

*Listener:*

## Um Reconhecedor de Pronúncia para Falantes do Português Brasileiro Aprendizes de Inglês

Projeto de Mestrado apresentado em 26 de junho de 2013,  
nos Seminários do NILC (ICMC/USP), em São Carlos.

**Gustavo Augusto de Mendonça Almeida (USP)**

[gustavoauma@gmail.com](mailto:gustavoauma@gmail.com)

**Orientadora: Profa. Dra. Sandra Maria Aluisio (USP)**

[sandra@icmc.usp.br](mailto:sandra@icmc.usp.br)

**Co-orientador: Prof. Dr. Aldebaro Klautau Jr. (UFPA)**

[aldebaro.klautau@gmail.com](mailto:aldebaro.klautau@gmail.com)

# 0. Overview



**Seção 1:** Introdução

**Seção 2:** Fundamentação Teórica

2.1: Aquisição de L2

2.2: Reconhecimento de Fala

2.3: Linguística de *Corpus*

**Seção 3:** Metodologia

3.1: Levantamento dos erros de pronúncia

3.2: Modelagem do reconhecedor

**Seção 4:** Avaliação

**Seção 5:** Considerações Finais

**Seção 6:** Referências bibliográficas

# 1. Introdução

**QUAL O NÍVEL DE CONHECIMENTO DE INGLÊS  
DOS BRASILEIROS?**



# 1. Introdução



Em uma pesquisa realizada pela *Global English* (2013), envolvendo 137.000 informantes, sobre o conhecimento de **inglês em empresas**, o Brasil ocupou a **71ª** posição em um ranking de **77 países**.

1		PHILIPPINES	7.95
2		NORWAY	7.06
3		NETHERLANDS	7.03
4		UNITED KINGDOM	6.81
5		AUSTRALIA	6.78
6		BELGIUM	6.45
7		FINLAND	6.39
8		SWEDEN	6.33
...			
69		VENEZUELA	3.39
70		TURKEY	3.30
71		BRAZIL	3.27
72		EL SALVADOR	3.24
73		CHILE	3.24

**Figura 1.** Ranking parcial da *Global English* (2013).

# 1. Introdução



O desempenho dos brasileiros correspondeu ao nível *beginner*, que constitui a pior das faixas consideradas pela pesquisa.

Essa faixa delimita indivíduos com conhecimento de inglês iniciante, com capacidades comunicativas bastante limitadas.

(GLOBAL ENGLISH, 2013)

BRAZIL

3.27

## BEGINNER

Can read and communicate using only simple questions and statements, but can't communicate and understand basic business information during phone calls.

## BASIC

Can understand business presentations and communication descriptions of problems and solutions, but can only take a minimal role in business discussions and the execution of complex tasks.

## INTERMEDIATE

Can take an active role in business discussions and perform relatively complex tasks.

## ADVANCED

Can communicate and collaborate much like a native English speaker.



Figura 2. Faixas de desempenho consideradas pela *Global English* (2013).

# 1. Introdução



No Índice de Proficiência em Inglês, estabelecido pela agência *Education First* (EF), o Brasil, em 2012, foi classificado na 46ª posição de 54 países, sendo agrupado na faixa de países com **proficiência muito baixa em inglês**.

