**Các công cụ Speech-to-Text**

1. **Google Cloud Speech-to-Text**
   * Hỗ trợ nhiều ngôn ngữ và phương ngữ.
   * Cung cấp khả năng nhận diện giọng nói theo thời gian thực.
   * Có thể xử lý các âm thanh dài và có khả năng nhận dạng theo ngữ cảnh.
2. **Microsoft Azure Speech-to-Text**
   * Hỗ trợ nhiều ngôn ngữ và giọng nói, cung cấp nhận diện giọng nói theo thời gian thực.
   * Khả năng tùy chỉnh mô hình với Custom Speech để cải thiện độ chính xác cho các ứng dụng cụ thể.
   * Tích hợp dễ dàng với các dịch vụ khác của Azure.
3. **IBM Watson Speech to Text**
   * Cung cấp khả năng chuyển đổi âm thanh thành văn bản theo thời gian thực.
   * Hỗ trợ nhiều ngôn ngữ và cung cấp khả năng tùy chỉnh từ điển và ngữ pháp.
   * Có thể xử lý các tập tin âm thanh đã ghi âm sẵn hoặc streaming audio.
4. **Amazon Transcribe**
   * Dịch vụ STT của AWS hỗ trợ nhiều ngôn ngữ và phương ngữ.
   * Có khả năng nhận diện giọng nói trong các cuộc hội thoại đa chiều và đánh dấu người nói.
   * Hỗ trợ các ứng dụng từ y tế, doanh nghiệp đến truyền thông với khả năng tích hợp cao.
5. **DeepSpeech (Mozilla)**
   * Mô hình STT mã nguồn mở dựa trên deep learning.
   * Được thiết kế để dễ dàng tích hợp vào các ứng dụng và có thể huấn luyện với dữ liệu của riêng bạn.
   * Hiệu quả trên cả phần cứng thông thường lẫn các thiết bị di động.
6. **Kaldi**
   * Mô hình STT mã nguồn mở được sử dụng rộng rãi trong nghiên cứu và ứng dụng công nghiệp.
   * Hỗ trợ nhiều ngôn ngữ và có khả năng tùy chỉnh mạnh mẽ.
   * Phù hợp với các hệ thống lớn đòi hỏi sự linh hoạt và khả năng tùy chỉnh.
7. **Vosk**
   * Công cụ STT mã nguồn mở với khả năng nhận diện giọng nói ngoại tuyến.
   * Hỗ trợ nhiều ngôn ngữ và có thể chạy trên các thiết bị nhỏ như Raspberry Pi.
   * Được tối ưu hóa cho các ứng dụng yêu cầu hiệu suất cao và tiêu thụ tài nguyên thấp.
8. **AssemblyAI**
   * Dịch vụ API STT dựa trên AI, cung cấp khả năng nhận diện giọng nói với độ chính xác cao.
   * Hỗ trợ nhiều ngôn ngữ và các tính năng nâng cao như nhận diện từ khóa, phát hiện người nói, và đánh dấu thời gian.
9. **Rev.ai**
   * Dịch vụ STT từ Rev, hỗ trợ nhiều ngôn ngữ và cung cấp khả năng chuyển đổi âm thanh thành văn bản với độ chính xác cao.
   * Tích hợp dễ dàng qua API và hỗ trợ nhiều định dạng âm thanh.

**Các mô hình Speech-to-Text**

1. **DeepSpeech (Mozilla)**
   * Mô hình STT dựa trên deep learning sử dụng RNN, được tối ưu hóa cho tốc độ và độ chính xác.
   * Phù hợp với cả các ứng dụng trực tuyến và ngoại tuyến.
2. **Wav2Vec 2.0 (Facebook AI)**
   * Mô hình STT sử dụng Transformer và learning representations trực tiếp từ âm thanh mà không cần transcript.
   * Mang lại độ chính xác cao trong nhận diện giọng nói, đặc biệt trong các ngữ cảnh thiếu dữ liệu.
3. **Jasper (NVIDIA)**
   * Mô hình STT dựa trên Convolutional Neural Network (CNN) được thiết kế cho độ chính xác cao và khả năng huấn luyện nhanh chóng.
   * Thường được sử dụng trong các ứng dụng yêu cầu thời gian thực và hiệu suất cao.
4. **Conformer (Google)**
   * Kết hợp giữa Convolutional Neural Networks (CNN) và Transformer, mô hình này cải thiện độ chính xác và tốc độ xử lý trong STT.
   * Được sử dụng rộng rãi trong các ứng dụng yêu cầu độ chính xác cao như trợ lý ảo và dịch vụ khách hàng.
5. **Listen, Attend and Spell (LAS)**
   * Mô hình STT dựa trên sequence-to-sequence với cơ chế attention, tạo ra văn bản từ chuỗi âm thanh.
   * Được sử dụng rộng rãi trong nghiên cứu và phát triển STT.
6. **QuartzNet (NVIDIA)**
   * Một phiên bản nhẹ và nhanh của Jasper, được tối ưu hóa cho các thiết bị có tài nguyên hạn chế.
   * Hiệu quả trong các ứng dụng thực tế, đặc biệt là trên thiết bị di động.