

Uso de interpretador de voz embarcado em *Raspberry PI* para atendimento de pessoas com deficiência visual em redes de fastfood

1nd Guilherme Rodrigues Costa

Faculdade do Gama of Universidade de Brasília (FGA)
Universidade de Brasília (UnB)
Brasília, Brasil
211039448@aluno.unb.br

2nd Victor Cruz de Oliveira

Faculdade do Gama of Universidade de Brasília (FGA)
Universidade de Brasília (UnB)
Brasília, Brasil
211031842@aluno.unb.br

Abstract—Este projeto apresenta o desenvolvimento de um sistema embarcado utilizando um Raspberry Pi 4, que emprega um interpretador de voz baseado em *Automatic Speech Recognition* (ASR) para o atendimento de clientes por meio da conversa do usuário com o sistema. O objetivo é melhorar a experiência do atendimento eletrônico em redes de fastFood para usuários com deficiência visual ou que tenham dificuldade com tecnologia. Ao integrar o sistema de ASR nas unidades de autoatendimento o usuário seria capaz de trata-las como assistentes virtuais, em que eles realizariam os seus pedidos por meio de uma conversa.

Index Terms—Raspiberry Pi, Assistente virtual, Voz para texto, Automatic speech recognition (ASR)

I. INTRODUÇÃO

Nos últimos anos é possível observar um crescimento na substituição de atendentes por unidades de autoatendimento em diversos setores, como redes de *fast food* e mercados, essa mudança se deve a uma tentativa das empresas de agilizar o atendimento ao cliente aliado a redução de custos para o empregador. Porém essa mudança brusca na forma como essas empresas lidam com clientes trouxe impactos na experiência de atendimento para pessoas com deficiências, em especial, as com deficiências visuais, se tornando necessário opções de acessibilidade para que este nicho de pessoas.

A partir de uma análise de mercado é possível observar que os modelos mais comuns desse equipamento, como os modelos Horazion e Quick Service da schalter [1], são baseados em *displays* interativos que utilizam do estímulo visual. Alguns totens, possuem opções de acessibilidade para pessoas com deficiência visual, como fonte aumentada e audiodescrição, cuja as instruções de acionamento vem escritas em braile na estrutura do totem. Por tanto o mercado do autoatendimento por meio da comunicação verbal com o totem ainda é um nicho a ser explorado no mercado.

Este projeto se propõe a desenvolver um sistema embarcado utilizando um Raspberry Pi 4 de um interpretador de voz baseado em *Automatic Speech Recognition* (ASR), focado no atendimento de pessoas com deficiência visual em redes de fast food. O objetivo é que a partir da comunicação verbal com o totem o usuário seja capaz de realizar os seus pedidos, melhorando assim a experiência do atendimento desse usuário.

A. Automatic speech recognition

ASR é o nome dado as algoritmos que utilizam de *Machine Learning* e Inteligência artificial para reconhecimento dinâmico da fala humana e transcreve-la em forma de texto. Com o passar dos anos essa tecnologia vem crescendo de forma exponencial, vem sendo integrada em diversas aplicações como por exemplo transcrição de áudio de podcasts no spotify, legendas dinâmicas em vídeos no youtube, e assistentes virtuais como a alexa e siri.

Existem diversos métodos de implementação de ASR, tradicionalmente se utiliza sistemas híbridos que integra nos seu processo de decodificação de áudio, o modelo léxico, que utiliza algoritmos para identificar a fonética das palavras, com o modelo de reconhecimento de padrões acústicos, *Acoustic Model* (AM), e com modelos de linguagem, *language model* (LM), que detecta quais são as sequencias de palavras são mais faladas para tentar prever qual a proxima palavra que o usuário ira dizer. Para cada língua é preciso de modelos diferentes devido a suas particularidades de linguagem e fonemas. AO integrar os três modelos simultaneamente os algoritmos ASR são capazes de decodificar dinamicamente os sinais de áudio e transcrever a fala em texto.

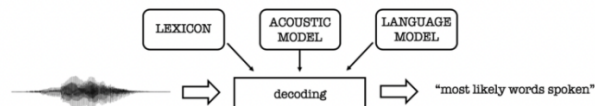


Fig. 1. ASR modelo Híbrido

B. Requisitos

Requisitos Funcionais

- RF0 Captação de áudio
- RF1 Reprodução de áudio
- RF2 Transcrição e interpretação dinâmica da fala humana
- RF3 Realização de pedidos a partir das falas do usuário

Requisitos não Funcionais:

- NF0 Resposta do sistema em tempo real
- NF1 baixo custo computacional

REFERENCES

- [1] Disponível em: https://www.schalter.com.br/totens-de-autoatendimento/?gad_source=1&gclid=CjwKCAjwwr6wBhBcEiwAfMEQs2_Co9UYBy6wAq2_hbRCK6VBqmHk4ny4HMhx1thp40sjc6nbFodo7hoCaS0QAvD_BwE (Acessador em Abril 05, 2024)
- [2] “Inteligência Artificial para Telefonia — PhoneBOT Leucotron,” Leucotron Telecom. Disponível em: <https://www.leucotron.com.br/phonebot-leucotron> (Acessado em Abril 03, 2024).
- [3] Li, Jinyu, et al. Robust Automatic Speech Recognition: A Bridge to Practical Applications. 1st ed. San Diego: Elsevier Science & Technology, 2015. Print.
- [4] Sriram, Anuroop et al. “Robust Speech Recognition Using Generative Adversarial Networks.” 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2018): 5639–5643. Print.
- [5] Gales, Mark. Foundations and Trends in Signal Processing. Vol. 1. Hanover, Mass: Now Publishers, 2008. Print. Foundations and Trends in Signal Processing.