

# AI & CHATBOT

Aula 07 – Reconhecimento de Fala

Prof. Henrique Ferreira (manhã)

Prof. Marcelo Grave (noite)

FIAP  
GRADUAÇÃO

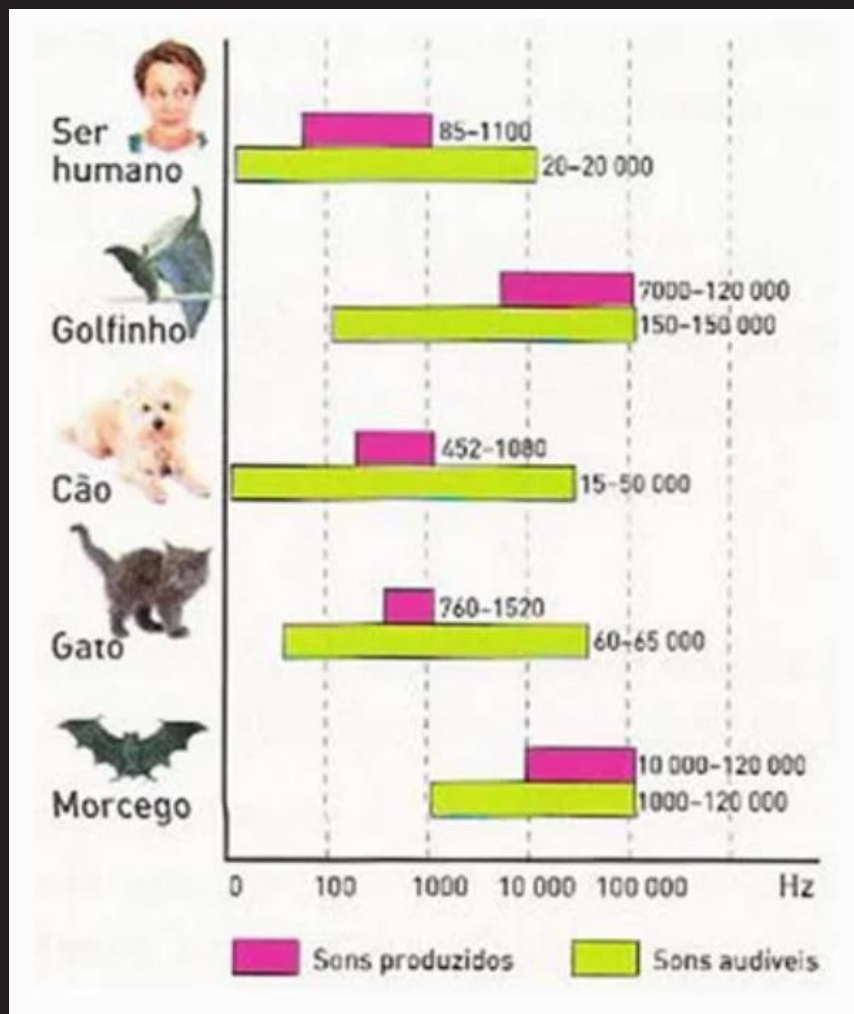
# O que é som?

Entendendo é que som e áudio

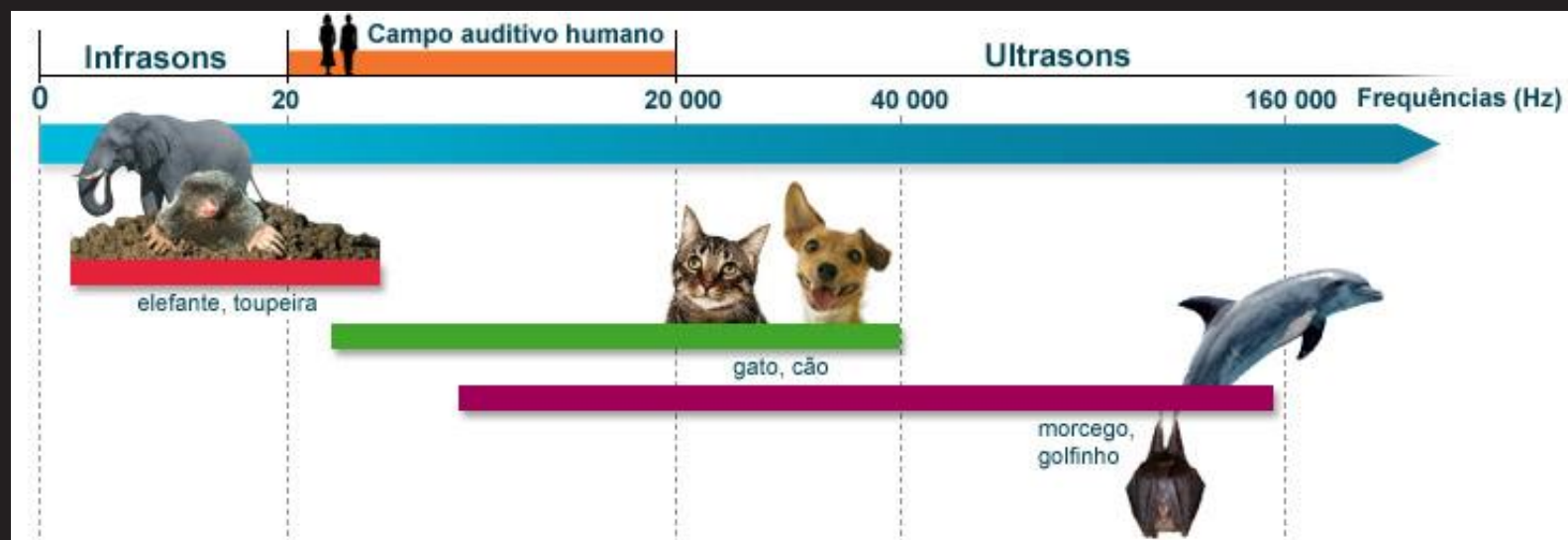
# O que é som?

- Som é uma onda de pressão no ar (ondas acústicas);
- Seres humanos e outros animais modulam as ondas de pressão do ar de maneira a codificar informação nelas;
- Fala é a modulação de onda sonoras através de uma linguagem natural.
- Seres humanos são capazes de perceber (ouvir) essas ondas de pressão e de decodificá-las (através da língua);
- Uma etapa fundamental da alfabetização é o desenvolvimento da capacidade de transformar sons em símbolos (letras, palavras e frases);

# O que é som?

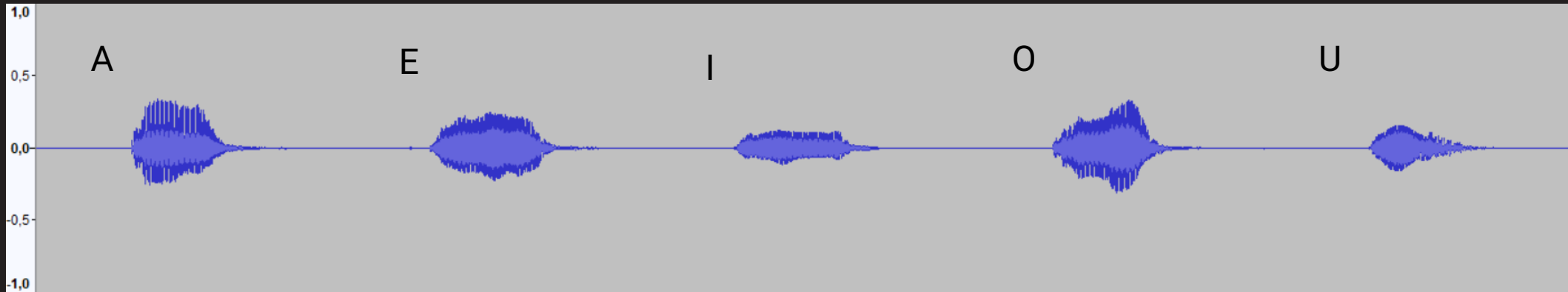


- Toda onda tem uma frequência de oscilação;
- Animais e dispositivos tem faixas de frequência de operação (produção e captação de som);

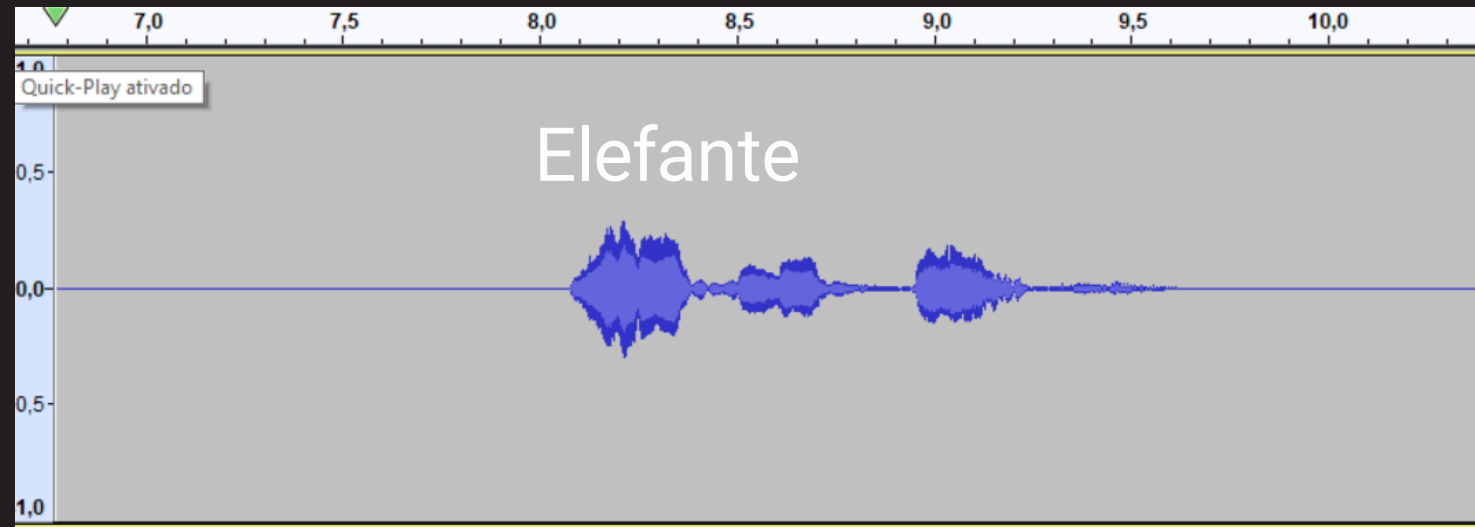


# O que é som?

Sinal de áudio digital no domínio do tempo:



# O que é som?



A

E

I

O

U

# Como som é armazenado no computador?

- **Arquivos digitais de áudio** podem ter vários formatos: wav, mp3, ogg, opus, entre outros;
- Os formatos envolvem um **codec** (um padrão de codificação) que pode configurar **compressão com perdas** ou não:
  - MP3 é um padrão de codificação de áudio com perdas!
  - Arquivos WAV são iguais aqueles gravados, sem perdas;
- Além das perdas propositalis (para diminuir o espaço de armazenamento) podem ocorrer perdas no processo de captação, devido a características do microfone e do sistema de amostragem (conversor analógico/digital);

# Reconhecimento de Fala

Introdução ao reconhecimento de fala por IAs



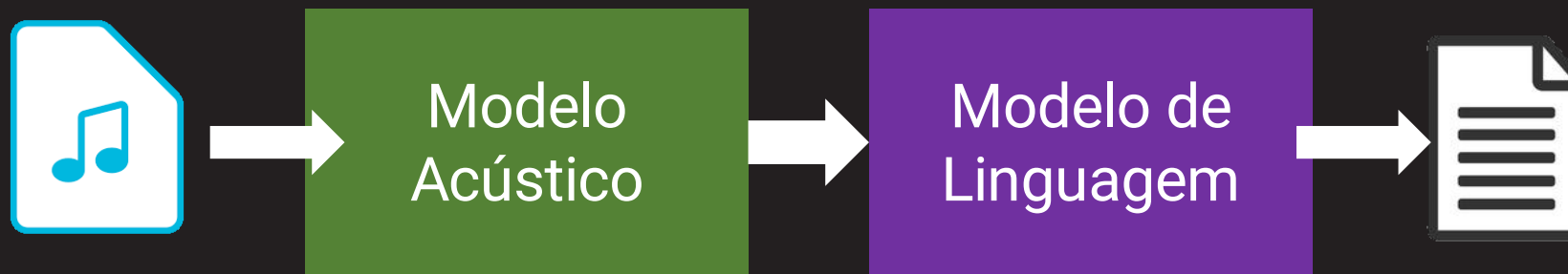
# Reconhecimento de Fala

- O reconhecimento de fala é uma subárea do campo de IA;
- Existem diferentes sistemas capazes de reconhecer fala;
- O reconhecimento de fala envolve transformar arquivos de áudio em arquivos de texto (transcrição);
- Para entender a fundo como esses sistemas funcionam precisamos entender o que é um sinal de áudio, como ele é representado no computador e que tipo de propriedades físicas podemos extrair desses sinais;
- Na disciplina vamos nos ater apenas a utilização de sistemas prontos (soluções de bancada). Não teremos tempo para nos aprofundar nos nossos próprios sistemas de IA e áudio;

# Reconhecimento de Fala

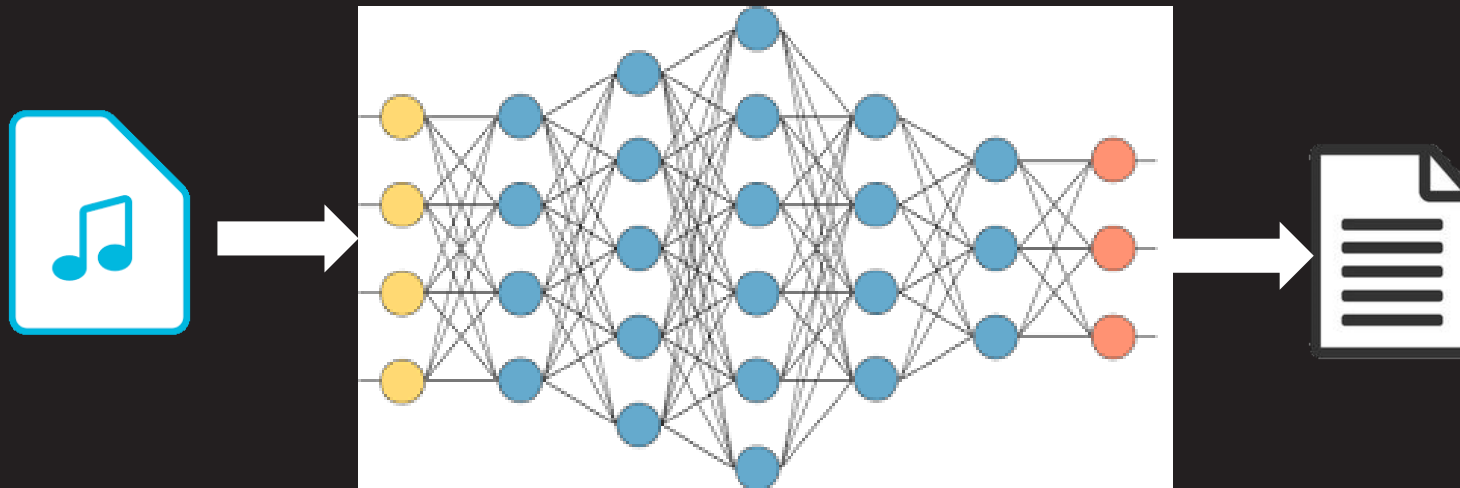
Historicamente, os sistemas de reconhecimento de fala usam duas tecnologias conectadas:

- O modelo acústico possui informações sobre as formas de onda do sinal e sua associação com fonemas;
- O modelo de linguagem possui informações sobre a língua, como a ocorrência estatística de fonemas conectados;



# Reconhecimento de Fala

Ultimamente tem surgido sistemas de redes neurais do tipo end-to-end:

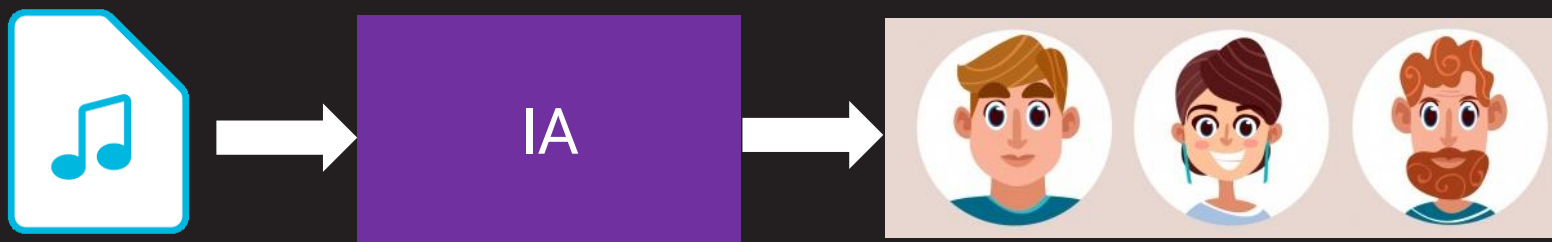


# Reconhecimento de Fala vs Reconhecimento de Voz

**Reconhecedor de Fala:** transforma áudio em texto



**Reconhecedor de Voz:** tenta identificar quem está falando



# Watson Speech-to-Text

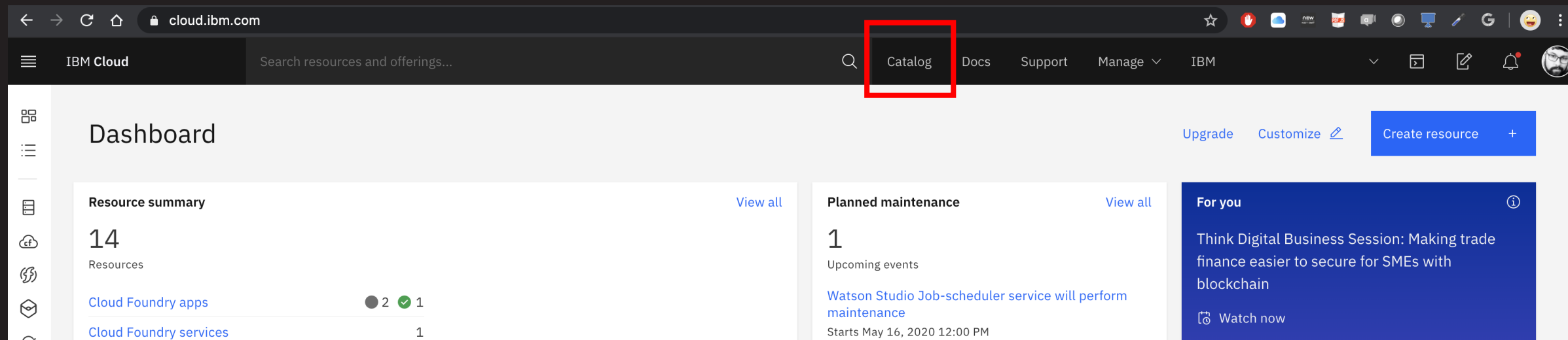
Reconhecedor de fala da IBM

# Watson Speech-to-Text (STT)

- O IBM Watson STT é o serviço de Reconhecimento de Fala disponível para várias línguas;
- Ele realiza a transcrição de áudio para texto;
- Tem suporte à vários formatos de arquivos de áudio;
- Possui funcionalidades adicionais para reconhecimento de keywords, rotulação de voz, customização do modelo de linguagem e do modelo acústico, entre outros;
- No nosso curso não será possível abordar todas as funcionalidades extras;

