AIによるトルコ語のテキスト読み上げと自由会話の音声認識
トルコ語母語話者と日本語母語話者の場合

第2回研究会　学習者コーパス分析Ⅱ：音声認識から談話標識まで

2021年12月19日

東京外国語大学世界言語社会教育センター　梅野　毅

発表概要

- AIによる音声認識エンジンについて
- トルコ語のテキスト読み上げと自由会話の音声認識の結果報告
- まとめ

AIによる音声認識技術について

- 5年ほど前にAIの音声認識能力が人間の能力を超えたという報道が多く見られた。
  - 2016年10月 Microsoftの音声認識技術が人間の能力を超える
    WER 5.9
    https://pc.watch.impress.co.jp/docs/news/1025715.html
    https://www.cbsnews.com/news/microsoft-speech-recognition-technology-understands-conversation-as-well-as-people-do/
  - 2017年、Googleの音声認識、WERが4.9に改善。
    https://robotstart.info/2017/05/22/google-speech-recognition-word-error-rate.html
- スマートスピーカーなど音声認識機能を組み込んだ家電が普及。
- 音声認識はアプリからCloud AIが主流に。

Cloud上の音声認識エンジンについて

- GCP（Google Cloud Platform）のSpeech-to-Text API
- AWS（Amazon Web Services）のAmazon Transcribe
- Microsoft AzureのSpeech Service

などが有名です（無料ですぐに使えます）。

GCP（Google Cloud Platform）のSpeech-to-Text API

- サポート言語：127言語
  https://cloud.google.com/blog/ja/products/ai-machine-learning/new-features-models-and-languages-for-speech-to-text
- 拡張モデル（テレフォニーモデルなど）を含めると289言語モデル。
  https://cloud.google.com/speech-to-text/docs/languages
- 料金は、月60分までは無料。以降、$0.006/15 秒（2.7円/1分）。
- アカウント作成不要の無料のデモもあり。（1分程度まで）。
  https://cloud.google.com/speech-to-text?hl=ja
- 使い始めるには、アカウントの作成＋クレジットカートの登録が必要。
- 音声は、マイク録音（ストリーミング）かファイルアップロードで処理。

AWS（Amazon Web Services）のAmazon Transcribe

- 対応言語：37言語
  https://docs.aws.amazon.com/transcribe/latest/dg/supported-languages.html#table-language-matrix
- 無料利用枠12 か月間、1 か月あたり 60 分（2.7円/分）。
  https://aws.amazon.com/jp/transcribe/pricing/
- 無料期間は1年だが有用枠は使用量が多いと割引がある。
- アカウントの作成が必要＋クレジットカートの登録が必要。
- 音声は、マイク録音（ストリーミング）かファイルアップロードで処理。