**Proficiência muito baixa**

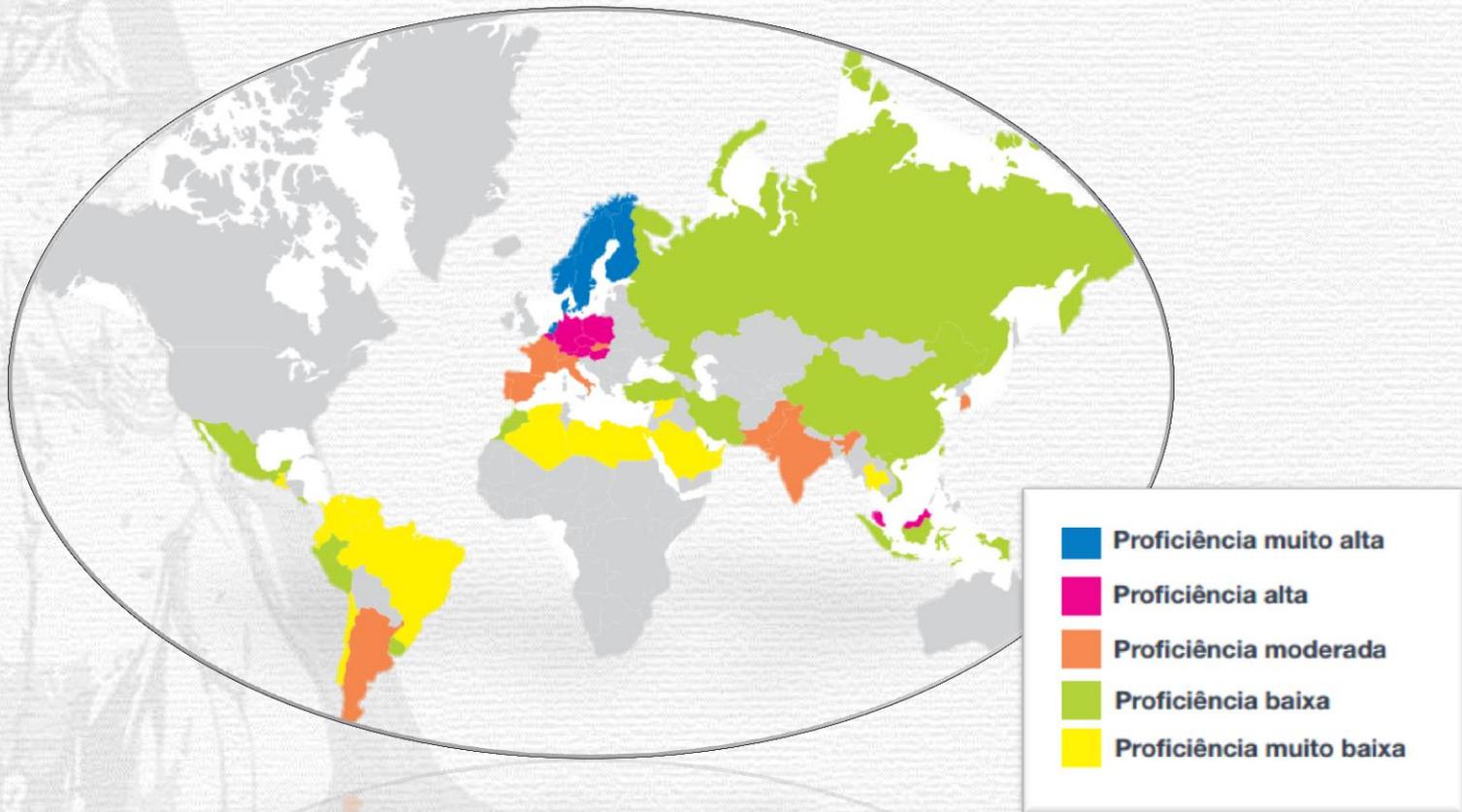
Classificação	País	EF EPI
39	Chile	48.41
40	Venezuela	47.50
41	El Salvador	47.31
42	Síria	47.22
43	Equador	47.19
44	Argélia	47.13
45	Kuwait	47.01
46	Brasil	46.86
47	Guatemala	46.66
48	Egito	45.92
49	Emirados Árabes Unidos	45.53
50	Colômbia	45.07
51	Panamá	44.68
52	Arábia Saudita	44.60
53	Tailândia	44.36
54	Libia	42.53

**Figura 3.** Ranking de países com proficiência muito baixa.

# 1. Introdução



## ÍNDICE DE PROFICIÊNCIA EM INGLÊS – EDUCATION FIRST (2012)



**Figura 4.** Mapa de Proficiência em Inglês.

(EDUCATION FIRST, 2012)

# 1. Introdução



Este projeto busca trazer contribuições para a melhoria desses índices. **O objetivo é desenvolver um reconhecedor de pronúncia para falantes do português brasileiro (PB) aprendizes de inglês, chamado *Listener*, que seja capaz de fornecer ao usuário *feedback*, em tempo real, sobre a qualidade de sua pronúncia.**

Recursos semelhantes já foram desenvolvidos para outras línguas, como o japonês (TSUBOTA et al., 2004), o espanhol (REIS & HAZAN, 2011), o holandês (STRIK et al., 2008; NERI et al., 2003) e o francês (GENEVALOGIC, 2006).

