# Mô tả Data từ Doanh nghiệp

|  |  |  |  |
| --- | --- | --- | --- |
| STT | File nhận | Ký hiệu | Mô tả |
| 1 | out-90839053905-6918-20240724-154737-1721810857.3210042.wav | 1.wav | Đoạn ghi âm tín hiệu kém, chập chờn, không rõ chữ, đứt quăng, có chỗ giọng nhỏ. |
| 2 | out-90918738554-6918-20240725-092444-1721874284.3219571.wav | 2.wav | Đoạn ghi âm rè ít, ít chập chờn nhưng đoạn đối thoại không bị mất thông tin, khách hàng có vẻ hài lòng. |
| 3 | out-90942926006-6918-20240725-095203-1721875923.3221099.wav | 3.wav | Đoạn ghi âm rè nhiều, chập chờn, khách hàng có vẻ không quan tâm lắm. |
| 4 | out-99990944829575-6948-20240725-171443-1721902480.3243254.wav | 4.wav | Đoạn ghi âm chập chờn, rè, bên khách hàng phản hồi không nghe. |
| 5 | out-99990947676025-6948-20240725-173122-1721903480.3243974.wav | 5.wav | Đoạn ghi âm rè, một chút chập chờn, không bị mất thông tin. |

# Thông tin các công cụ/mô hình hiện tại

**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. **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.
6. **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.
7. **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.
8. **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.
   * 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.
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.

# Khảo sát thực tế

|  |  |  |
| --- | --- | --- |
| STT | Công cụ/ Mô hình | Kết quả khảo sát |
| 1 | Google Cloud Speech-to-Text | Hỗ trợ tiếng việt |
| 2 | Microsoft Azure Speech-to-Text | Hỗ trợ tiếng việt |
| 3 | IBM Watson Speech to Text |  |
| 4 | Amazon Transcribe | Không hỗ trợ tiếng việt |
| 5 | DeepSpeech (Mozilla) |  |
| 6 | Kaldi |  |
| 7 | Vosk |  |
| 8 | AssemblyAI |  |
| 9 | Rev.ai |  |
| 10 | Wav2Vec 2.0 (Facebook AI) |  |
| 11 | Jasper (NVIDIA) |  |
| 12 | Conformer (Google) |  |
| 13 | Listen, Attend and Spell (LAS) |  |
| 14 | QuartzNet (NVIDIA) |  |

