**UNIVERSIDADE PAULISTA**

**MURILO DE SOUZA RIBEIRO**

**ROBERT DANILO C. CAETANO**

**CAPTAÇÃO E ANÁLISE DE DADOS VOCAIS, FERRAMENTA "ÁSPIDE RECOGNIZER”**

**RIBEIRÃO PRETO**

**2019**

**MURILO SOUZA RIBEIRO**

**ROBERT DANILO C. CAETANO**

**CAPTAÇÃO E ANÁLISE DE DADOS VOCAIS, FERRAMENTA "ÁSPIDE RECOGNIZER”**

Trabalho de conclusão de curso para obtenção do título de graduação em Ciência da Computação apresentado à Universidade Paulista – UNIP

Orientador: Prof. Jean Miller Scatena

**RIBEIRÃO PRETO**

**2019**

**SUMÁRIO**

[1 INTRODUÇÃO 4](#_Toc5036121)

[2 OBJETIVO 5](#_Toc5036122)

[**2.1 Objetivo Geral 5**](#_Toc5036123)

[**2.2 Objetivos específicos 5**](#_Toc5036124)

[3 MATERIAIS E MÉTODOS 6](#_Toc5036125)

[4 REFERENCIAL TEÓRICO 7](#_Toc5036126)

## [5 REFERÊNCIAS BIBLIOGRÁFICAS 8](#_Toc5036127)

# 1 INTRODUÇÃO

Se existe algo mais fundamental do que a escrita não sabemos, mas o que realmente temos ciência é que desde os primórdios esse é o meio indispensável de registro de informações, comunicação, experiências, tratados e etc. A escrita nos agrega valores inafiançáveis, seja por meio da história e paradigmas de pessoas intelectuais que já partiram do plano terrestre, mas que deixaram legados que vêm auxiliando até o tempo contemporâneo.

Sabendo disso a computação vem nos acompanhando paralelamente cobrindo as necessidades da sociedade cada vez mais carentes desde seu processo de evolução natural, consequentemente a décadas atrás era impensável a possibilidade de uma máquina ter a capacidade de reconhecer a voz humana e ser intérprete da sua linguagem, seja ela qual for, de modo a propiciar soluções de problemas e até mesmo, mais conforto e praticidade no cotidiano. Hoje o gelo foi quebrado e a grande maioria já está adaptado com esse processo tecnológico avançado, que acabou deixando de ser novidade.

Ferramentas e recursos a todo momento são desenvolvidas e disparadas para descomplicar a vida da sociedade, focado nesse paradigma o processo de busca por um novo recurso foi pensado e conseguimos trabalhar encima das necessidades de caráter social, com um recurso primordial e fundamental, ou seja, o registro de informações de modo refinado do audível para transcrição, por esse motivo confeccionamos o software responsável pela conversão de arquivos de áudio ou captura direta através de um microfone, para formato textual.

De acordo com Lima (2017) ”Uma equipe que não estabelece uma boa comunicação entre seus membros passa a impressão de estar à deriva, dando oportunidade para o erro e, consequentemente, o retrabalho”. Indubitavelmente mesmo que a tecnologia nos ajude constantemente, um dos maiores problemas dos tempos atuais, ainda é a comunicação, embora pareça simples nos comunicarmos, esse contexto vai além de uma conversa bem organizada e formal, pois a comunicação precisa ser coerente e preferencialmente que proporcione um elo eficaz entre os interlocutores, através do entendimento. E tendo em vista esse problema, propusemos que a ferramenta nos ajude também na análise das comunicações e isso contribui para o nascimento de um futuro projeto, capaz de identificar e analisar possíveis erros na comunicação entre indivíduos, através de suas expressões verbais, falta de clareza e coesão nos dialetos eletrônicos ou pessoais. E nessa aplicação pode ser encontrada a solução desse empecilho, presente grande parte das vezes em nossos relacionamentos a temível incomunicação.

# 2 OBJETIVO

## 2.1 Objetivo Geral

O intuito deste trabalho, de forma geral, é o estudo e desenvolvimento de uma ferramenta capacitada para realizar a conversão de áudio em texto, independentemente do método utilizado na entrada de dados, seja ela por arquivo ou captura de voz.

## 2.2 Objetivos específicos

Dentre os objetivos destacamos:

* Estudar as formas de conversão de áudio em texto;
* Estudar as características e o funcionalidades da biblioteca Speech Recognition;
* Estudar as características e o funcionalidades do CMUSphinx;
* Desenvolver a ferramenta utilizando a linguagem Python;
* Simplificar o processo de registro de informações relevantes.