No entanto, **para o PB, há ainda uma lacuna** a ser explorada.

# 1. Introdução



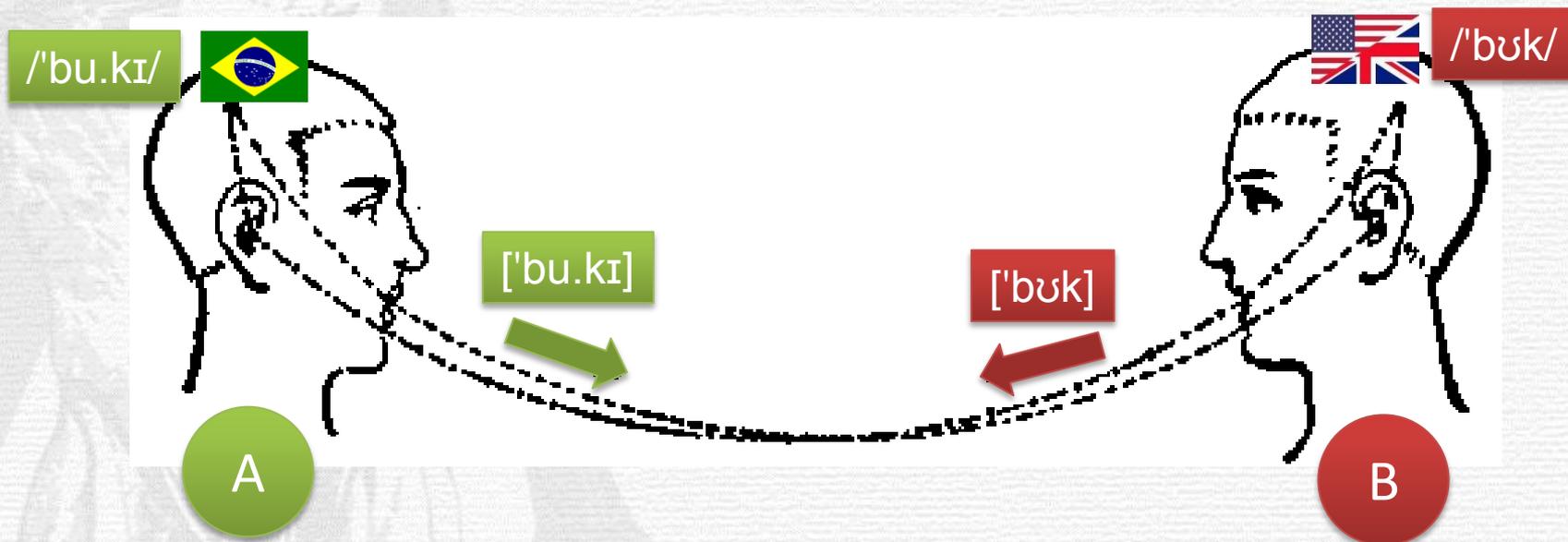
A hipótese de pesquisa é que é possível **construir um reconhecedor de fala eficiente para analisar a pronúncia de inglês de falantes nativos do PB,** através de:

- (i) uma **classificação de erros** de pronúncia que leve em conta a transferência de padrões de L1 para L2;
- (ii) um **modelo acústico** que agregue dados de fala do inglês tanto de nativos, quanto de aprendizes;
- (iii) um **dicionário de pronúncia** que contenha a transcrição das pronúncias desviantes do aprendiz;
- (iv) um **modelo de língua** que condiga com a sintaxe do aprendiz.

A eficiência do *Listener* será verificada a partir de medidas tradicionais para **avaliação intrínseca** de reconhecedores de fala.

