

MBA⁺

Artificial Intelligence & Machine Learning

MBA⁺

Audição Cognitiva

Prof . André Silva de Carvalho

Email: profandre.carvalho@fiap.com.br

www.linkedin.com/in/andresilvadecarvalho

<http://lattes.cnpq.br/6876528572507972>

2019



Atividade Final

Vamos analisar o áudio

Para realizarmos esta atividade deveremos utilizar os arquivos de áudio disponibilizados na área da atividade final.

São dois áudios diferentes denominados:

- 1) AUDIO_UM
- 2) AUDIO_REAL

Conforme apresentado em aula o formato do dado pode comprometer sobremaneira o processo de análise do áudio.

Para identificar o formato do áudio é solicitado fornecer, no mínimo, os dados abaixo de cada áudio:

- Formato
- Taxa de bits
- Número de canais
- Taxa de amostragem

Como medir a qualidade da transformação dos dados

Para esta finalidade utilizaremos o cálculo conhecido como Taxa de Erro de Palavras (WER)

$$WER = 100\% \times \frac{S + D + I}{|W|}$$

Onde:

S = substituição

D = exclusão

I = inserção

|W| = número de palavras total

O primeiro erro, conhecido como **substituição (S)**, acontece quando uma palavra incorreta é reconhecida no lugar de uma palavra falada corretamente.

O segundo erro, conhecido como **exclusão (D)**, ocorre quando uma palavra falada não é reconhecida.

Finalmente, o terceiro erro, conhecido como **inserção (I)**, acontece quando palavras extras são estimadas pelo reconhecedor.

| A avaliação

A partir dos áudios fornecidos realizar a transcrição e apresentar para cada áudio o valor dos elementos do cálculo

WER:

S =

D =

I =

W =

| A atividade 1

Como são dois áudios deverão ser apresentados os parâmetros para cada áudio:

AUDIO_UM

S =

D =

I =

W =

AUDIO_REAL

S =

D =

I =

W =

A atividade 2

Comparando a transcrição com o arquivo original pede-se para avaliar se a ideia central das mensagens estão preservadas, considerando que o arquivo transcrito sofreu efeitos do erro da transcrição pelo speech-to-text da IBM. Apresente a ideia central de cada texto.

Justifique a sua análise apresentando o modelo de análise utilizado.

A base dos áudios

FIAP

ESPECIAL

“O idoso não é um ser de outro planeta. É um consumidor normal, com hábitos normais, que interage nas redes, sociais, usa tablet, computadores e smartphones. Eu piloto drones!”, assegura o publicitário Roberto Duailibi, o “D” da agência DPZ&T, que aos 83 anos de idade continua trabalhando intensamente, viajando, dando palestras e produzindo muito conteúdo. “O mundo mudou e todos os países avançados já se deram conta de que essa nova realidade precisa ser vista e aproveitada do ponto de vista do consumo e do marketing.”

Um bom exemplo disso é a campanha da Vodafone, que há dois anos transformou duas idosas viúvas da Romênia em celebridades por meio do projeto Sunday Grannies. Na ocasião, a McCann Bucharest criou uma página no Facebook para as vovós divulgarem seus dotes culinários e convidarem os estudantes da região para uma refeição caseira em sua casa. A iniciativa fez tanto sucesso que a

Será disponibilizado a fonte em texto dos dois áudios.

Esta será a base de análise da qualidade da transcrição

O arquivo em formato pdf estará na área relativa ao trabalho

É o que comprova Marcelo Oste, diretor de marketing da CVC: “Em 2010, os idosos representavam 15% das vendas de pacotes na CVC. Hoje, esse público já equivale a 20% das vendas e dos embarques. Eles já somam quase 1 milhão de turistas por ano. E a perspectiva é de contínuo crescimento”.

Ainda assim, são poucas as campanhas publicitárias e ações de marketing direcionadas a esse target, que muitas vezes tem dinheiro e tempo para gastar, mas não encontra no mercado nacional produtos e serviços capazes de satisfazer os seus desejos. “O mercado de hoje ainda não sabe o que fazer para a camada da população mais velha”, afirma Armando Ferrentini, presidente da Editora Referência, que publica o jornal *Propmark*. “O mundo dos negócios não acordou para esse target, pois, acredito, as marcas não querem se associar aos mais velhos.”

