

Construção de Aplicativos Usando PLN e Recursos para Reconhecimento e Síntese de Fala Disponibilizados pelo Projeto FalaBrasil da UFPA

Aldebaro Klautau



Grupo FalaBrasil

Laboratório de Processamento de Sinais (LaPS)

Universidade Federal do Pará (UFPA)

<http://www.laps.ufpa.br/falabrasil>

Sumário

Introdução

Definição

Motivação

Aspectos do Reconhecimento automático de voz

Ferramentas

Grupo FalaBrasil

Recursos Disponibilizados

Aplicações Desenvolvidos pelo Grupo FalaBrasil

Aplicações Desenvolvidas por Terceiros Utilizando Recursos do FalaBrasil

Conclusão

Interação com a comunidade

Conclusão

Sumário

Introdução

Definição

Motivação

Aspectos do Reconhecimento automático de voz

Ferramentas

Grupo FalaBrasil

Recursos Disponibilizados

Aplicações Desenvolvidos pelo Grupo FalaBrasil

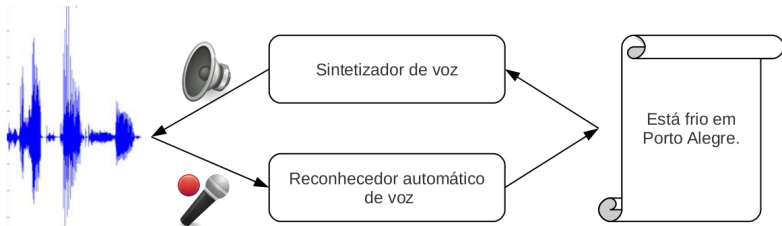
Aplicações Desenvolvidas por Terceiros Utilizando Recursos do FalaBrasil

Conclusão

Interação com a comunidade

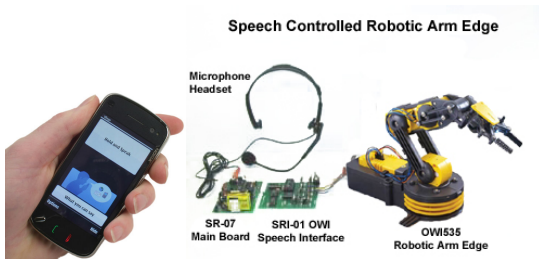
Conclusão

Reconhecimento automático de voz (ASR) e síntese de voz (TTS)

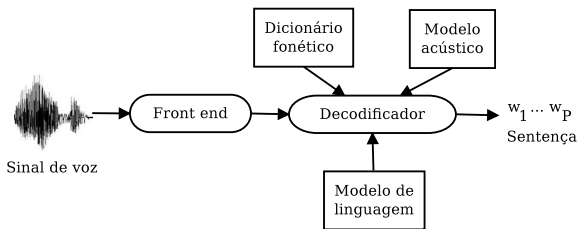


Por que reconhecimento automático de voz?

- Automação
- Acessibilidade
 - Mais de 9 milhões de pessoas com dificuldades físicas, segundo o IBGE.



Reconhecimento automático de voz



- Comando e controle vs Ditado, “Distributed versus desktop”
- *front end*: extrai parâmetros do sinal de voz
- Modelo de linguagem: probabilidade de uma dada sequência de palavras
- Modelo acústico: probabilidade de um sinal de voz ser um determinado fone
- Dicionário fonético: mapeamento entre palavras e fones
- Decodificador une os recursos para reconhecer o som

Ferramentas para o desenvolvimento de ASR

- HTK: treinamento e decodificação de HMM
- SRILM: treinamento de modelos de linguagem
- Decodificadores:
 - Julius
 - HDecode
 - Sphinx-4
 - PocketSphinx
- APIs para o desenvolvimento de aplicativos:
 - Microsoft Speech API
 - Java Speech API
- **Coruja**: Reconhecimento Automático de Voz em Português Brasileiro