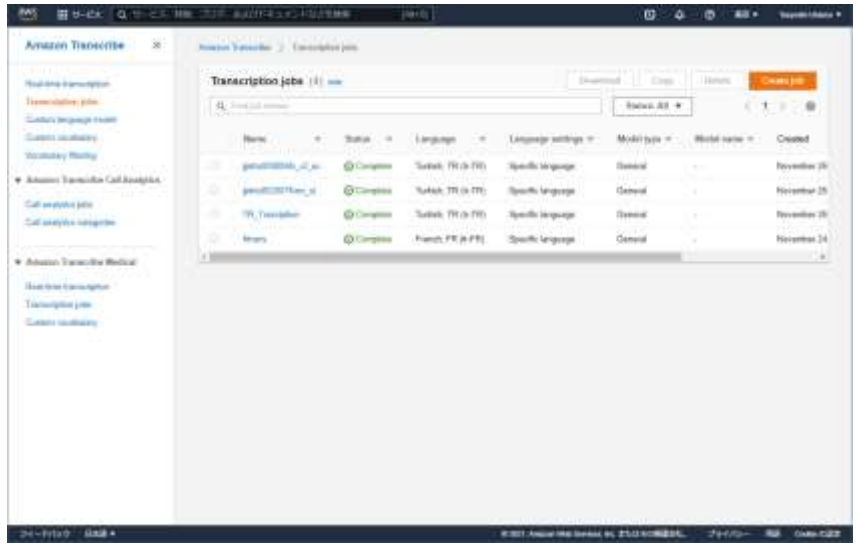
Microsoft AzureのSpeech Service

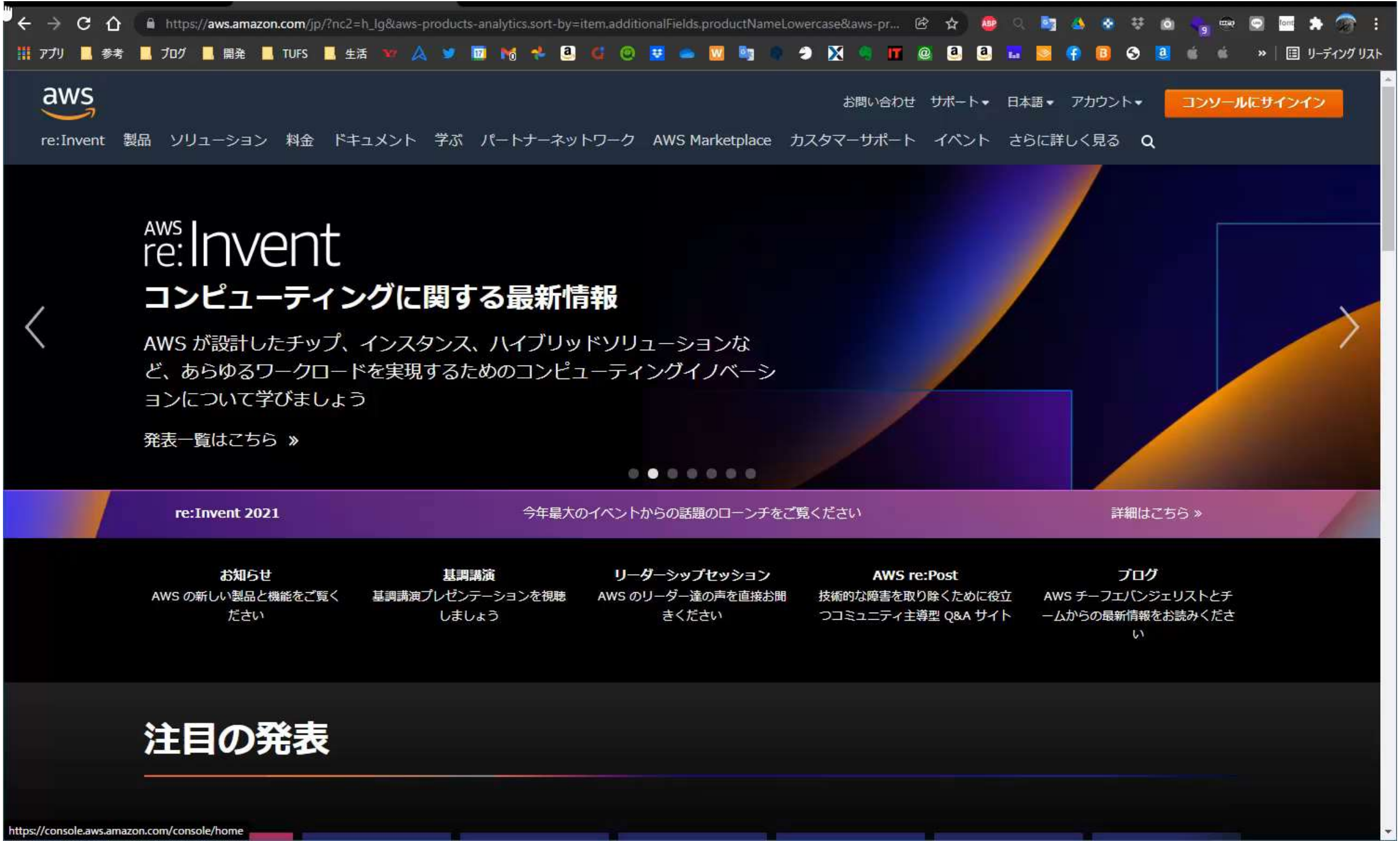
- サポート言語：125言語
  https://docs.microsoft.com/ja-jp/azure/cognitive-services/speech-service/language-support
- 料金
  https://azure.microsoft.com/ja-jp/pricing/details/cognitive-services/speech-services/

