



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
FACULDADE DE ENGENHARIA DA COMPUTAÇÃO E
TELECOMUNICAÇÕES**

**Controle de TV e Servidor LAMP em Sistemas Embarcados:
Uma Solução para Clientes Web e Android com Suporte à
Reconhecimento e Síntese de Voz**

**Belém – Brazil
Abril/2015**

Controle de TV e Servidor LAMP em Sistemas Embarcados: Uma Solução para Clientes Web e Android com Suporte à Reconhecimento e Síntese de Voz

Cassio Trindade Batista
cassio.batista.13@gmail.com
201106840003

Pedro Henrique C. F. Soares
pedrofigueiredoc@gmail.com
201106840007

Gabriel Peixoto de Carvalho
gaburiero.c@gmail.com
201106840010

Thiago Barros Coelho
tbarroscoelho@gmail.com
201106840040

Projeto apresentado à disciplina Projeto de Hardware de Interfaceamento como requisito de avaliação. Professores: Jeferson Leite e Adalbery Castro.

Belem – Brazil
Abril/2015

Sumário

1	Introdução	4
2	Objetivos	4
2.1	Reconhecimento Automático de Voz	4
2.2	Síntese de Voz	4
2.3	Servidor LAMP	4
3	Justificativa	5
4	Revisão Teórica	5
5	Metodologia	5
5.1	Preparação do servidor	5
6	Orçamento	5
7	Dificuldades e Soluções	5

Lista de Figuras

1 Introdução

Tecnologias Assistias

A automação residencial está cada dia facilitando mais a vida dos moradores. Atualmente, boa parte das tarefas do cotidiano são realizadas através de soluções automáticas e práticas, resultando em um ambiente confortável e agradável.

Open source

Servidor próprio de voz, independente de internet

A [1].

2 Objetivos

O objetivo principal consiste em criar um protótipo portátil que seja capaz de controlar um aparelho de televisão através do envio remoto de sinais. Além disso, o sistema será configurado como um servidor que disponibiliza um serviço genérico de reconhecimento de fala, de modo que o aparelho de TV mencionado possa ser remotamente controlado através da voz do usuário; e um serviço de síntese de fala, provendo *feedback* das ações de acordo com o entendimento do sistema ASR.

2.1 Reconhecimento Automático de Voz

Para que o reconhecimento automático de voz seja possível, o *software* Julius deverá ser instalado no servidor. Julius é um software capaz de processar e decodificar áudio em aproximadamente tempo real para tarefas de ditado de até 60 mil palavras.

Para que o Julius possa realizar o reconhecimento em Português Brasileiro, serão necessários basicamente dois recursos: um modelo acústico e um dicionário fonético. Modelos acústicos genéricos para PT_BR podem ser encontrados na página do Grupo FalaBrasil, bem como o software que cria o dicionário fonético (conversor grafema-fonema ou G2P). Entretanto, embora a taxa de acerto dos modelos seja razoavelmente boa, é possível melhorá-la através da criação ou treino de novos modelos.

O processo de treino será realizado pelo software HTK (acrônimo para kit de ferramentas dos modelos ocultos de Markov, livremente traduzido do inglês), o qual é capaz de extrair segmentos de fala de um arquivo de áudio e assinalar uma referência à ele. Tal referência é retirada do dicionário fonético, previamente criado com o software G2P.

2.2 Síntese de Voz

O software eSpeak é a principal referência em síntese de voz em ambientes Linux. Graças à disponibilização de uma API do eSpeak no site oficial do desenvolvedor, o sistema, além de conseguir “ouvir e entender”, também será capaz de “falar”. O download dos modelos para PT_BR é feito automaticamente com o das bibliotecas necessárias.

2.3 Servidor LAMP

```
***PResdro***  
***Predro***  
***Pesro***  
***Porredo***  
***PResdro***
```

3 Justificativa

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), no censo realizado em 2010, aproximadamente 23,9% dos brasileiros declaram ter alguma deficiência [?]. Esse número ainda é mais impressionante quando se pensa que cerca de um quarto de uma população de 190 milhões de habitantes é portadora de alguma necessidade especial. Ainda segundo os dados, 6,9% (13,3 mi) dos brasileiros apresentam algum grau de deficiência motora, enquanto 18,8% (35,7 mi) afirmam serem cegas ou terem alguma dificuldade para enxergar.

Essa pesquisa tem como finalidade apresentar uma solução para diminuir a exclusão digital vivenciada especialmente por pessoas com necessidade motora ou visual, as quais estão à margem da utilização de dispositivos eletrônicos por conta da falta de acessibilidade. A tecnologia de reconhecimento de fala torna acessível a utilização de qualquer dispositivo eletrônico por usuários incapazes de realizar movimentos específicos com membros superiores, como segurar um controle físico e apertar botões ou digitar, por exemplo. Além disso, os portadores de necessidades visuais também poderão ser ajudados, já que nem todos os controles possuem referências reconhecíveis pelo tato.

4 Revisão Teórica

Arduino vs BeagleBoneBlack:

```
**thaigo**  
**thaigo**  
**thaigo**  
**thaigo**  
**thaigo**
```

5 Metodologia

5.1 Preparação do servidor

O servidor, por ser o elemento chave na consolidação do projeto, deve ser o módulo a ser prioritariamente configurado, afim de ser preparado para atender às devidas requisições, bem como executar qualquer tipo de aplicação solicitada. Sendo assim, a instalação da plataforma Angstrom foi tomada como o primeiro passo. Angstrom [?] é um sistema operacional, baseado em Linux, preparado exclusivamente para plataformas embarcadas, tais como o Raspberry Pi ou a própria Beagle Board.

6 Orçamento

Produto	Preço (Total em R\$)
1 × BeagleBone Board Black	200,00
1 × Smartphone com Android	800,00
2 × InfraRed Leds	2,50
Total	1502.50

7 Dificuldades e Soluções

Mano

Referências Bibliográficas

- [1] X. Huang, A. Acero, and H. Hon, *Spoken Language Processing*. Prentice-Hall, 2001.