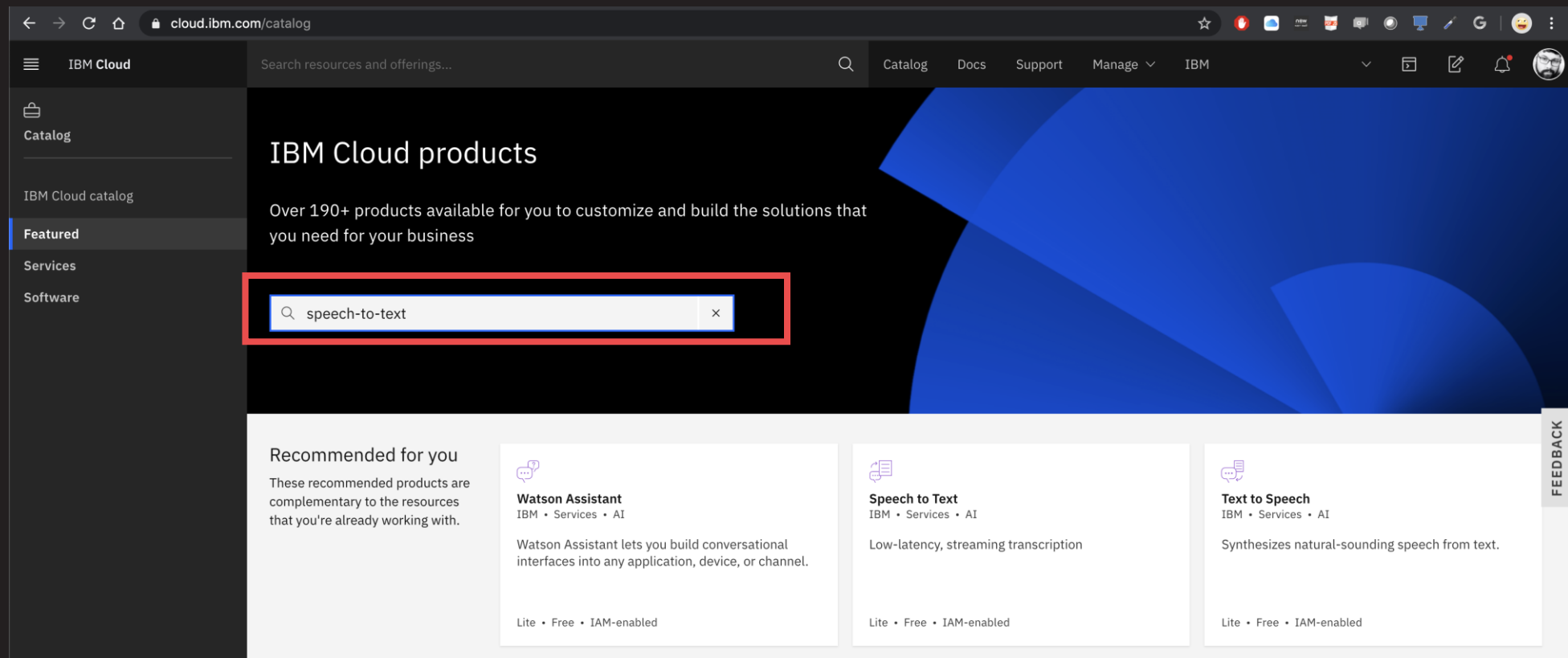
# Criando o Serviço Watson Speech-to-text (STT)

- Para criar o serviço Watson STT e pegar a chave de autorização (apikey) o primeiro passo é fazer o login em <https://cloud.ibm.com> e clicar em Catálogo (“Catalog”):



# Criando o Serviço Watson Speech-to-text (STT)

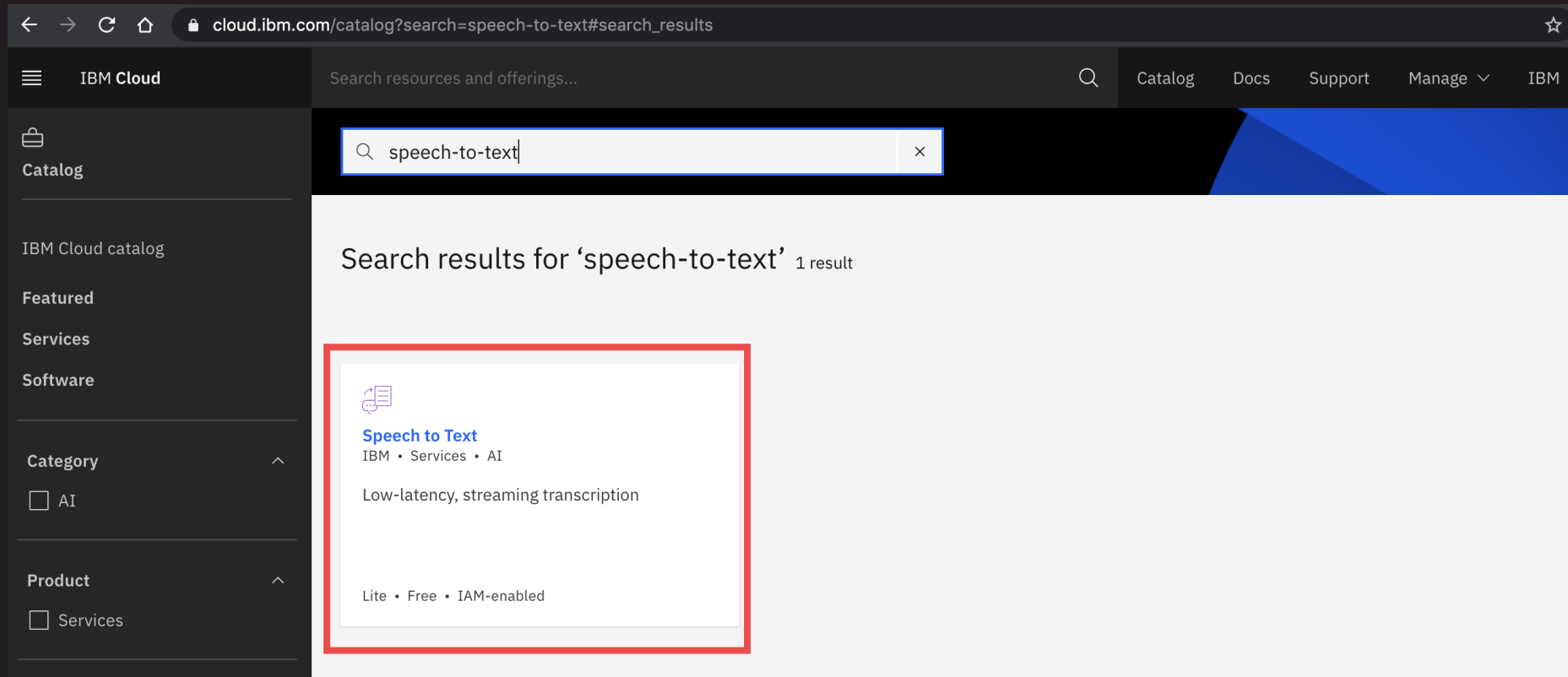
- No catálogo digite Speech-to-text:





# Criando o Serviço Watson Speech-to-text (STT)

- Selecione o serviço do Watson Speech-to-text:



# Criando o Serviço Watson Speech-to-text (STT)

- Clique em Create. Não é necessária nenhuma configuração nesse momento.

**Speech to Text**  
Author: IBM • Date of last update: 04/24/2020 • [Docs](#) • [API docs](#)

**Create** About

Select a region

Select a region

Dallas

Select a pricing plan

Displayed prices do not include tax. Monthly prices shown are for country or region: [United States](#)

Plan	Features	Pricing
Lite	500 Minutes per Month  The Lite plan gets you started with 500 minutes per month at no cost. When you upgrade to a paid plan, you will get access to Customization capabilities. Lite plan services are deleted after 30 days of inactivity.	Free
Standard	Standard Minutes Multi-Tiered	Click to view tiers and pricing detail
Premium	Everything in Standard plus... Usage and Training Data is Private • Stored in an Isolated Single Tenant Environment High Availability and Service Level Uptime Guarantee HIPAA - Washington DC Only IBM Cloud Service Endpoints	