GCP Speech-to-Text APIの使い方

AWSの使い方

0. アカウント作成 https://aws.amazon.com/jp/register-flow/
1. ディスクを作り、音声をアップロード（S3）
2. Amazon Transcriptionサービスを起動しファイルを選ぶ。
3. 文字化実行（Create Job ）




https://aws.amazon.com/jp/?nc2=h_lg&aws-products-analytics.sort-by=item.additionalFields.productNameLowercase&aws-pr...
アプリ
参考
ブログ
開発
TUFS
生活
リーディング リスト
aws
お問い合わせ
サポート
日本語
アカウント
コンソールにサインイン
re:Invent
製品
ソリューション
料金
ドキュメント
学ぶ
パートナーネットワーク
AWS Marketplace
カスタマーサポート
イベント
さらに詳しく見る
AWS re:Invent
コンピューティングに関する最新情報
AWS が設計したチップ、インスタンス、ハイブリッドソリューションなど、あらゆるワークロードを実現するためのコンピューティングイノベーションについて学びましょう
発表一覧はこちら »
re:Invent 2021
今年最大のイベントからの話題のローンチをご覧ください
詳細はこちら »
お知らせ
AWS の新しい製品と機能をご覧ください
基調講演
基調講演プレゼンテーションを視聴しましょう
リーダーシップセッション
AWS のリーダー達の声を直接お聞きください
AWS re:Post
技術的な障害を取り除くために役立つコミュニティ主導型 Q&A サイト
ブログ
AWS チーフエバンジェリストとチームからの最新情報をお読みください
注目の発表
https://console.aws.amazon.com/console/home

トルコ語のテキスト読み上げと自由会話の音声認識の結果報告

- 読み上げタスクの単語数が181語であったため、比較のため自由会話についても文字起こし後の１８１語までで分析しました。
- 単語誤り率の計算には本ロジェクトのWebページを使用。
  http://www.coelang.tufs.ac.jp/interlang/
  ユーザ名：xxadmin
  パスワード：lang#inter%235

読み上げタスクの単語認識率

読み上げ原稿とAIにより文字化されたテキストの比較

- トルコ人による読み上げタスクのAI認識率
  レーベンシュタイン距離 = **30**
  単語誤り率 : **17%** = 30 ／ 181
  単語認識率 : **83%** = 1 − 0.17
- 日本人による読み上げタスクのAI認識率
  レーベンシュタイン距離 = **76**
  単語誤り率 : **42%** = 76 ／ 181
  単語認識率 : **58%** = 1 − 0.42