Na tentativa de quebrar esse tabu, recentemente duas grandes montadoras da indústria automobilística colocaram os idosos como protagonistas de suas ações de marke-

Como transcrever

Utilizaremos a demo da IBM que realiza a transcrição online.

<https://speech-to-text-demo.ng.bluemix.net>

Transcribe Audio

- Use your microphone to record audio. For best results, use broadband models for microphone input.
- Upload pre-recorded audio (.mp3, .mpeg, .wav, .flac, or .opus only).
- Play one of the sample audio files.*

*Both US English broadband sample audio files are covered under the Creative Commons license.

The returned result includes the recognized text, word alternatives, and spotted keywords. Some models can detect multiple speakers; this may slow down performance.


Voice Model:


US English broadband model (16KHz) ▼


Keywords to spot:


IBM,admired,AI,transformations,cognitive,Artificial Intelligence,dæ

☒ Detect multiple speakers

 Record Audio

 Upload Audio File

 Play Sample 1

 Play Sample 2

Text

Word Timings and Alternatives

Keywords (0/0)

JSON

Como transcrever

This system is for demonstration purposes only and is not intended to process Personal Data. No Personal Data is to be entered into this system as it may not have the necessary controls in place to meet the requirements of the General Data Protection Regulation (EU) 2016/679.

By using this application, you agree to the [Terms of Use](#)

Transcribe Audio

- Use your microphone to record audio. For best results, use broadband models for microphone input.
- Upload pre-recorded audio (.mp3, .mpeg, .wav, .flac, or .opus only).
- Play one of the sample audio files.*

*Both US English broadband sample audio files are covered under the Creative Commons license.

The returned result includes the recognized text, word alternatives, and spotted keywords. Some models can detect multiple speakers; this may slow down performance.

O Watson na versão demo somente trabalha com os formatos:

- Mp3
- Mpeg
- Wav
- Flac
- Opus

Voice Model:

US English broadband model (16KHz)



https://cloud.ibm.com/docs/services/speech-to-text?topic=speech-to-text-output#speaker_labels

Keywords to spot:

IBM,admired,AI,transformations,cognitive,Artificial Intelligence,d

Como transcrever

https://speech-to-text...50200000=85734591550272329045&cm_mc_sid_52640000=20417861550272329072

Apps Portal Corporativo ...

Brazilian Portuguese narrowband model (8KHz)
Korean broadband model (16KHz)
French broadband model (16KHz)
US English narrowband model (8KHz)
Korean narrowband model (8KHz)
Brazilian Portuguese broadband model (16KHz)
French narrowband model (8KHz)
Mandarin broadband model (16KHz)
Japanese broadband model (16KHz)
GB English narrowband model (8KHz)
Spanish broadband model (16KHz)
Modern Standard Arabic broadband model (16KHz)
Mandarin narrowband model (8KHz)
Japanese narrowband model (8KHz)
Spanish narrowband model (8KHz)
US English Short Form narrowband model (8KHz)
German narrowband model (8KHz)
German broadband model (16KHz)
GB English broadband model (16KHz)
US English broadband model (16KHz)

US English broadband model (16KHz)

Keywords to spot:
IBM,admired,AI,transformations,cognitive,Artificial Intelligence,d

☐ Detect multiple speakers

Record Audio Upload Audio File Play Sample 1 Play Sample 2

Text Word Timings and Alternatives Keywords (0/0) JSON

Preferências de cookies

Como transcrever

É muito fácil e usar.

Transcribe Audio

- Use your microphone to record audio. For best results, use broadband models for microphone input.
- Upload pre-recorded audio (.mp3, .mpeg, .wav, .flac, or .opus only).
- Play one of the sample audio files.*

*Both US English broadband sample audio files are covered under the Creative Commons license.

The returned result includes the recognized text, word alternatives, and spotting speakers; this may slow down performance.