**Create**

Add to estimate

View terms

Summary

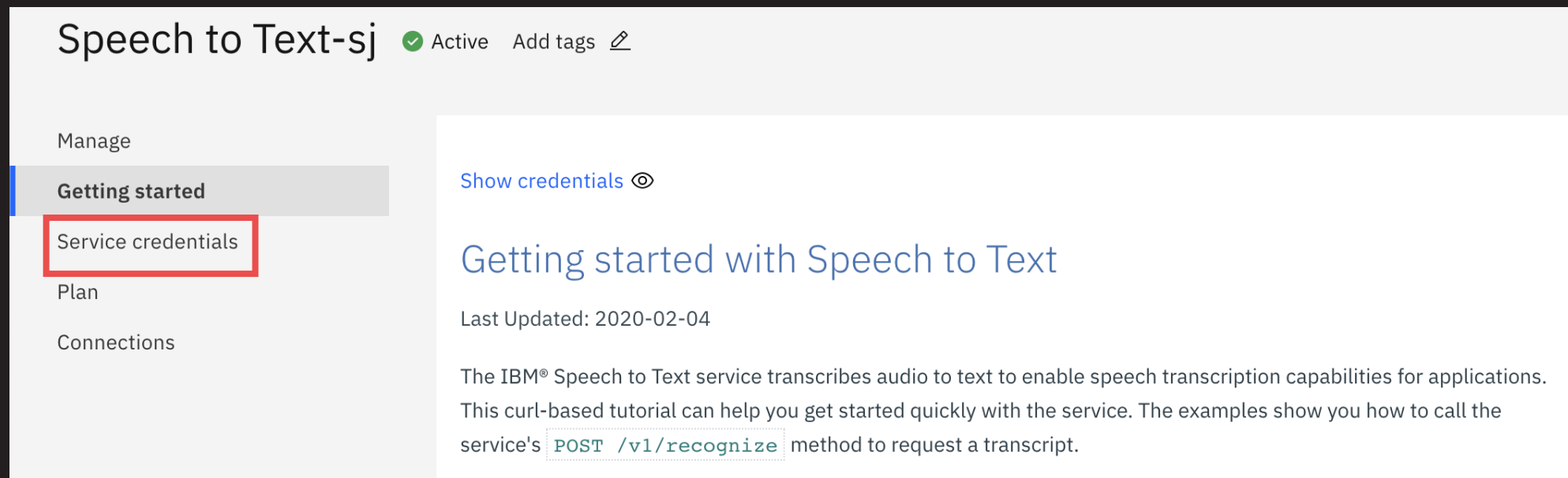
**Speech to Text** Free

Region: Dallas  
Plan: Lite  
Service name: Speech to Text-sj  
Resource group: Default

FEEDBACK

# Criando o Serviço do Watson STT

- Pronto, você criou o seu STT. Agora basta clicar em “Service Credentials” para começarmos a configurar o serviço do Node-red.



Speech to Text-sj ✓ Active [Add tags](#) [✎](#)


Manage

**Getting started**

**Service credentials**

Plan

Connections

[Show credentials](#) 

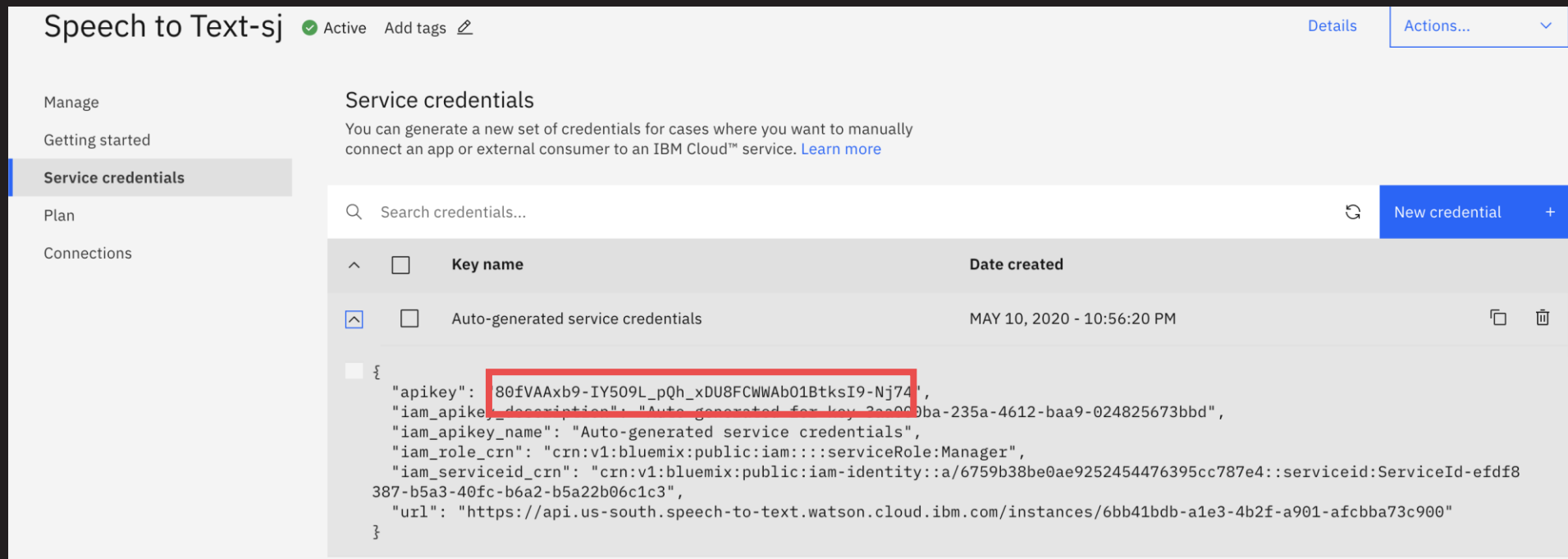
## Getting started with Speech to Text

Last Updated: 2020-02-04

The IBM® Speech to Text service transcribes audio to text to enable speech transcription capabilities for applications. This curl-based tutorial can help you get started quickly with the service. The examples show you how to call the service's `POST /v1/recognize` method to request a transcript.

# Criando o Serviço do Watson STT

- Agora basta copiar e colar o valor “apikey” sem as aspas no atributo do nó do Watson STT do Node-RED.



Speech to Text-sj Active Add tags

Manage  
Getting started  
**Service credentials**  
Plan  
Connections

Service credentials

You can generate a new set of credentials for cases where you want to manually connect an app or external consumer to an IBM Cloud™ service. [Learn more](#)

Search credentials...

New credential +

Key name	Date created
Auto-generated service credentials	MAY 10, 2020 - 10:56:20 PM

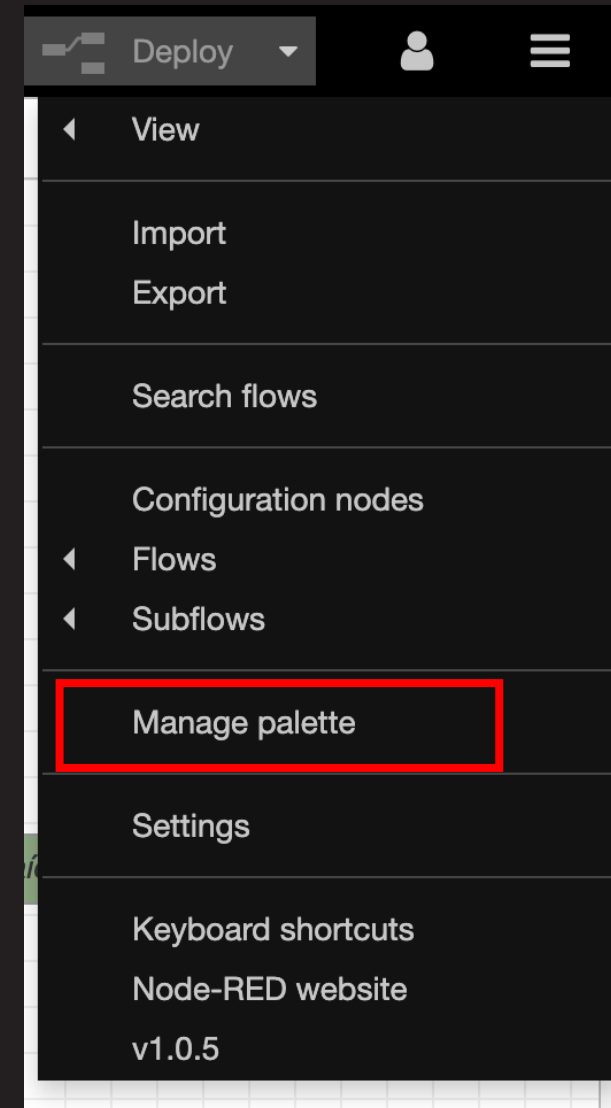
```
{  "apikey": "80fVAAxb9-IY509L_pQh_xDU8FCWWAb01BtksI9-Nj74",  "iam_apikey_description": "Auto-generated for key 2aa0c0ba-235a-4612-baa9-024825673bbd",  "iam_apikey_name": "Auto-generated service credentials",  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Manager",  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a/6759b38be0ae9252454476395cc787e4::serviceid:ServiceId-efdf8387-b5a3-40fc-b6a2-b5a22b06c1c3",  "url": "https://api.us-south.speech-to-text.watson.cloud.ibm.com/instances/6bb41bdb-a1e3-4b2f-a901-afcbba73c900"}
```