トルコ人による読み上げタスクの原稿との比較で誤りとされた部分を人力検証

読み上げタスク原稿	トルコ人読み上げ音声AI文字化	赤の数値は間違いでないのにWERに加算された数値＝9
kınalıada'da bir ev kınalıada'ya	kınalıada'da bir ev kınalı ada[y1] 'ya	[y1]間違いと言えない 2
ömrümde inmedim ama orayı öyle	ömrümde inmedim ama orayı öyle	
severim ki neden bilmem belki de	severim ki neden bilmem belki de	
orada kendisiyle hiç konuşmadığım	orada kendisi ile[y2] hiç konuşmadığım	[y2]正解（正書法の問題）2
bir arkadaşım otururdu da onun	bir arkadaşım otururdu da onun	[y3]間違い
için kınalı'nın önünden geçerken	için kınalının[y3] önünden geçerken	[y5]間違い
hep o arkadaşımın hiç gitmediğim	hep o arkadaşımın hiç gitmediğim	[y6]間違い
hiç gitmeyeceğim evini düşünürüm arkadaşım sakin sessiz iyi bir	hiç gitmeyeceğim evini düşünürüm arkadaşım sakin sessiz iyi bir	[y7]間違い
kızdır sabahleyin ilk vapurla işine iner	kızdır sabah [y4] ilk vapurla eşine[y5] iner	[y4]間違いと言えない 2
son vapura elinde paketlerle döner	sonra[y6] profilinde[y7] paketlerle[y8] döner	[y8]間違いと言えない 1
bazı defa son vapurda	bazı defa son vapurda	
beraberizdir onun iskelenin	beraberiz dir[y9] onun iskelenin	[y9]間違いと言えない 2
kalabalığında kaybolduğunu	kalabalığında kaybolduğunu	[y10]1語聞き取れていない
sonra projektörcü ışığını yolcularla	sonra projektör [y10] yolcularla	[y11]間違い
karanlık yola tuttuğu zaman az bir	karanlık yolla [y11] tuttuğu zaman az bir	[y12]間違い
şey sallanarak hızlı hızlı sağa saptığını	şey sağlanarak[y12] hızlı hızlı sağ[y13] saptığını	[y13]間違い
görür yine kaybederim küçük	görür yine kaybederim küçük	[y14]1語聞き取れていない
kaplamaları simsiyah kesilmiş bir ahşap	kaplamaları simsiyah kesilmiş bir	[y15]間違い
evde oturduğunu sanıyorum evden	ahşap evde oturduğunu [y14] evden	[y16]1語聞き取れていない
deniz görünmüyor olmalı yahut	deniz görünmüyor olmalı yahut	[y17]2語聞き取れていない
belki de bir iki penceresinden çakal	belki de bir iki penceresinden çakal	[y18]間違い
ayva nar hünnap ağaçlarını görürüm	ayva nar hünnap ağaçlarını görürüm	[y19]間違い
bahçede bir de çıkrıklı kuyu olacak	bahçede bir de çıkrıklı kuyu olacak	[y20]間違い
kırkını aşmış şişmanca yeşil gözlü	kırkını açmış[y15] şişmanca yeşil gözlü	[y21]1語聞き取れていない
bir kadın olan anasını kırmızı elma	bir kadın olan anasını kırmızı elma	
yüzüyle küf yeşili gözleriyle görür	yüzüyle küf yeşili gözleriyle görür	
ben de severim böyle bahçeyi	ben de severim böyle bahçeyi	
evini anasını tarif ederken gördüm	evini [y16] tarif ederken gördüm	
sanmayın ben görmeden	sanmayın ben görmeden	
severim bahçeleri insanları evleri	severim bahçeleri insanları evleri	
eve bu küçük bahçeden girilir evin alt	eve bu küçük bahçeden girilir evin alt	
katında kendileri oturur üst katını	katında kendileri oturur [y17]	
yazın kiraya verirler bir meryem	yazın kiraya verirler bir meryem	
ana kandili önündeki isa	ana kandili önündeki isa	
resminden küçücük sarımtrak	resminden kız çocuk[y18] sarımtırak	
aynaya kadar her şeyde ağır	ay ne[y19] kadar her şeyde alır[y20]	
günlük kokusuna benzer	günlük kokusuna benzer	
bir ortodoks hava eser	bir ortodoksu [y21] eser	

読み上げタスクの単語認識率（補正後）

- トルコ人による読み上げタスクのAI認識率
  レーベンシュタイン距離 = 21（ 30 −9 ）
  単語誤り率：**12%**
  単語認識率：**88%**（83%から5%アップ）
- 日本人による読み上げタスクのAI認識率
  レーベンシュタイン距離 = 74（76−2）
  単語誤り率：41%
  単語認識率：59%（58%から1%アップ）

