



소개

3분

인공 지능(AI) 솔루션이 구두 명령을 받아들이고 음성 응답을 제공하는 서비스가 더욱 늘어날 것으로 기대됩니다. "조명 끄기"와 같은 명령을 내리고 "오늘 비가 올 것인가?"와 같은 질문에 구두 답변을 요청하는 등 말로 통제할 수 있는 홈 시스템 및 자동차 시스템이 더욱 증가하게 될 것입니다.

이러한 종류의 상호 작용이 가능하려면 AI 시스템이 다음 두 가지 기능을 지원해야 합니다.

- **음성 인식** - 음성 입력을 감지하고 해석하는 기능.
- **음성 합성** - 음성 출력을 생성하는 기능.

음성 인식

음성 인식은 말해진 단어를 텍스트 표현으로 전사하는 방법을 통해 처리할 수 있는 데이터로 변환하는 것과 관련이 있습니다. 말해진 단어는 오디오 파일의 녹음된 음성 또는 마이크에서 나오는 라이브 오디오 형식이 될 수 있습니다. 음성 패턴은 오디오에서 분석되어 단어에 매핑되는 인식 가능한 패턴을 결정합니다. 이를 위해 소프트웨어는 일반적으로 다음을 비롯한 여러 유형의 모델을 사용합니다.

- 오디오 신호를 음소(특정 사운드를 나타내는 단위)로 변환하는 '음향' 모델.
- 음소를 단어로 매핑하는 '언어' 모델(일반적으로 음소에 따라 가장 가능성이 높은 단어 시퀀스를 예측하는 통계 알고리즘 사용).

인식된 단어는 일반적으로 다음과 같이 다양한 목적에 사용할 수 있는 텍스트로 변환됩니다.

- 녹화된 동영상 또는 라이브 비디오에 대한 자막 제공
- 전화 통화 또는 회의 내용 대본 만들기
- 자동화된 메모 받아쓰기
- 추가 처리를 위해 의도한 사용자 입력 결정

음성 합성

음성 합성은 대부분 음성 인식의 반대입니다. 음성 합성은 일반적으로 텍스트를 음성으로 변환하여 데이터를 음성화하는 것과 관련이 있습니다. 음성 합성 솔루션에는 일반적으로 다음 정보가 필요합니다.

- 읽을 텍스트.

- 말을 음성화하는 데 사용할 음성.

음성을 합성하기 위해 시스템은 일반적으로 텍스트를 '토큰화'하여 개별 단어로 분할하고 각 단어에 음성 발음을 할당합니다. 그런 다음 오디오 형식으로 변환될 음소를 만들기 위해 음성 전사를 '운율' 단위(예: 구, 절 또는 문장)로 세분화합니다. 그런 다음 음고 및 음색과 같은 매개 변수를 결정하는 음성을 적용하고 스피커로 출력되거나 파일에 기록될 수 있는 오디오 파형을 생성하여 이러한 음소를 오디오로 합성합니다.

다음과 같이 다양한 목적으로 음성 합성의 출력을 사용할 수 있습니다.

- 사용자 입력에 대한 음성 응답 생성.
- 전화 시스템용 음성 메뉴 만들기.
- 핸즈프리 시나리오에서 메일 또는 문자 메시지를 소리 내어 읽기.
- 기차역 또는 공항과 같은 공공장소에서 공지 사항 방송.

다음 단원: Azure에서 음성 시작

계속 >