# Reconhecendo Fala com Node-RED

Integrando o Watson STT com Node-RED

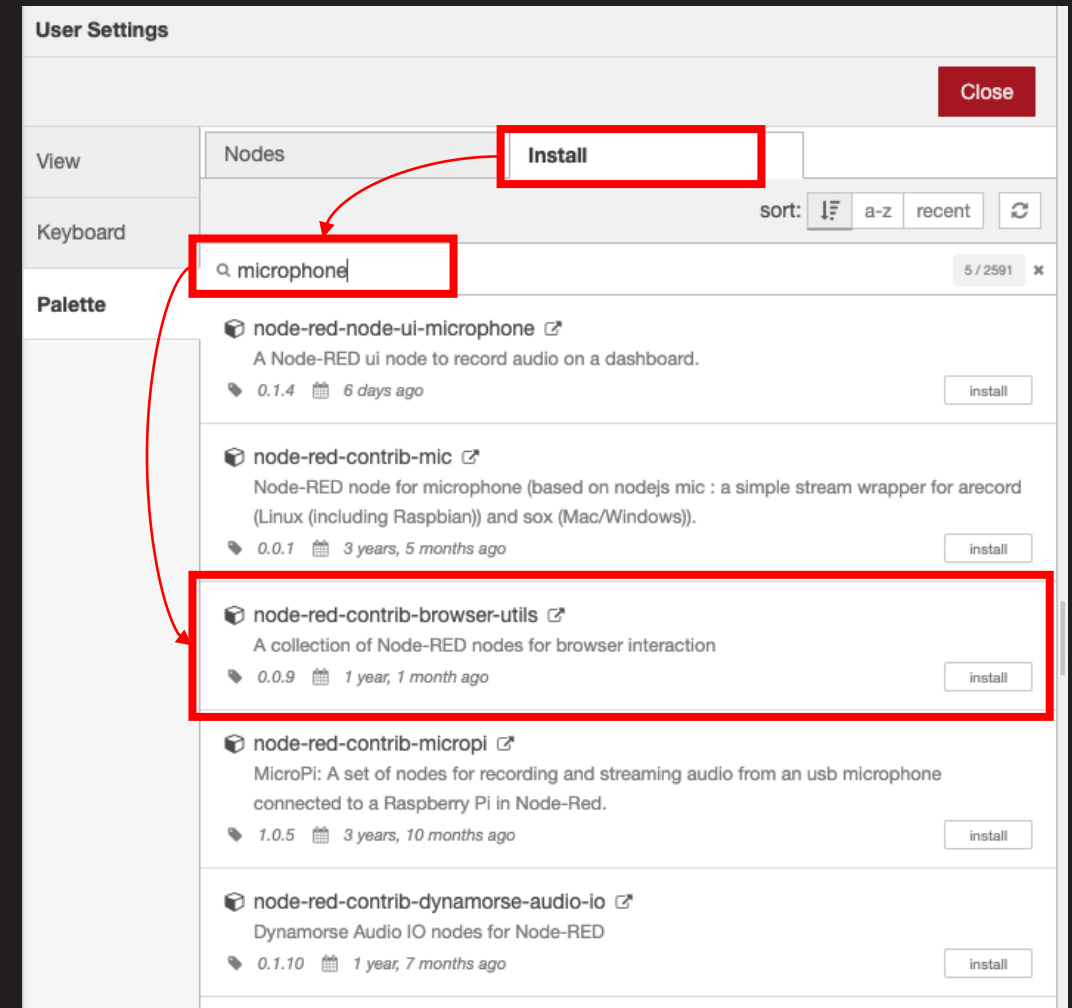
# Adicionando os nós necessários

- Vamos adicionar o nó que corresponde ao microfone para poder usá-lo para testar nossa aplicação.
- Esse nó não existe no Node-RED padrão em vamos instalá-lo manualmente clicando em “Manage palette”



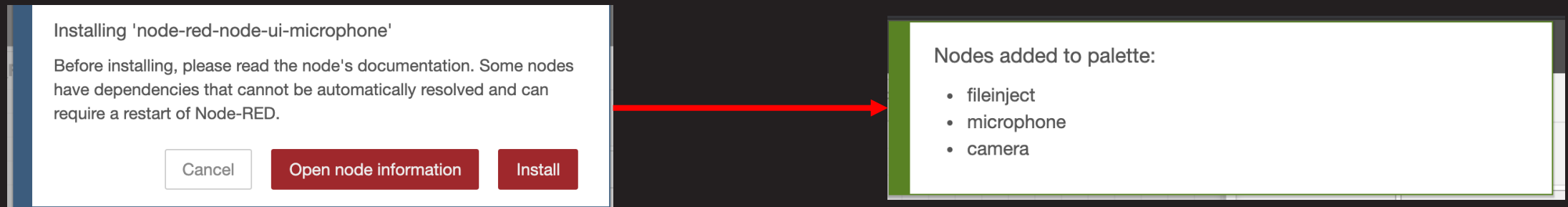
# Adicionando os nós necessários

- Agora clique na aba “Install”, depois digite “microphone” e vamos instalar o pacote “node-red-contrib-browser-utils” clicando em “install”



# Adicionando os nós necessários

- Confirme a instalação clicando em “Install” na janela que se abrirá no topo da tela e logo o pacote estará instalado e uma mensagem em verde, como a mostrada abaixo, aparecerá.





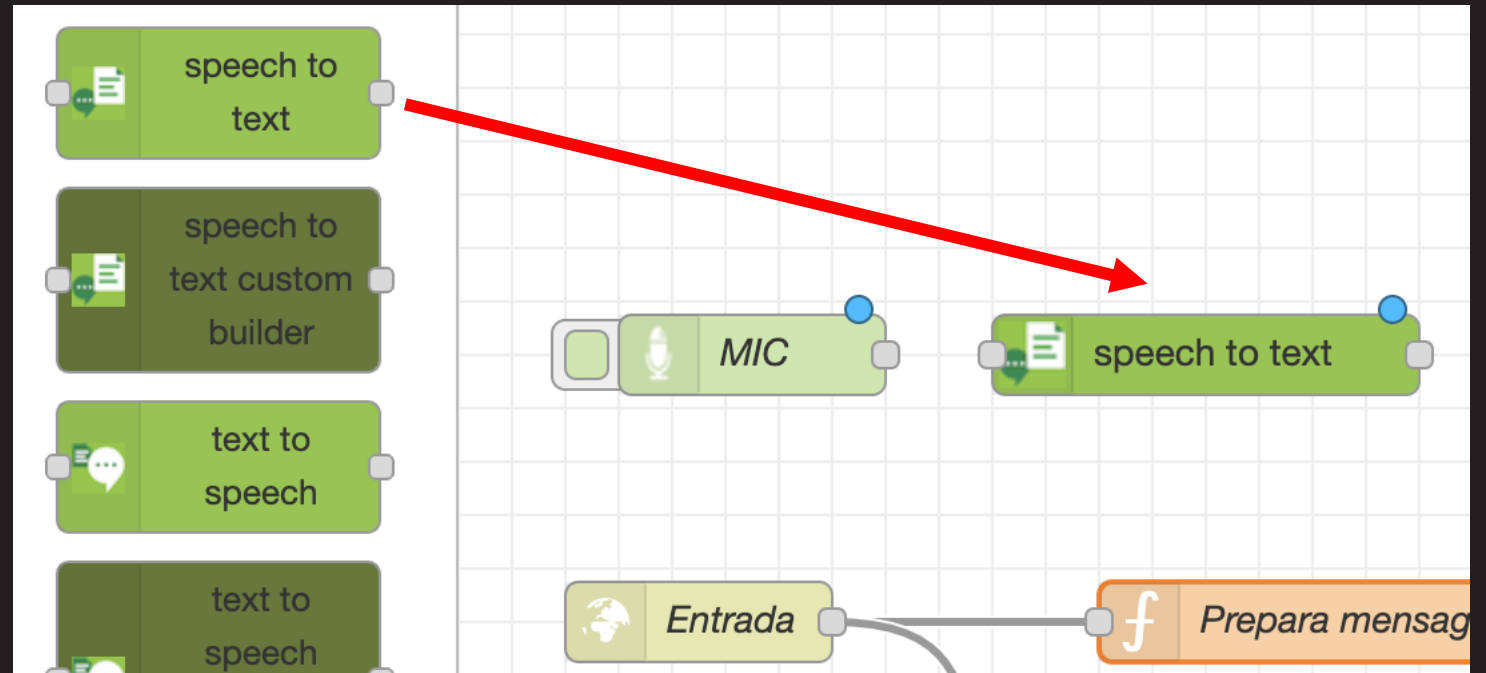
# Adicionando os nós necessários

- O microfone está instalado e agora você pode incluí-lo no fluxo e renomeá-lo.



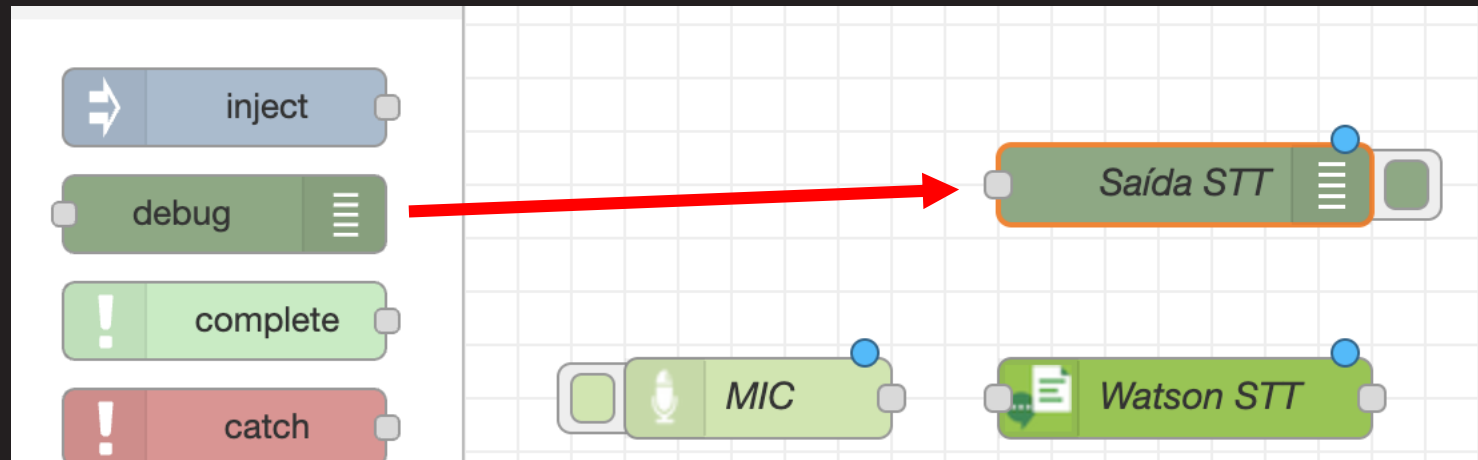
# Adicionando os nós necessários

- Agora podemos incluir o serviço Speech-to-text no fluxo e renomeá-lo.
- Esse serviço será o responsável por fazer a transformação do áudio em texto.