# 3 MATERIAIS E MÉTODOS

Utilizaremos para a criação desta ferramenta a linguagem de programação Python, já que é uma linguagem amplamente utilizada em aplicações de reconhecimento, seja de texto, imagem ou áudio, por ser uma linguagem de fácil manipulação e escrita.

Python possui uma grande variedade de bibliotecas para a manipulação e compreensão de áudio, nesta ferramenta serão utilizadas as bibliotecas Tkinter, o CMUSphinx.

Tkinter é uma biblioteca nativa do Python que será utilizada para a criação de interface gráfica desta ferramenta.

CMUSphinx é um conjunto d ferramentas dedicadas a compreensão da voz. Ele é composto essencialmente por quatro ferramentas; O Pocketsphinx que é a biblioteca de reconhecimento de voz;

O Sphinxtrain responsável pelo treino do modelo acústico; A Sphinxbase é uma biblioteca de suporte para as demais; e por fim a Sphinx4 que é uma biblioteca de reconhecimento de voz que pode ser modificada, com seu funcionamento voltado para aplicações desktop. No CMUSphinx existem vários modelos de linguagem e ainda permite a criação de novos caso seja necessário.

Nesta aplicação utilizaremos o Pocketsphinx e a Sphinxbase para a interpretação e manipulação do áudio.

# 4 REFERENCIAL TEÓRICO

É preciso ter como base um conhecimento bem sólido referente as técnicas empregadas nas relações entre homem e máquina, tocante com as situações dos dias atuais é tolerável questionar-se a respeito da conversação e compreensão dos computadores, smartphones e até mesmo carros com reconhecimento de fala, será real possíveis tais tecnologias nos proporcionar essa imensa comodidade? É por isso que tendo embasamento em definições conceituais, justificamos e tentamos extrair o mais proveitoso conteúdo dessas tecnologias, junto com seu funcionamento e exemplificações do mesmo.

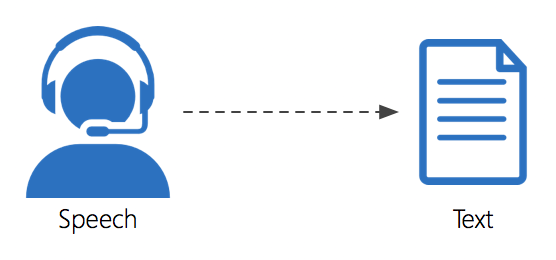
O carro chefe dessa evolução toda e também o nosso alvo!

Imagem 1 – Speech To Text (Speech Recogntion)

|  |
| --- |
| Fonte:<https://www.simplifiedpython.net/speech-recognition-python/> |

O Speech Recognition (SR), também conhecido por Automatic Speech Recognition (ASR), permite fazer a captação da voz do utilizador e transforma a voz em formato texto de forma a poder ser interpretada pelo componente Language Understanding (LU). O reconhecimento da fala pode ser feito pela seguinte forma: a fala é captada e em seguida é dividida em partes com cortes feitos nos intervalos de silêncio. Depois, é feito o reconhecimento do significado de cada parte obtida com base em várias combinações de palavras que são combinadas com o áudio de cada parte. A melhor combinação é a escolhida. (Rato, 2016, p. 10)

Um dos principais recursos para a relação entre a interface homem máquina é o Speech Recognition, por ser uma biblioteca, possui vários módulos integrados capazes de fornecer uma gama de possibilidades aplicáveis no tratamento dos áudios captados, seja por arquivo ou recepção via dispositivo de entrada.

Aliado a área de acessibilidade é prestado auxílio até para pessoas portadoras de deficiência auditiva e visual, tendo essas situações como inspiração, partimos nesse projeto a um estudo aprofundado de seu funcionamento e suas derivações que contém API’s das mais renomadas organizações como por exemplo a Google.

# 5 REFERÊNCIAS BIBLIOGRÁFICAS

Lima, H. (2017). ***A Importância da Comunicação nas Organizações*.** Acesso em 31 de 03 de 2019, disponível em Endomarketing.tv: https://endomarketing.tv/importancia-da-comunicacao-nas-organizacoes/#.XKD315hKiM9

Rato, J. C. (Setembro de 2016). ***Conversação Homem-máquina****.* Acesso em 01 de 04 de 2019, disponível em IC Online: https://iconline.ipleiria.pt/bitstream/10400.8/2375/1/jo%C3%A3o%20Rato-Mestrado%20em%20Eng.Inform%C3%A1tica-Computa%C3%A7%C3%A3o%20M%C3%B3vel.pdf