



**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

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

## Lista de Figuras

# 1 Introdução

A interface homem-máquina encontra-se cada vez mais amigável. O que antes era portado somente por empresas e pessoas com poder financeiro diferenciado e acima da média, em termos de tecnologia, é hoje muito mais simples e agradável para usuários domésticos sem profundo conhecimento no assunto. Diversos estudos vêm sendo desenvolvidos a fim de melhorar ainda mais essa comunicação, de modo que a máquina se aproxime mais de ações como pensar e falar, típicas do ser humano.

Síntese e reconhecimento automático de voz (do inglês *text-to-speech* e *automatic speech recognition*, respectivamente, TTS e ASR) [?, 1] tornam a interface citada acima muito mais prática e natural, de forma que a comunicação de fato se assemelha àquela estabelecida entre duas pessoas. O ASR refere-se ao sistema que, tomando o sinal de fala digitalizado como entrada, é capaz de gerar o texto transcrito na saída. Já um sistema TTS realiza a função contrária, na qual um sinal analógico de voz é sintetizado de acordo com o texto posto na entrada. São inúmeras as aplicações que utilizam tais sistemas envolvendo processamento de voz. Este trabalho tem um foco em uma área especial: automação doméstica com foco em acessibilidade.

De acordo com [?], tecnologias assistivas (TA) “englobam produtos, recursos, metodologias, estratégias, práticas e serviços que objetivam promover a funcionalidade, relacionada à atividade e participação de pessoas com deficiência, incapacidades ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social”.

Servidor próprio de voz, independente de internet

Vale ressaltar que todas as APIs e softwares utilizados para criação dos sistemas e dos recursos utilizados (com exceção do HTK, o qual será visto mais adiante) possuem licença *open source* e são encontrados disponíveis livremente na Internet.

## 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.

- Recursos de Acessibilidade ao computador
  - Equipamentos de entrada e saída (síntese de voz, Braille), auxílios alternativos de acesso (ponteiros de cabeça, de luz), teclados modificados ou alternativos, acionadores, softwares especiais (de reconhecimento de voz, etc.), que permitem as pessoas com deficiência a usarem o computador.
- Sistemas de Controle de Ambiente
  - Sistemas eletrônicos que permitem as pessoas com limitações moto-locomotoras, controlar remotamente aparelhos eletro-eletrônicos, sistemas de segurança, entre outros, localizados em seu quarto, sala, escritório, casa e arredores.
- Auxílios para cegos ou com visão subnormal
  - Auxílios para grupos específicos que inclui lupas e lentes, Braille para equipamentos com síntese de voz, grandes telas de impressão, sistema de TV com aumento para leitura de documentos, publicações etc.

## 4 Revisão Teórica

Arduino vs BeagleBoneBlack vs Raspberry Pi:

**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 (Convertido em R\$)
1 × BeagleBone Board Black	200,00
1 × Smartphone com Android	200,00
2 × InfraRed Leds	2,50
Total	402.50

## 7 Dificuldades e Soluções

Mano

## Referências Bibliográficas

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