# Adicionando os nós necessários

- E adicionar um Debug para começar a entender o que o Watson STT retornará e renomeá-lo para auxiliar na identificação da saída.



# Adicionando os nós necessários

- Bem, agora o que falta é configurar o Watson STT. Os valores a serem configurados são:
  - Language: Portuguese Brazilian
  - Quality: Broadband Model
  - Marcar “Place output on msg.payload”
  - O único valor externo a ser capturado é API-Key demarcado em vermelho.

OBS: O restante não precisa configurar ainda.

Edit speech to text node

Delete Cancel Done

Properties

Name Watson STT

Username Username

Password Password

API Key .....

Service Endpoint https://api.us-south.speech-to-text.watson.cloud

Language Portuguese Brazilian

Quality BroadbandModel

Max Alternative Transcripts 1

Keywords

Keywords Threshold 0,5

☐ Word Confidence

☐ Speaker Labels

☐ Smart Formatting

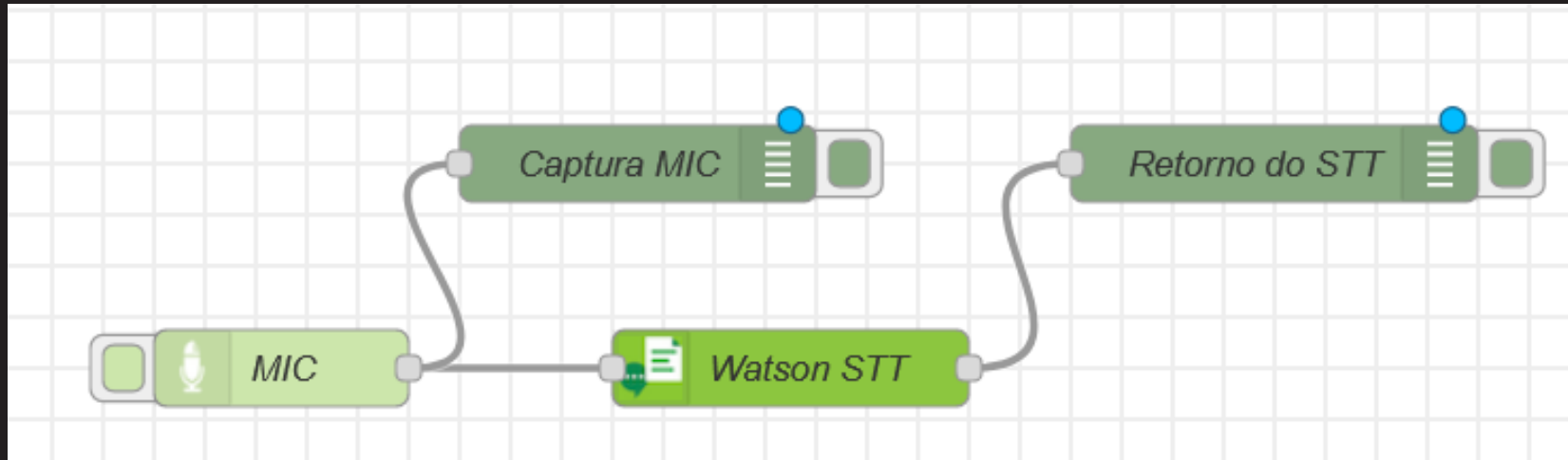
☒ Place output on msg.payload

☐ Disable Audio format pre-check

☐ Streaming Mode

# Fluxo completo

- O fluxo final deve ficar assim:



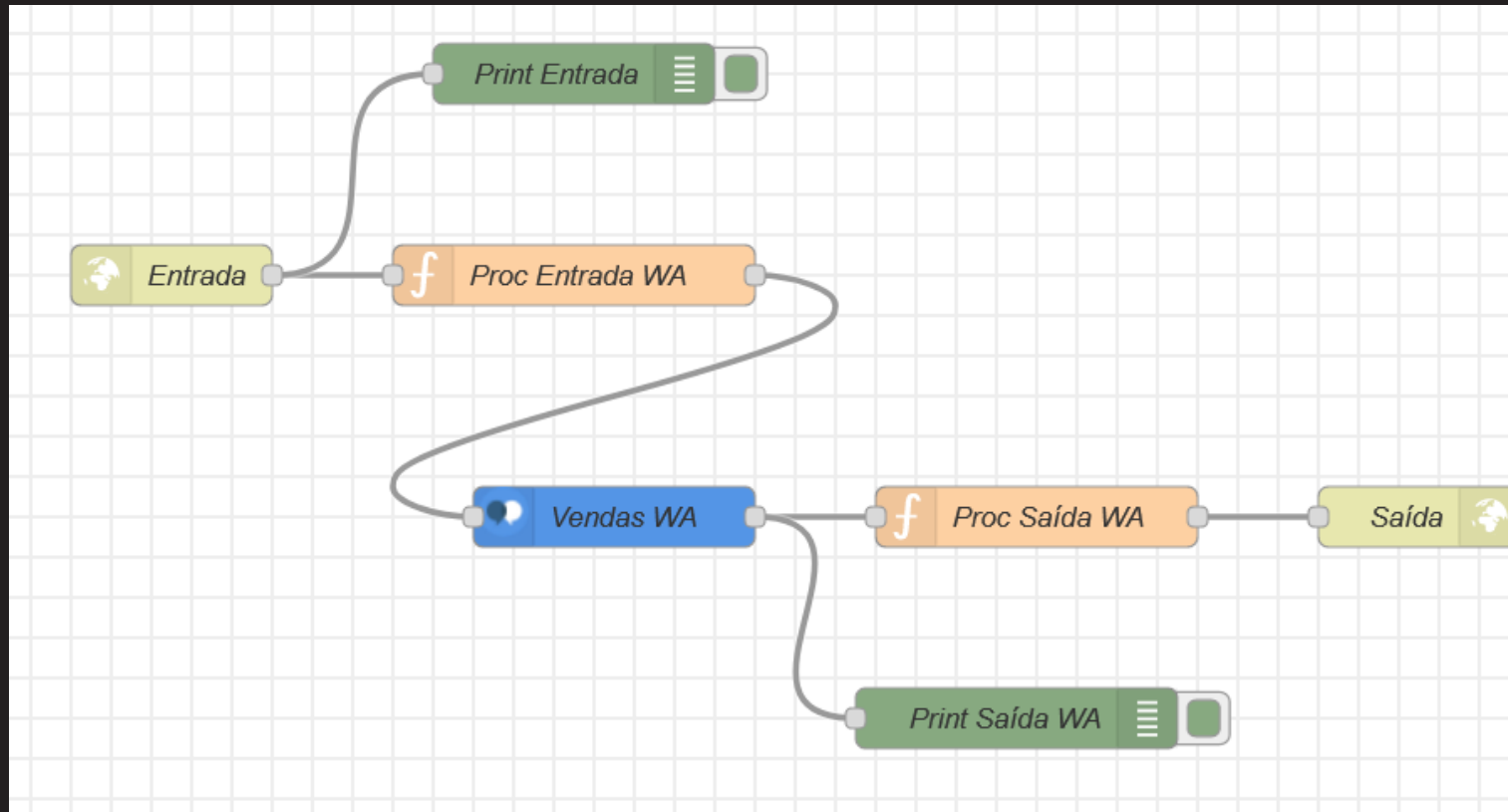
- Dê Deploy e teste!