日本人の場合、同じ語が2回認識されている。言いよどみ、言い直しがそのまま文字化されている可能性。

自由会話の単語認識率

- トルコ人による自由会話のAI認識率
  レーベンシュタイン距離 = **121**
  単語誤り率：**70%**
  単語認識率：**30%（会話全体では52%）**
- 日本人による自由会話のAI認識率
  レーベンシュタイン距離 = **165**
  単語誤り率：**93%**
  単語認識率：**7%（会話全体では39%）**

GCP VS AWS 読み上げタスクの単語認識率

- GCP:トルコ人による読み上げタスクのAI認識率
  レーベンシュタイン距離 = **30**
  単語誤り率:**17%** = 30 ／ 181
  単語認識率:**83%** = 1 − 0.17
- AWS:トルコ人による読み上げタスクのAI認識率
  レーベンシュタイン距離 = **31**
  単語誤り率:**17%** = 31 ／ 181
  単語認識率:**83%** = 1 − 0.17

考察および今後の利用について

- AIによる音声認識の性能を最大限利用できるのは、話者の言語が学習された言語と正確に一致する場合。
  一致した場合は、1時間の音声を5分程度で95%以上の認識率の文字化が可能。
  Chambers先生のスピーチ
  http://www.coelang.tufs.ac.jp/dev/SpeechAPI/en-IN/message.html
  http://www.coelang.tufs.ac.jp/mt/en/ch_message/
- 音響モデル、言語モデル両方に言えるが、学習されたモデルにマッチしない場合は期待した認識率を発揮できない。
  例えば、日本人のトルコ語学習者。方言を使用する母語話者。
  http://www.coelang.tufs.ac.jp/dev/SpeechAPI/en-gb-sct/
  http://www.coelang.tufs.ac.jp/dev/SpeechAPI/en-gb-wls/

- GCP Speech-to-Textの認識モデルの細分化
  対応言語モデル289。英語だけでも37も認識モデルが存在する。
  音響モデルの違い（電話音声、ビデオ音声・・・）、言語モデルの違い（コマンドと検索）
  https://cloud.google.com/speech-to-text/docs/languages
  英語（オーストラリア）、英語（オーストラリア）、英語（カナダ）、英語（カナダ）、英語（ガーナ）、英語（ガーナ）、英語（香港）、英語（香港）、英語（インド）、英語（インド）、英語（アイルランド）、英語（アイルランド）、英語（ケニア）、英語（ケニア）、英語（ニュージーランド）、英語（ニュージーランド）、英語（ナイジェリア）、英語（ナイジェリア）、英語（パキスタン）、英語（パキスタン）、英語（フィリピン）、英語（フィリピン）、英語（シンガポール）、英語（シンガポール）、英語（南アフリカ）、英語（南アフリカ）、英語（タンザニア）、英語（タンザニア）、英語（英国）、英語（英国）、英語（英国）、英語（英国）、英語（米国）、英語（米国）、英語（米国）、英語（米国）、英語（米国）
  学習者の文字化で利用するのであれば、学習者の音声を集めて追加学習を行う必要がありそう。
- AIの性能評価として今回の実験について
  AIは言いよどみ、言い直しなども文字化してしまうため、AIの評価として低めの数値が出てしまった部分もある。
  学習者のトルコ語の認識率が一様に低めとなったことは否定できない。

- 会話モジュールの時間情報作成に応用
  http://www.coelang.tufs.ac.jp/dev/SpeechAPI/

- Moodle録音プラグインに応用
  https://mdle2.tufs.ac.jp/moodleJP/mod/puhrec/view.php?id=248

まとめ

AIによる音声認識は使いよう

- 母語話者ではないものはまだ苦手。
  学習者でなくとも方言も苦手。
  追加学習を行った専用のモデルがリリースされるまではあまり期待できる結果はでない。
- 自由会話の認識は特に苦手。
- 使えるところでは積極的使う価値あり。