリック" と "連結" のトレーニング方法を含む標準レベルを使用して音声をトレーニングできるようにするには 2,000 より多くの発話を、またカスタム ニューラル音声をトレーニングするには 300 より多くの発話を、選択する必要があります。

   > [!NOTE] 重複したオーディオ名はトレーニングから削除されます。選択したデータセット内の複数の .zip ファイルに同じオーディオ名が含まれていないことを確認してください。
   >
   > [!TIP] 高品質の結果を得るためには、同じ話者のデータセットを使用する必要があります。異なるトレーニング方法には、異なるトレーニング データ サイズが必要です。"統計的パラメトリック" 方法を使用してモデルをトレーニングするには、少なくとも 2,000 個の異なる発話が必要です。"連結" 方法の場合は 6,000 個の発話ですが、"ニューラル" の場合の最小データ サイズ要件は 300 個の発話です。

5. 次のステップで、**トレーニング方法** を選択します。

   > [!NOTE] ニューラル音声をトレーニングする場合は、ボイス タレントのプロファイルと共に、自分の音声データがカスタム音声モデルのトレーニングに使用されることをボイス タレントが承認している音声同意ファイルを、指定する必要があります。カスタム ニューラル音声を利用するためのアクセスには制限があります。責任ある AI の要件について理解し、こちらのアクセスを適用してください。

   このページでは、テスト用のスクリプトのアップロードを選択することもできます。テスト スクリプトは、1 Mb 未満の txt ファイルである必要があります。サポートされているエンコード形式は、ANSI/ASCII、UTF-8、UTF-8-BOM、UTF-16-LE、または UTF-16-BE です。発話の段落ごとに、個別の音声になります。すべての文を 1 つの音声に結合したい場合は、1 つの段落にします。

6. **[Train](トレーニング)** をクリックして、実際の音声モデルの作成を開始します。

[トレーニング] の表に、この新しく作成されたモデルに対応する新しいエントリが表示されます。この表には、次の状態も表示されます。処理中、成功、失敗。

表示される状態は、ここに示すように、ご自分のデータセットから音声モデルへの変換プロセスを反映しています。

State	意味
処理中	実際の音声モデルを作成中です。
成功	実際の音声モデルは作成が済み、デプロイ可能です。
失敗	気が付かなかったデータの問題やネットワークの問題など、さまざまな理由により、トレーニング中に実際の音声モデルが失敗しました。

トレーニング時間は、処理されるオーディオ データの量と、選択したトレーニング方法によって異なります。30 分から 40 時間かかる可能性があります。実際のモデルのトレーニングが成功したら、そのテストを開始できます。

> [!NOTE] Free サブスクリプション (F0) ユーザーは、1 つの音声フォントを同時にトレーニングできます。Standard サブスクリプション (S0) ユー

ザーは、3 つの音声を同時にトレーニングできます。制限に達した場合は、少なくとも 1 つの音声フォントのトレーニングが終わるまで待ってから、やり直します。

[!NOTE] カスタム ニューラル音声のトレーニングは無料ではありません。価格を確認してください。

[!NOTE] サブスクリプションあたりのトレーニングできる音声モデルの最大数は、Free サブスクリプション (F0) ユーザーの場合はモデル 10 個、Standard サブスクリプション (S0) ユーザーの場合は 100 個です。

ニューラル音声トレーニング機能を使用している場合、リアルタイムのストリーミング シナリオ向けに最適化されたモデルをトレーニングするか、または非同期の長いオーディオ合成用に最適化された HD ニューラル モデルをトレーニングするかを選択できます。

実際の音声モデルをテストする

トレーニングごとに、モデルのテストに役立つ 100 個のサンプル オーディオ ファイルが自動的に生成されます。音声モデルが正常に作成されたら、展開して使用する前にテストすることができます。

1. **[テキスト読み上げ]** > **[Custom Voice]** > **プロジェクトの名前** > **[モデル]** に移動します。
2. テストするモデルの名前をクリックします。
3. モデルの詳細ページの **[テスト]** タブで、サンプルのオーディオ ファイルが見つかります。

音声の品質は、トレーニング データのサイズ、録音の品質、トランスクリプト ファイルの正確さ、トレーニング データに録音された音声が目的のユース ケースに合わせて設計された音声の性格とどの程度一致しているかなど、さまざまな要因に依存します。テクノロジの機能と制限、およびモデルの品質を向上させるためのベスト プラクティスの詳細については、こちらを確認してください。

カスタム音声エンドポイントを作成して使用する

音声モデルの作成とテストが正常に終了したら、カスタム Text-to-Speech エンドポイントに展開します。その後は、REST API で Text-to-Speech 要求を行うときの通常のエンドポイントの代わりに、このエンドポイントを使います。ご自分のカスタム エンドポイントは、フォントをデプロイするときに使ったサブスクリプションからのみ呼び出すことができます。

新しいカスタム音声エンドポイントを作成するには、**[テキスト読み上げ]** > **[Custom Voice]** > **[エンドポイント]** に移動します。**[エンドポイントの追加]** を選択し、ご自分のカスタム エンドポイントの **名前** と **説明** を入力します。次に、このエンドポイントに関連付けるカスタム音声モデルを選択します。

[追加] をクリックすると、エンドポイントの表にご自分の新しいエンドポイントのエントリが表示されます。新しいエンドポイントのインスタンス化には、数分かかることがあります。展開の状態が **[Succeeded](成功)** の場合、エンドポイントを使用する準備ができています。

常に使用するのでない場合は、エンドポイントを **中断** して **再開** することができます。中断後にエンドポイントが再アクティブ化されるとき、エンドポイントの URL は同じままになるので、アプリのコードを変更する必要はありません。

また、エンドポイントを新しいモデルに更新することもできます。モデルを変更するには、必ず新しいモデルの名前を更新するモデルと同じにします。

[!NOTE] Free サブスクリプション (F0) ユーザーは、1 つだけモデルをデプロイできます。Standard サブスクリプション (S0) ユーザーは、それぞれが独自のカスタム音声を使用する最大 50 個のエンドポイントを作成できます。

[!NOTE] 実際のカスタム音声を使用するには、音声モデルの名前を指定し、HTTP 要求に直接カスタム URI を使用し、同じサブスクリプションを使用して TTS サービスの認証を通過する必要があります。

ご自分のエンドポイントがデプロイされると、エンドポイント名はリンクとして表示されます。リンクをクリックすると、エンドポイント キー、エンドポイントの URL、サンプル コードなど、ご自分のエンドポイントに固有の情報が表示されます。

Custom Voice ポータルを使用して、エンドポイントのオンライン テストを行うこともできます。ご自分のエンドポイントをテストするには、**[Endpoint detail](エンドポイントの詳細)** ページから **[Check endpoint](エンドポイントの確認)** を選択します。エンドポイントのテスト ページが表示されます。読み上げるテキストをテキスト ボックスに (プレーンテキストまたは SSML 形式のどちらかで) 入力します。カスタム音声フォントで読み上げられるテキストを聞くには、**[Play](再生)** を選択します。このテスト機能は、カスタム音声合成の実際の使用量に対して課金されます。

カスタム エンドポイントの機能は、テキスト読み上げ要求に使用される標準のエンドポイントと同じです。詳しくは、REST API に関するページをご覧ください。

次のステップ

- ガイド:音声サンプルを録音する
- Text-to-Speech API リファレンス
- Long Audio API