# Processando a Fala

Integrando o Watson Assistant e o Watson STT com Node-RED

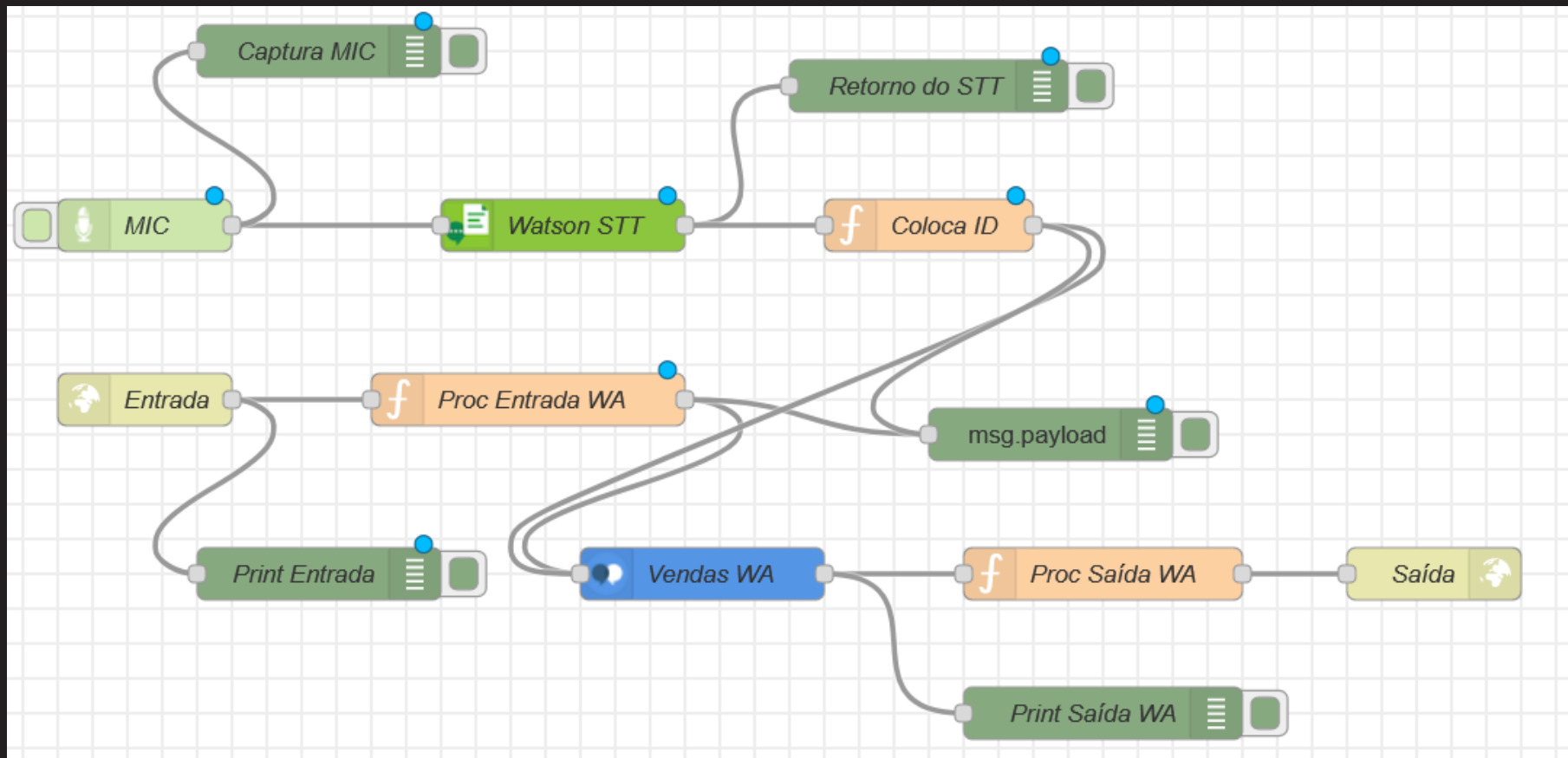
# Integrando o WA e o STT

- Importe no Node-RED o fluxo de integração com a entrada e saída HTTP (último fluxo da aula 05), como da figura abaixo:



# Integrando o WA e o STT

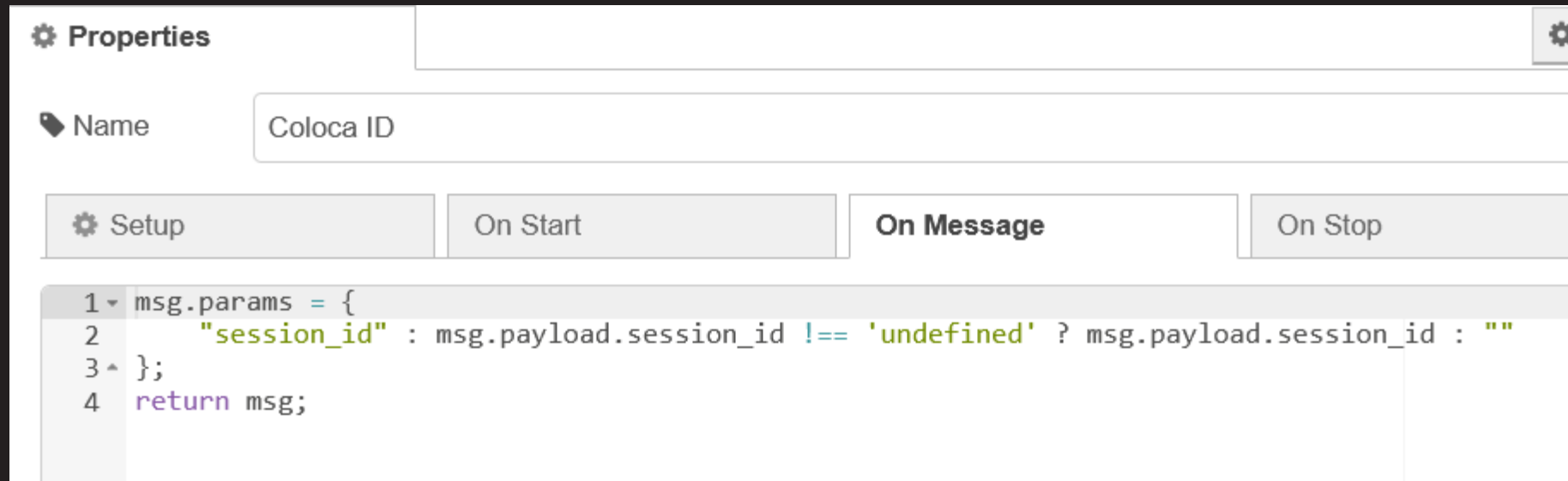
- Adicione um nó de microfone, um nó do STT e um nó de function para preparar a entrada para o WA.





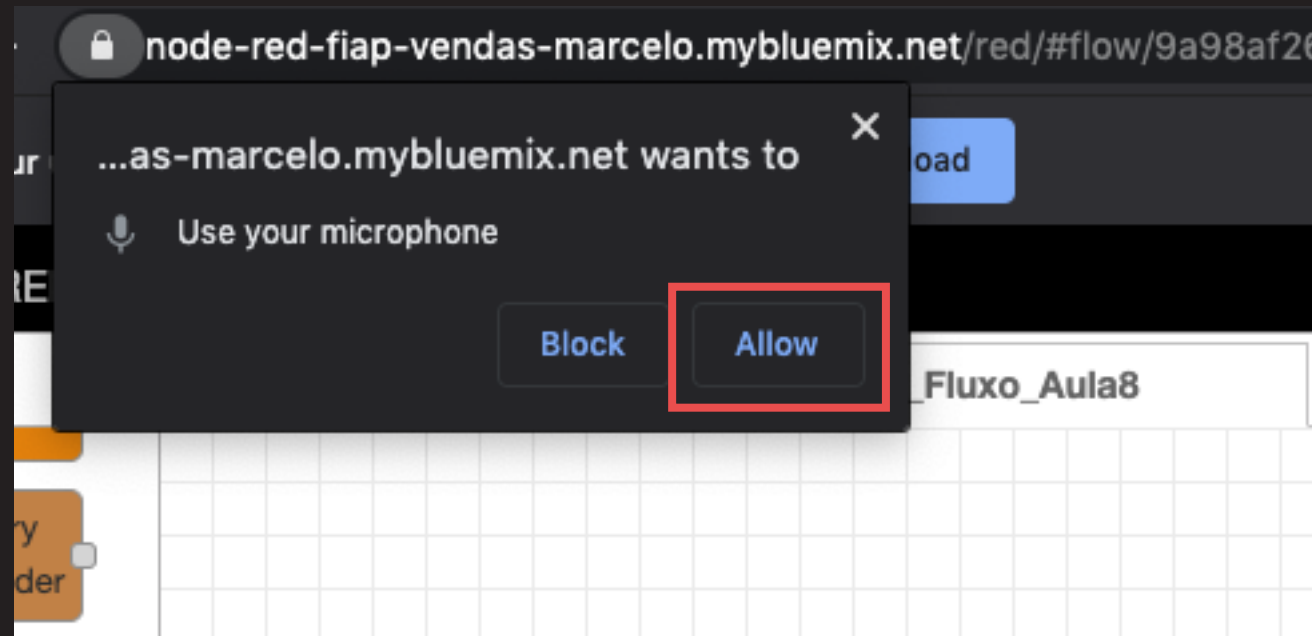
# Integrando o WA e o STT

- Para colocar o ID da sessão basta inserir esse código no nó de function:



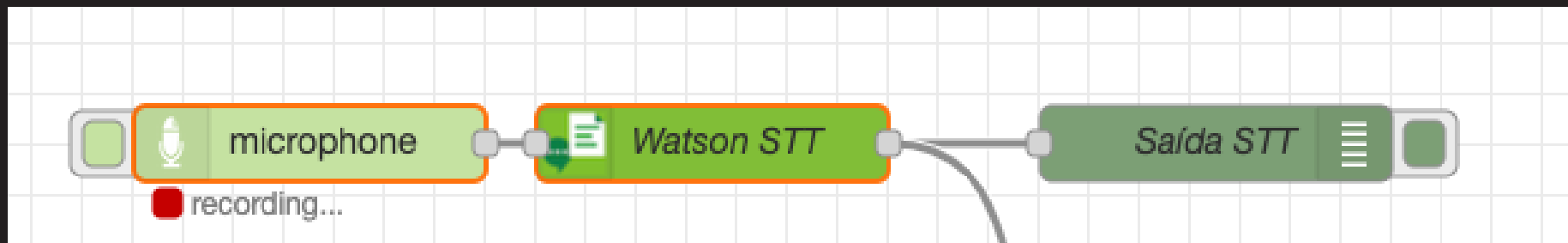
# Testando o assistente

- No momento em que clicar no microfone deverá autorizar o browser a usar o microfone clicando em “Allow” ou “Permito”.



# Testando o assistente

- Depois que que permitir poderá gravar seu áudio. Enquanto o microfone esteve ativo aparecerá um “recording” abaixo do nó, como demonstrado na figura abaixo:



# Testando o assistente

- Depois que terminar de gravar e pressionar o botão novamente o Watson STT fará o processamento do áudio. Quando o áudio chegar por completo no STT aparecerá “requesting” em azul. Após processar o áudio, o STT enviará o resultado para o Watson Assistant seguindo o fluxo do código.