Sumário

Introdução

Definição

Motivação

Aspectos do Reconhecimento automático de voz

Ferramentas

Grupo FalaBrasil

Recursos Disponibilizados

Aplicações Desenvolvidos pelo Grupo FalaBrasil

Aplicações Desenvolvidas por Terceiros Utilizando Recursos do FalaBrasil

Conclusão

Interação com a comunidade

Conclusão

Recursos Disponibilizados

- Corpus de áudio livre com mais de 10 horas
- Scripts para treino e adaptação de modelos
- Conversor grafema para fonema (G2P)
- Modelos acústicos e de linguagem
- Engine para reconhecimento: Coruja (decodificador Julius)

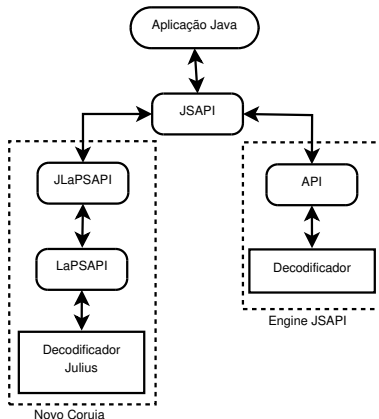
Modelos acústico e de linguagem: LaPSAM e LaPSLM

Decodificador	Modelos independentes		Modelos dependentes	
	CWR(%)	xRT	CWR(%)	xRT
Julius	60.42	0.7	77,7	0.7
HDecode	70.63	0.9	84,6	0.8
IBM ViaVoice	70.71	-	82.7	-

- Testes utilizando o LaPSBenchmark
- Para o teste com modelos adaptados foram utilizados 10 minutos de áudio para cada um dos dois locutores

Engine para reconhecimento de voz: Coruja

- Desenvolvimento de aplicativos com reconhecimento de voz em Português Brasileiro
- Suporte a linguagens .NET, C/C++ e Java
- Implementa parte da JSAPI



Aplicações Desenvolvidos pelo Grupo FalaBrasil

- SpeechOO
- Simon em Português Brasileiro
- Solução para atendimento eletrônico em centrais telefônicas
- FFTranscriber
- Coruja Navigator
- Rotas
- voiceconet
- Dicionário fonético
 - 65 mil palavras transcritas utilizando 38 fones
- mlf2textgrid

Simon em Português Brasileiro

- Automação predial, voltada para teatros e auditórios
- Controle do computador por Voz
- Atualmente com cenários para controle:
 - do sistema operacional
 - do mouse
 - do teclado
 - do navegador Firefox



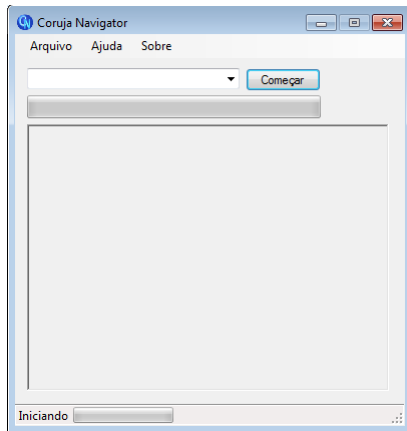
Solução para atendimento eletrônico em centrais telefônicas

- Reconhecimento de voz para atendimento telefônico
- Integração com o software Asterisk

Coruja Navigator

- Navegação Web não visual e sem as mãos
- Funcionamento:
 - Usuário fala um site
 - Software entra no site e então o lê para o usuário destacando os links
 - Usuário pode falar um link para ser acessado
- Atualmente funciona para o site de notícias Folha de São Paulo

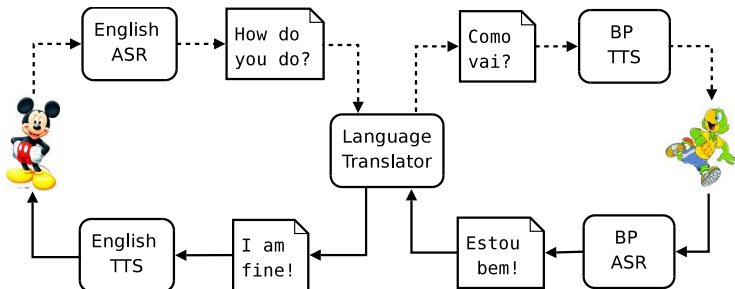
Coruja Navigator



Speech-to-Speech Machine Translation

- Permite diálogo entre falantes nativos do PB e do Inglês
- Funcionamento:
 - Usuário dita uma frase em PB
 - O sistema reconhece e traduz a fala para o Inglês
 - O texto gerado é então passado como entrada para um TTS em Inglês
 - O processo inverso acontece de forma similar
- A versão atual possui limitações, mas é operacional

