

Assistente pessoal de aprendizado via áudio

Assistente com interface sonora via Raspberry PI

Renato da Costa Motta Junior 09/0130723

Universidade de Brasília -UnB

Brasília -DF, Brasil

renato.motta.jr@gmail.com

Resumo—Este trabalho consiste de um assistente de voz para treino e fixação da tabuada, das quatro operações básicas. O projeto ambiciona ter uma interface inteiramente sonora com usuário. As perguntas são geradas aleatoriamente, mas uma graduação de dificuldade é indicada para futuros trabalhos.

JUSTIFICATIVA

Segundo dados do Saeb 2017, 7 em cada 10 alunos do ensino médio tem conhecimento insuficiente em português e matemática. Ainda segundo esta pesquisa do MEC o percentual negativo só aumenta ao longo da vida escolar.

Por que isso ocorre? Diversos fatores estão envolvidos, mas parece razoável afirmar que quanto mais os estudos se tornam uma obrigação, mais a criança perde o interesse em aprender.

Tendo em vista a crescente necessidade de aprendizagem de crianças, e a crescente utilização de meios tecnológicos entre as crianças desde pequenas como celulares, tablets entre outras.

Portanto o projeto traz como proposta um sistema de assistente virtual que irá auxiliar crianças nas fases iniciais de alfabetização, utilizando um recurso tecnológico para ativar a curiosidade e incentivar o aprendizado.

INTRODUÇÃO

O reconhecimento de fala (*speech recognition*), que não deve ser confundido com o reconhecimento de voz (*voice recognition*), é a capacidade de um sistema reconhecer discursos falados. Uma das principais aplicações dessa tecnologia é o chamado “*speech-to-text*” (ou “discurso para texto” em tradução direta) onde as palavras identificadas são convertidas para texto dentro de um sistema digital, como por exemplo um computador[1]. Alunos da Universidade do Paraná desenvolveram um produto nacional que faz essa conversão para palavras do português brasileiro, o Coruja. O

mesmo é de código aberto, sendo que seu repositório pode ser encontrado online, um dos lugares é o link em anexo[2]. Uma outra solução envolvendo o mesmo laboratório de pesquisa é o SpeechOO, que escreve as palavras identificadas diretamente no Writer do LibreOffice[3], mas esse não será abordado neste trabalho.

O ensino de matemática, sendo considerado essencial em todo o mundo, tem por convenção inicial no Brasil a tabuada dos números naturais até o dez. Também chamada de Tabuada de Pitágoras, a mesma tem origem nas tábuas de contas dos comerciantes da antiguidade[4]. Existem diversas discussões sobre o método de aprendizado da matemática, inclusive da própria tabuada. Mas fato é que o elemento lúdico é muito benéfico, em especial quando se lida com crianças, e é nesse contexto que entra nossa ferramenta.

OBJETIVO

Criar ferramenta que auxilie no aprendizado de crianças em início da fase escolar (primeiros anos do ensino fundamental). A interface por meio de áudio tem por objetivo aumentar o interesse da criança, por tornar a ferramenta mais humanizada e interativa. O protótipo será construído com a utilização de periféricos diretamente em placa construindo assim um protótipo funcional em Raspberry Pi.

BENEFÍCIOS

- Aceleração do aprendizado em crianças;
- Melhoria em atividades para deficientes visuais;
- Inclusão do elemento lúdico no aprender;
- Vencer a resistência da criança ao estudo pela sensação de interação com a aplicação.
- Melhorar o quadro de analfabetismo no país.

REQUISITOS

- Conhecimento de linguagem C;
- Conhecimento de NLP (Processamento de Linguagem Natural);
- API de conversão de fala para texto;
- Aquisição e emissão de áudio;
- A lógica do sistema será implementada em uma RaspBerry Pi e contará com uma lógica que de transformar as strings de saída em áudios (biblioteca eSpeak);

Quantidade	Componentes
1	monitor (testes)
1	microfone USB
1	teclado matricial
1	auto falante
1	fonte 5v

Tabela 1: Componentes

O diagrama de blocos abaixo mostra como a proposta de projeto funcionará primeiramente será captado o sinal de áudio da voz do usuário dando início ao sistema, ao o início do programa o usuário será apresentado a uma interface gráfica onde aparecerá contas de tabuadas em que o usuário irá escolher as opções para resolver as operações. estamos testando a saída de áudio pois está muito baixa, estamos projetando um circuito de amplificador de potência para melhorar a saída de áudio.

O programa irá ficar esperando a resposta via áudio ou em último caso via teclado, a raspberry pi 3 Model B ficará responsável pelo software dedicado a que ficará dedicado a transcrição da fala

em caracteres alfanuméricos para que assim possa ser usado pelo programa de forma a seguir a lógica representada no diagrama abaixo.