# Testando o assistente

- Podemos visualizar o resultado do na barra lateral per meio dos debugs que incluímos no fluxo.
- Notem que no Debug Saída STT, o que está no “payload” é o texto transcrito pelo STT. Como o texto está no payload será reconhecido automaticamente pelo Watson Assistant. E se caso não tivemos ativado a opção “*Place output on msg.payload*”?
  1. O texto seria reconhecido automaticamente?
  2. Se não fosse reconhecido automaticamente o que deveríamos fazer para garantir isso?

```
5/11/2020, 5:34:53 AM node: Saída STT
msg: Object
  ▾ object
    payload: "quero comprar um celular "
    _msgid: "9c2529b4.f3d228"
    transcription: "quero comprar um celular "
    ▸ fullresult: object

5/11/2020, 5:34:54 AM node: Saída Watson
msg: Object
  ▾ object
    ▸ payload: object
      _msgid: "9c2529b4.f3d228"
      transcription: "quero comprar um celular "
    ▸ fullresult: object

5/11/2020, 5:34:54 AM node: Resposta
msg: string[18]
"No response object"

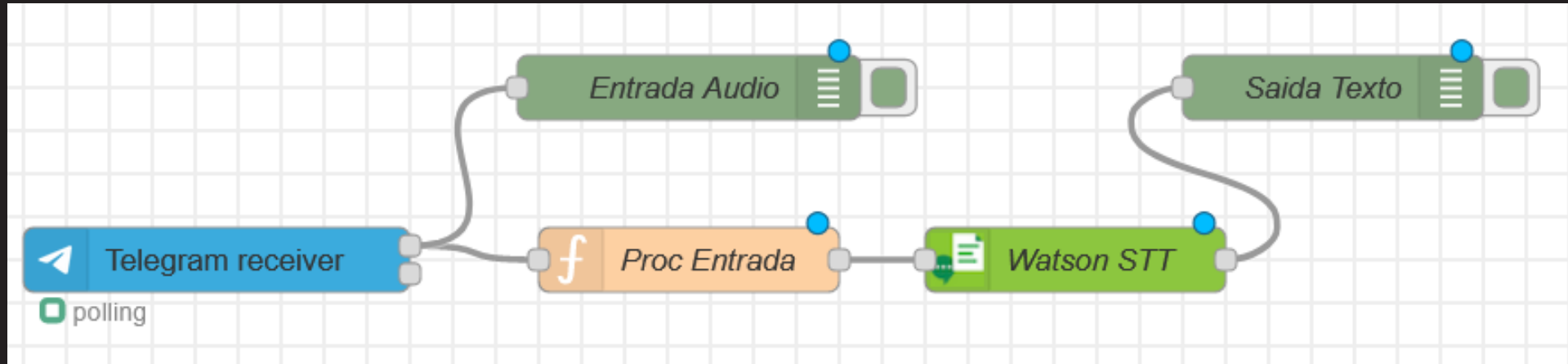
5/11/2020, 5:34:54 AM node: Prepara saída
msg.payload: Object
  ▾ object
    ▾ respostas: array[1]
      ▾ 0: object
        response_type: "text"
        text: "Qual celular você está procurando?"
        session_id: "3321683c-bccc-4fb5-85ca-e79a0cf1e9d4"
```

# Recebendo voz pelo Telegram

Integração entre o Telegram e o STT

# Integrando Telegram e STT

- Prepare o seguinte fluxo:



# Integrando Telegram e STT

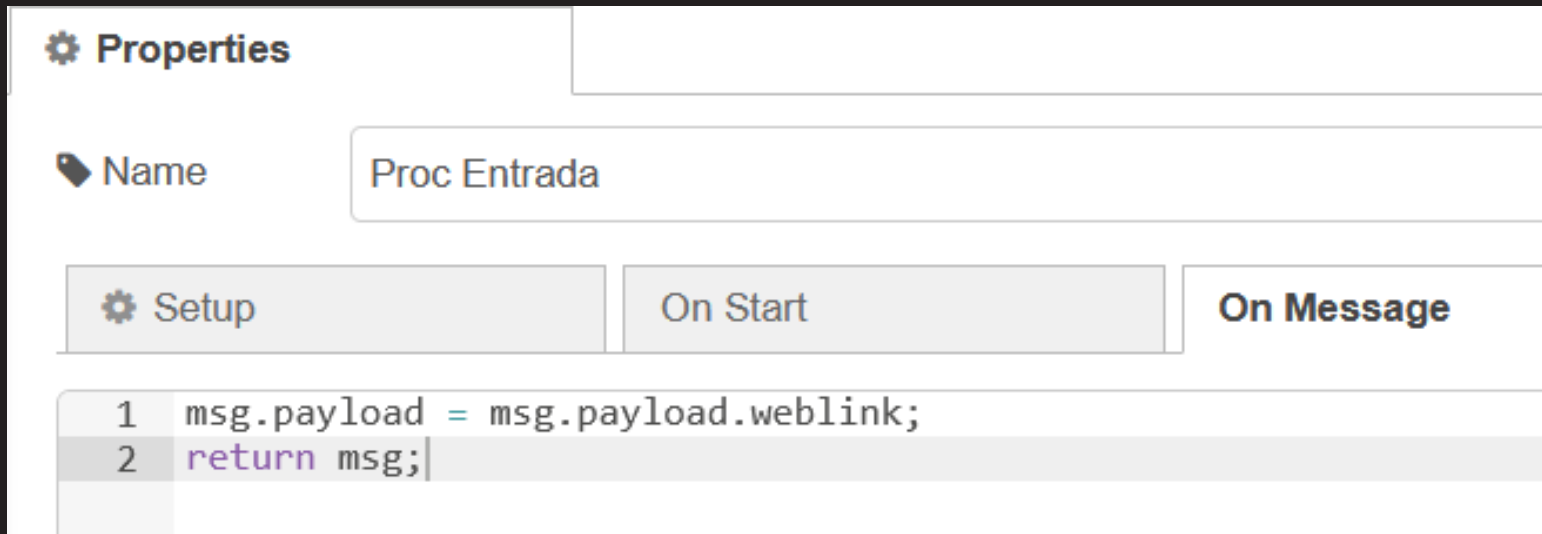
- Antes de programar o nó de function, dê um Deploy e envie uma mensagem de voz para o bot através do Telegram;
- Olhe para a payload (ela deve ser parecida com esta daqui do lado);
- Abra o link que está no atributo weblink da payload;
- O que ele é?

```
msg.payload : Object
  ▼ object
    chatId: 1297157419
    messageId: 314
    type: "voice"
    content:
      "AwACAgEAAxkBAAIBOmBz3L
      LTft3R549wnX-
      jE_1SWxUxAAJAAQAC4MqhRy
      4zT4aSGAssHgQ"
    caption: undefined
    date: 1618205874
    blob: true
    weblink:
      "https://api.telegram.o
      rg
      /file/bot1613485476:AAG
      4CKSTSRChFP_rBuqD9BmfmX
      32buto4sc/voice
      /file_3.oga"
```



# Integrando Telegram e STT

- Olhando para o nó de debug de entrada, percebemos que a função de processamento da entrada deve ser:



# Testando

```
msg.payload : Object
  ▶ { chatId: 1297157419,
    messageId: 316, type:
    "voice", content:
    "AwACAgEAAxkBAAIBPGBz3uBF
    l1960k...", caption:
    undefined ... }
```

```
msg.payload : string[39]
"olá quer ajuda para
comprar um celular "
```

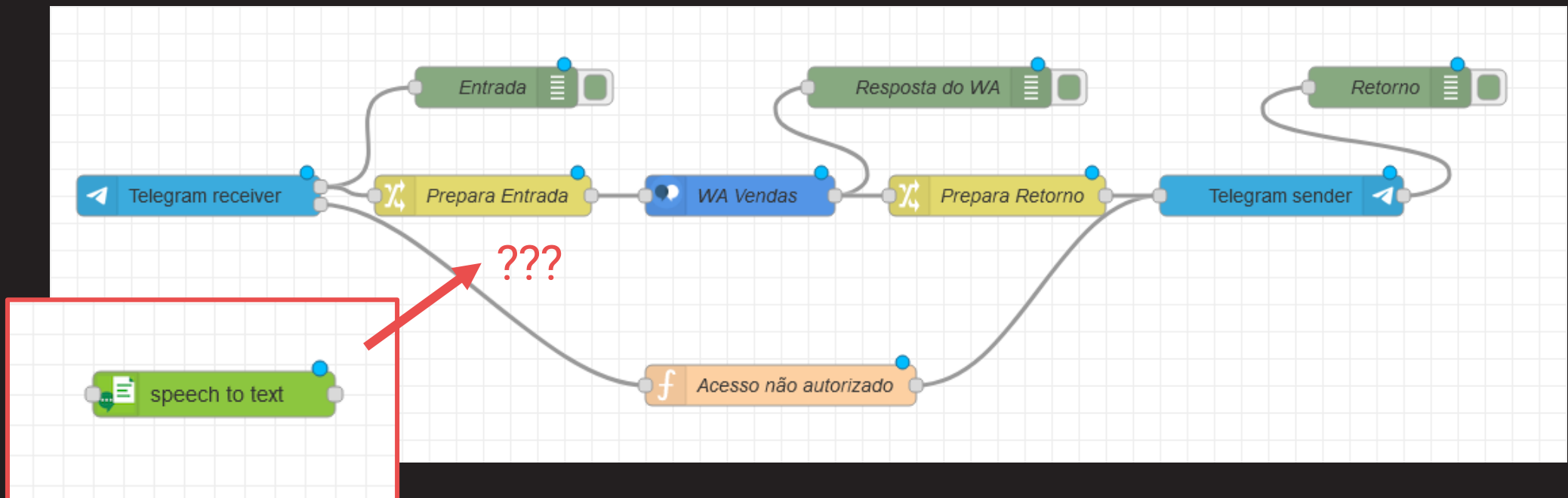


# Descanso

Do professor =D

# Exercícios

1. Carregue o fluxo de integração com o Telegram e o WA. Como você faria para fazer uma integração geral, isto é, entre Telegram, Watson Assistant e Watson STT, assumindo que pode ser recebido áudio e/ou texto?



# Próximos Passos

O que veremos na próxima aula

# Na próxima aula...

- Ensinando o bot a falar

**Copyright © 2021**

**Slides do Prof. Henrique Ferreira, com adaptações dos  
slides dos Prof. Marcelo Grave - FIAP**

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