Realize aqui o upload dos áudios

Voice Model:

US English broadband model (16KHz)

Keywords to spot:

IBM,admired,AI,transformations,cognitive,Artificial Intelligence,dæ

☒ Detect multiple speakers



Record Audio



Upload Audio File



Play Sample 1



Play Sample 2

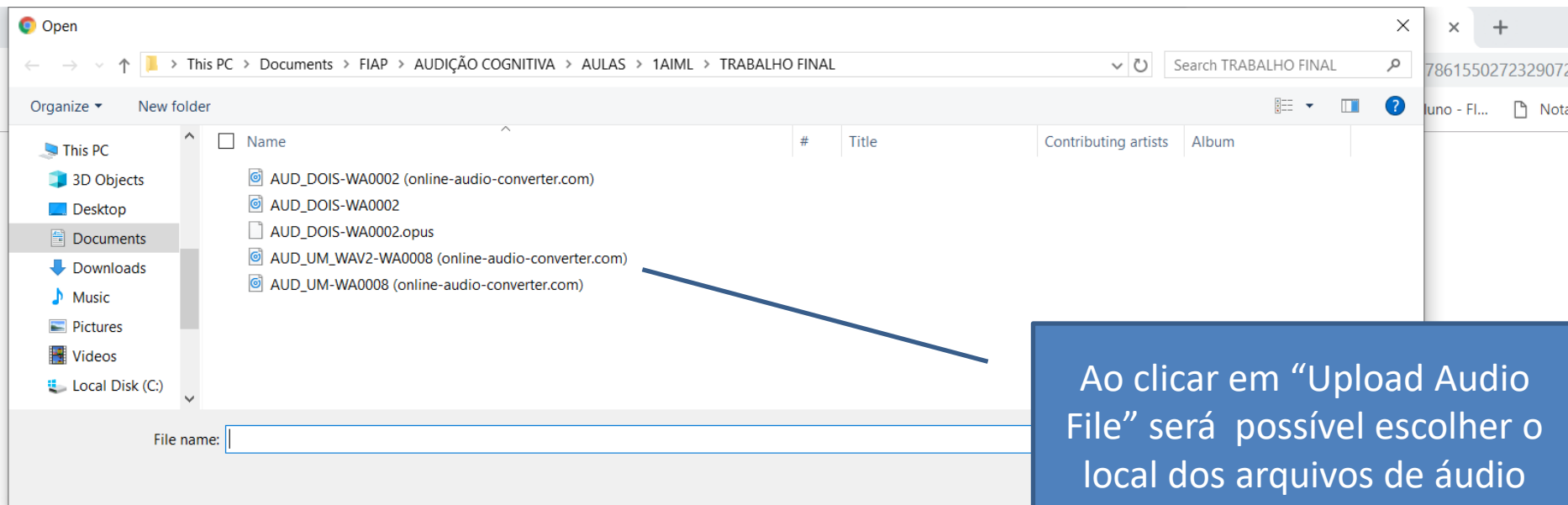
Text

Word Timings and Alternatives

Keywords (0/0)

JSON


Como transcrever





Voice Model: US English broadband model (16KHz) ▼


Keywords to spot: IBM,admired,Al,transformations,cognitive,Artificial Intelligence,dæ

☒ Detect multiple speakers

 Record Audio

 Upload Audio File

 Play Sample 1

 Play Sample 2

Text Word Timings and Alternatives Keywords (0/0) JSON

Como transcrever

Transcribe Audio

- Use your microphone to record audio. For best results, use broadband models for microphone input.
- Upload pre-recorded audio (.mp3, .mpeg, .wav, .flac, or .opus only).
- Play one of the sample audio files.*

*Both US English broadband sample audio files are covered under the Creative Commons license.

The returned result includes the recognized text, word alternatives, and spotted keywords. Some multiple speakers; this may slow down performance.

Voice Model:

Brazilian Portuguese narrowband model (8KHz)

Keywords to spot:

cozinha, inovadoras receitas, criatividade, sistema, treinado por es

☐ Detect multiple speakers (Not supported on current model)



Record Audio



Upload Audio File



Play Sample 1



Play Sample 2

Text

Word Timings and Alternatives

Keywords (0/6)

JSON

Para

Quando o upload for concluído o processo de transcrição ira iniciar automaticamente

Como transcrever

The returned result includes the recognized text, word alternatives, and spotted keywords. Some models can detect multiple speakers; this may slow down performance.