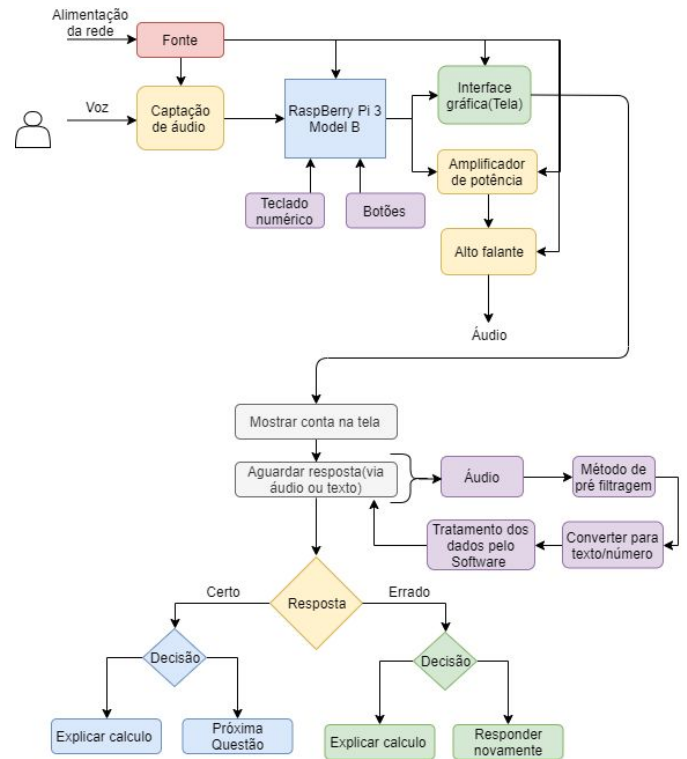


Figura 1: Diagrama de blocos do programa



Figura 2: Microfone USB



Figura 3: Alto falante com conector P2

RESULTADOS

A entrada do programa provém do conversor de fala para texto (Speak-to-Text). A princípio utilizaríamos o Coruja, com o modelo pré-treinado Julius. Porém, essa abordagem foi descontinuada pela dificuldade de interagir com os programas, e de realizar re-treinos com novos *datasets*.

A alternativa selecionada foi um modelo de aprendizado profundo que utiliza realimentação residual. Este modelo próprio, porém, falhou por falta de base de dados, atingindo apenas 20% de acurácia para as 29 classes.

A partir destes infortúnios, e considerando o tempo restante para a entrega do projeto, optamos por validar o mesmo, editando diretamente um documento de texto; emulando o que seria entregue pela API de reconhecimento de fala.

Dentro de suas limitações a validação foi satisfatória, permitindo a observância do funcionamento do programa, em diversos exemplos, sendo todos corretos.

Com isso fica provado que sendo dado um texto correto, ou seja, contendo a resposta do usuário, o programa sempre funciona dentro do esperado, mesmo quando há espúrios ruidosos entre as palavras. Isso só é verdade, porém, quando não há espúrios dentro da própria palavra.

O programa possui um *timeout* que encerra a espera por uma resposta quando 5 segundos da informação da pergunta ao usuário. Figura

No caso de acerto é dita uma parabenização. Em caso de erro há também uma reação seguida da correção, apresentando a resposta correta. Ambos os casos seguem novamente para o menu, onde há a geração de uma nova pergunta a ser pronunciada.

A seguir vemos algumas telas de teste:

```

surnagon@surnagon-Stefanini: ~/Projs/MathSpeaker/obj/x86_64
File Edit View Search Terminal Help
2 X 10
^C
(base) surnagon@surnagon-Stefanini:~/Projs/MathSpeaker/obj/x86_64$ ./mathspeaker

operacao = 2
5 - 3
a criança nao respondeu
operacao = 2
3 - 2
a criança nao respondeu
operacao = 2
9 - 7
a criança nao respondeu
operacao = 4

```

Existem muitas otimizações a serem feitas, além é claro da correção das funções dos periféricos. A inclusão de resto de divisão por exemplo seria uma adição simples que traria benefícios ao reforço do aprendizado da criança. Além de uma gama maior de mensagens, trazendo uma humanização que sem dúvidas cativa as crianças e é um dos pontos fortes deste projeto.

A ideia simples da tabuada utilizada neste projeto pode ser transferida para qualquer tipo de questionário. apesar de isso requer uma lógica muito robusta e ferramentas de inteligência artificial bem mais avançadas.

REFERENCIAS

- [1] Christensson, Per. "Speech Recognition Definition." TechTerms. (January 10, 2014). Disponível em: <https://techterms.com/definition/speech_recognition>. Acesso em: 01 de novembro de 2019.
- [2] Link disponível em: <<http://www.voxforge.org/home/forums/other-languages-forum/portuguese/fala-brasil---speech-recognition-for-the-brazilian-portuguese>>. Acesso em 01 de novembro de 2019.
- [3] BATISTA, P. S. "Avanços em Reconhecimento de Fala para Português Brasileiro e Aplicações: Ditado no LibreOffice e Unidade de Resposta Audível com Asterisk". Campus Universitario do Guamá, UFPA. Belém, Pará. Brasil. 2013.
- [4] "Historia da Tabuada". 5 de novembro de 2012. Disponível em: <<https://umnovojeitodeaprendertabuada.blogspot.com/2012/11/historia-da-tabuada.html>>. Acesso em 01 de novembro de 2019.
- [5] V. Farjado, F. Foreque "7 de cada 10 alunos do ensino médio têm nível insuficiente em português e matemática, diz MEC". Disponível em: <<https://glo.bo/2HB8ZnR>>. Acesso em 30 de setembro de 2019.
- [6] Governo do Estado do PARANÁ: Os desafios da escola pública paranaense na perspectiva do professor PDE. Versão online ISBN 978-85-8015-093-3 CADERNOS PDE
- [7] Amazon Alexa:Disponivel em <https://www.amazon.jobs/pt/business_categories/alexa>, acessado em 30/08/2019
- [8] Instalando IDE Qt Creator e pacotes Qt versão 5. Disponível em <www.filipeflop.com> acessado em 21/10/2019.
- [9] ERICO,Alisson; SHINOHARA, Cindi S.; SARMENTO, Cristiano D.Sistema de reconhecimento de voz para automatização de uma plataforma elevatória. 2014. 97f. Trabalho de Conclusão de Curso (Graduação -Curso de Engenharia Industrial Elétrica Ênfase em Automação).Universidade Tecnológica Federal do Paraná, 2014.