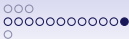
Speech-to-Speech Machine Translation





voiceconet

- Adaptação do modelo acústico do FalaBrasil a voz do usuário
- Funcionamento
 - Usuário acessa a página
`www.laps.ufpa.br/falabrasil/voiceconet.html`
 - Grava pelo menos 3 minutos de áudio
 - Adapta o modelo a sua voz
 - O modelo adaptado é disponibilizado para download pronto para funcionar com o decodificador Julius



mfl2textgrid

- Alinha áudio e texto e converte para TextGrid do Praat

The image displays two windows from the mfl2textgrid application. The left window, titled '2. TextGrid 989', shows a spectrogram of the audio file 'ministério'. The spectrogram has a vertical axis for frequency (0 Hz to 5000 Hz) and a horizontal axis for time (0.177978 to 142.668760 seconds). A blue dashed line tracks a formant across the spectrogram, with a label '75 Hz' at the end. Below the spectrogram, the text 'ministério' is displayed with vertical bars under each letter. The right window, titled 'mfl2textgrid', is a control panel. It contains several input fields with 'Procurar...' buttons: 'Arquivo TXT:' (E:/989.txt), 'Arquivo WAV:' (E:/989.wav), 'Salvar LAT em:' (E:/989.lat), 'Arquivo LAT:' (E:/989.lat), and 'Salvar TextGrid em:' (E:/989.TextGrid). There are two green progress bars, both at 100%. Buttons include 'Criar arquivo LAT' and 'Converter arquivo'.

Aplicações Desenvolvidas por Terceiros Utilizando Recursos do FalaBrasil

- Automação residencial (agente doméstico)
- Leitura de RSS controlada por voz, para utilizar ao dirigir
- Web service para reconhecimento de voz
- Controle de avatar por voz
- Indexação de vídeos e audios
- Transcrição de entrevistas
- Controle de robôs
- Assistente de direção
- Contabilização de jogos de basquete
- Contabilização de corrida de Kart
- Controle de Thunderbird e Outlook

Sumário

Introdução

Definição

Motivação

Aspectos do Reconhecimento automático de voz

Ferramentas

Grupo FalaBrasil

Recursos Disponibilizados

Aplicações Desenvolvidos pelo Grupo FalaBrasil

Aplicações Desenvolvidas por Terceiros Utilizando Recursos do FalaBrasil

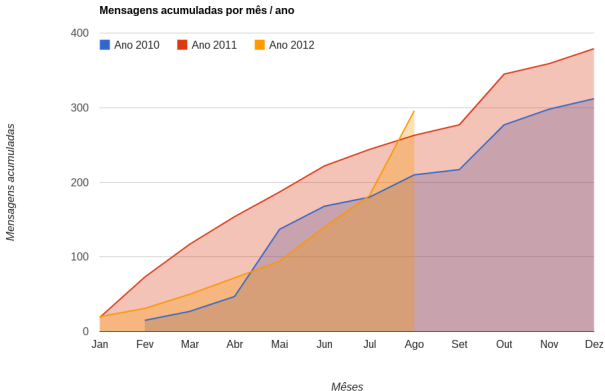
Conclusão

Interação com a comunidade

Conclusão

Grupo de discussão

- Mais de 250 membros
- Mensagens acumuladas



Conclusão

- Disponibilização de recursos tanto para PB como para outras línguas
- Todos recursos e produtos disponibilizados gratuitamente na página do Grupo FalaBrasil
<http://www.laps.ufpa.br/falabrasil>
- Sentimento: observar uso comercial do que se repassou gratuitamente
- Qual será o impacto do software livre em NLP?
- Pesquisa em modelos de linguagem discriminativos

Obrigado!

Contato: a.klautau@ieee.org