Voice Model:

Brazilian Portuguese narrowband model (8KHz)

Keywords to search

cozinha, ino

☐ Detect multiple speakers (Not supported on current model)



Record Audio



Upload Audio File



Play Sample 1



Play Sample 2

Apesar de não ser possível realizar o ajuste na transcrição alterando as palavras transcritas erradas é possível avaliar as palavras alternativas e as probabilidades

Text

Word Timings and Alternatives

Keywords (0/6)

JSON

para³ a revolução³ não⁵ longevidade <eps>² está transformando profundamente os mercados² de² atuação das
empres 33.43s - 33.8s voltar⁴ é o² consumidor precisam definir implementar estratégias específicas²
grave⁷ de compor a² faixa² demográfica da maior do maior crescimento e poder de
compra • linha: 82.6%
• lilian: 12.5%
a estra • unha: 1.8%
• lilian: 1%
parte d essa³ da da⁵
verdadeiro⁴ cenas² linha⁴ formado a³ revolução da nos vestiários²
dica⁵ silêncio⁶ porque⁴ talvez não excessivo² jean³ pender³ as mão
o² sul⁴ apresentado pelas² novas³ gerações²

Ainda sobre dados...



MediaInfo é um programa grátis que permite ter acesso a informações sobre os arquivos de áudio. Muito útil para avaliar o arquivo.

Arquivos para o trabalho final

MediaArea.net/MediaInfo - C:\Users\AndréCarvalho\Downloads\PTT-20190209-WA0002.opus

Arquivo Exibir Opções Depurar Ajuda Idioma

▼ C:\Users\AndréCarvalho\Downloads\PTT-20190209-WA0002.opus

▼ Geral

Nome completo: C:\Users\AndréCarvalho\Downloads\PTT-20190209-WA0002.opus
Formato: Ogg
Tamanho do arquivo: 480 KiB
Duração: 3 min 28s
Taxa de Bits Total: 18.8 kb/s

▼ Áudio

ID: 100 (0x64)
Formato: Opus
Duração: 3 min 28s
Nº de canais: 1 canal
Channel layout: C
Taxa de amostragem: 16.0 kHz
Biblioteca usada: WhatsApp

MediaArea.net/MediaInfo - C:\Users\AndréCarvalho\Downloads\AUD-20190209-WA0008.m4a

Arquivo Exibir Opções Depurar Ajuda Idioma

▼ C:\Users\AndréCarvalho\Downloads\AUD-20190209-WA0008.m4a

▼ Geral

Nome completo: C:\Users\AndréCarvalho\Downloads\AUD-20190209-WA0008.m4a
Formato: MPEG-4
Perfil do Formato: Apple audio with iTunes info
ID do Codec: M4A (M4A/mp42/Isom)
Tamanho do arquivo: 2.49 MiB
Duração: 5 min 15s
Taxa de Bits Total, Modo: Constante
Taxa de Bits Total: 66.2 kb/s
Data da codificação: UTC 2019-02-09 13:52:50
Data rotulada: UTC 2019-02-09 13:52:50
Programa usado: com.apple.VoiceMemos (iOS 12.1.4)

▼ Áudio

ID: 1
Formato: AAC LC
Formato/Informações: Advanced Audio Codec Low Complexity
ID do Codec: mp4a-40-2
Duração: 5 min 15s
Source_Duration/String: 5 min 15s
Modo da taxa de bits: Constante
Taxa de bits: 64.0 kb/s
Nº de canais: 1 canal
Channel layout: C
Taxa de amostragem: 48.0 kHz
Taxa de quadros: 46.875 FPS (1024 SPF)
Tamanho da Faixa: 2.43 MiB (98%)
Source_StreamSize/String: 2.43 MiB (98%)
Título: Core Media Audio
Data da codificação: UTC 2019-02-09 13:52:50
Data rotulada: UTC 2019-02-09 13:52:50



**Obrigado pela jornada
juntos!**

MBA⁺

Copyright © **2016** Prof.

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).