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Assembly AI

A screenshot of a computer

Description automatically generated

**/**

**wav2vec2-base-vietnamese-160h**

****

|  |  |
| --- | --- |
| **File** |  |
| **1.wav** | **dạ sao đó xin nghe dạ em choem gọi tr tên mãng vin nó ú á chị á tim này củ mình đang sử dụng là hô xim bao g chị dạ ám to thì đổi xim đúng gờ giùm em nhá tới là bên vi na th sẽ hạ chạm ntim mình cchị nói nó cứ bị đức quản á em sao chị ủa nghe nghe được chưa chị nghe chứa ạ à diờ dạ được à dạ tại cái tim của mình đ đang sử dụng là cái phô tim bal củ ì phái ở tó nhằm so sẽ trạ thống bao gọ thô xim không bử dụng được nữa ochị em chủ chị vám in a chị đổi đên xim bốn gờ giùm em nha tháng bảy nà b gùm em mình ở thuật nà đạ em emem em hướng dẫn cái điểm giao dịch gần nhất chư chị dạ ở quận hai ạ phưần an phú quận hai vậy mình g cái cửa hàng một cố bảy trần não nha đó làm ở trong hảm mộn năm một dạ ủa nhưng mà chị cho em hỏi xíu cái xim này đứng tên năm lăm đúng khôn gống r tại cái này thì ton cho quán á nên là chị gọivới nhận cộc gọi cứ không có lạm cái gì trên cái điện thoại đãhếtvậ cái này đi đổi xim mà am lm em theo máy đổi g chị cứ chứ khổng cót ùm ậ n à dạ là ở tháng bảy l cắp là trong bảy li tốm giùm em chứ nó đang hạ chạm từ từ đó rợ lúc đó mình khôngcó sóng mình không gọi được luôn á à dạ dạ em cám ơn chị ừnhưng mà biết chmình hỏi cái điện thoại mình đang à là điện hai gờ luôn đúng kha dạ đúng rồi vậy thì mình phải đổi n cái điện thoại ba gờ luôn nha tại vì á điện thoại hai gờ nó nó cũng không in ác luôn á dạ dạ đạ e nói lại à dồ rồi vậ cám ơn chị nha ạ** |
| **2.wav** | **transcript: a lô a lô dạ em chào chị em gọi từ bên mộ khoảng chín hác àn của vin na pôn á chị ơi ừ mchị đang sử dụng cái xim vin na bên em và nhận xim bao giờ cũ á chị ờ là bên nhà mạng co th hạ chặm thóng bao giờ thì xim mình ó chị không có không nh tranh thủ bình ghét cửa hàng vin nam mình đổi xim giúp em nha ch đổi xim khá hả em dạ vẫn là cái số đó thôi chị chị đổi thức phô xim thôi để nó nhanh hơn với lại nó không bị kháa xim lại thôi chị à rồinhưng mà chị nhưng mà chị cho em hỏi xíu im này của anh bình đứng tên đúng không ạ đúng rồi chị lên chị đổi tn chị luôn à vậy chị chị đói anh bình lên vừa đổi tên gùm đổi đổi xim luôn giùm em nha ừ tại a đên ạ lên tranh thủ trong tháng bảy đến nă bôn đó có đó chị ở khu vực nào ậ để em hướng dẫn cógiấy cửa hàng ồ nhá cách lái cách lái chị chị gần trà chị chị nghe không ạ chị nghe em nói không rồi chị nghe rồi dạ chị mình có gần bên đường trần não không ó chị a ạ cụn hơi sao trần bão phía rưới này không có hải quận hai đó có hai điểm một là bên trần não nớ một bên nguyễn ni trin chị gần bên nào nhất ạ bị c rên bện cới trên phấy ạ dạ bốn chín hai chị ha ờ bốn chín thai vậy cho trên rồi thủ trong tuần giùm em nhưng mà thứ bảy này bên cửa hàng tr em không làm nha chị ừ chịtrong tháng bảy nó là được nha chị cảm ơn à ủ ôn bôn thai dạ đúng rồi đối diện con cưng đó chị ờ rổi cậy ô kê dạ ạ em cá án ơn** |
| **3.wav** | **transcript: a lô lô a lô d anh nghe a lô dạ anh nghe không ạ a l a lô dạ anh ơi em em gọi từ bên bộ phận chăm sóc khách hàng của vin na thô ó anh nghe không anh da lô a lô dạ dạ dạ dạ anh ơi mình đang sử dụng cái xim vin na bên em mà dạn phôi xim ba gờ cũ á anh sắp tới mình nhà mạng vin na à t t hạ chậm thóng bào là ì bết mình trong sử dụng được anh h anh tranh thủ anh ná cửa hàng vin na anh đổi xim bốn gờ giùm em nha a lô anh nghe không ạ ồ dạ a lô dạ dạ ạ dạ dạ thì mình tranh thủ mình ghé cửa hàng vin na mình đổi đang phô xim bốn gờ giùm em a ải sắp tới nhà mng đó tá hạ chặm sóng c xim mình skhông sử dụng đrồi rồi dạ anh ơi mình ở khu vực nào để em hướng dẫn cửa hàng gần vi na gần giúpo anh bì e ninh bình hạ nrồi dạ dạ anh gém cái ca hàng vin a đin bình anh đổi thì trong phán bạn em nha thông qua tháng tám sợi là chặm són hạ mình không sử dụng được dạ cám ơn anh ạ** |
| **4.wav** | **transcript: ôa lô chào chị chị ơiem gọi đến từ bộ phận chăm sóc khách hàng của vi na phôn á chị nhờ chị cầm căn cước công dân với lại cái điện thoại của mìnhe cô chị em nói không à a lô chị ơi em đến từ vi na phôn á chị chị à bêm** |
| **5.wav** | **transcript: a lô anh ôi vậy chào anh em vọi đến từ bộ phần chăm sóc khách hàng của vi na phôn á anh giờ anh cầm a cớ công dân với điện thoại của mình ra điểm giao dịch gần nhất để mình đổi từ phôi xim hai gờ lên bốn nười nhá anh dạ mình thực hiện trước ba mốt tháng bảy giùm em cảm ơn anh cầm bôn bảy ạ dạ trước ba mốt tháng bảy giùm em tchủ nhật là hông dạ không ch được không làm sáng thứ bảy bên em có làm cửa hàng ba đđoàn kết tờn bình họ tà có thủ đức ừ tám giờ cho tới khoảng mười một rưỡi hơn anh dạ cảm ơn anh** |

**PhoWhisper: Automatic Speech Recognition for Vietnamese**