



Azure에서 음성 시작

3분

Microsoft Azure는 다음과 같은 API(애플리케이션 프로그래밍 인터페이스)를 포함하는 **음성** 인지 서비스를 통해 음성 인식 및 음성 합성 기능을 모두 제공합니다.

- **Speech-to-Text API**
- **Text-to-Speech API**

음성 서비스용 Azure 리소스

애플리케이션에서 음성 서비스를 사용하려면 Azure 구독에 적절한 리소스를 프로비저닝해야 합니다. 다음 종류의 리소스 중 하나를 프로비저닝하도록 선택할 수 있습니다.

- **음성 리소스** - 음성 서비스를 사용하려는 경우 또는 다른 서비스와 별도로 리소스에 대한 액세스 및 청구를 관리하려는 경우 이 리소스 종류를 선택합니다.
- **Cognitive Services 리소스** - 다른 인지 서비스와 함께 음성 서비스를 사용할 계획이고 이러한 서비스에 대한 액세스 및 청구를 함께 관리하려는 경우 이 리소스 종류를 선택합니다.

Speech to Text API

Speech to Text API를 사용하여 오디오의 실시간 또는 일괄 전사를 텍스트 형식으로 수행할 수 있습니다. 전사용 오디오 소스는 마이크 또는 오디오 파일에서 나오는 실시간 오디오 스트림일 수 있습니다.

Speech to Text API에서 사용하는 모델은 Microsoft에서 학습한 범용 언어 모델을 기반으로 합니다. 모델에 사용하는 데이터는 Microsoft 소유이며 Microsoft Azure에 배포됩니다. 이 모델은 대화 및 받아쓰기라는 두 가지 시나리오에 최적화되어 있습니다. Microsoft에서 미리 빌드된 모델이 필요한 항목을 제공하지 않는 경우 음향, 언어 및 발음을 포함한 사용자 지정 모델을 만들고 학습할 수도 있습니다.

실시간 전사

실시간 Speech to Text를 이용하여 오디오 스트림에서 텍스트를 전사할 수 있습니다. 프레젠테이션, 데모 또는 사람이 말하는 다른 모든 시나리오에 실시간 전사를 사용할 수 있습니다.

실시간 전사가 작동하려면 마이크에서 들어오는 오디오 또는 오디오 파일과 같은 다른 오디오 입력 소스를 애플리케이션에서 청취할 수 있어야 합니다. 애플리케이션 코드는 오디오를 서비스로 스트리밍하여 전사된 텍스트를 반환합니다.

전사 일괄 처리

모든 Speech to Text 시나리오가 실시간인 것은 아닙니다. 파일 공유, 원격 서버 또는 Azure 스토리지에 오디오 녹음이 저장되어 있을 수 있습니다. SAS(공유 액세스 서명) URI가 있는 오디오 파일을 가리키고 비동기적으로 전사 결과를 받을 수 있습니다.

일괄 작업이 '최선의 노력 기준'으로 예약되므로 전사 일괄 처리는 비동기 방식으로 실행해야 합니다. 일반적으로 작업은 요청 후 몇 분 이내에 실행을 시작하지만 작업이 실행 중 상태로 변경되는 시점에 대한 추정은 없습니다.

Text to Speech API

Text to Speech API를 사용하면 텍스트 입력을 가청 음성으로 변환할 수 있으며, 이를 컴퓨터 스피커를 통해 직접 재생하거나 오디오 파일에 쓸 수 있습니다.

음성 합성 목소리

Text to Speech API를 사용하는 경우 텍스트를 발음하는 데 사용할 음성을 지정할 수 있습니다. 이 기능을 통해 유연하게 음성 합성 솔루션을 개인화하고 개성을 부여할 수 있습니다.

이 서비스에는 '표준' 음성뿐만 아니라 '신경망'을 활용하여 억양과 관련된 음성 합성의 일반적인 한계를 극복하는 '신경' 음성을 포함하여 여러 언어 및 지역 발음을 지원하는 여러 미리 정의된 음성이 포함되어 있어 보다 자연스러운 음성이 생성됩니다. 사용자 지정 음성을 개발하고 Text to Speech API와 함께 사용할 수도 있습니다.

지원되는 언어

음성 텍스트 변환 및 Text to Speech API는 다양한 언어를 지원합니다. 아래 링크를 사용하여 지원되는 언어에 대한 세부 정보를 찾습니다.

- [Speech to Text 언어](#).
- [Text to Speech 언어](#).

다음 단원: 연습 - 음성 서비스 사용

계속 >