## 2.1. Aquisição de Segunda Língua (ASL)

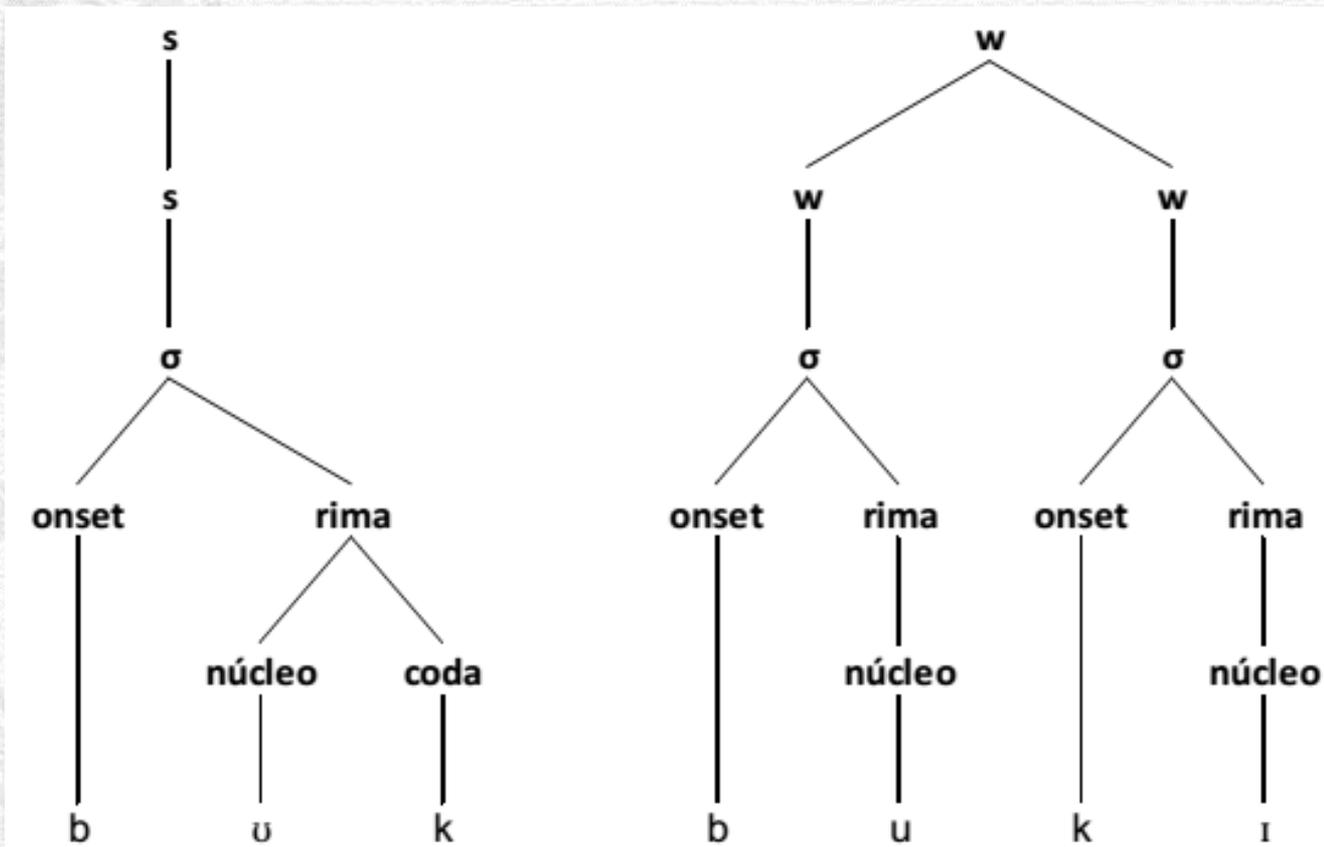
“ Quando nos deparamos com uma língua estrangeira, a tendência natural é que interpretemos seus sons a partir dos sons de nossa própria língua. Analogamente, quando falamos uma língua estrangeira, tendemos a utilizar os sons e os padrões sonoros de nossa língua nativa na realização. (WELLS, 2000)



**Figura 5.** Esquema do processo de comunicação.

(SAUSSURE, 1916; WELLS, 2000)

## 2.1. Aquisição de Segunda Língua (ASL)



**Figura 6.** Realização da palavra 'book' na pronúncia padrão do inglês (esq.) e com transferência do PB para o inglês (dir.) – Representação autossegmental segundo Selkirk (1982).

## 2.1. Aquisição de Segunda Língua (ASL)

No que diz respeito à **pronúncia** de uma língua estrangeira, há, principalmente, a **transferência** de **padrões do sistema fonológico** da L1 para a L2 e, também, a transferência de **padrões de correspondência entre letra e som** da L1 para a L2.



Forma Ortográfica	Forma Fonética <i>AmE</i>	Forma Fonética <i>PB &gt;&gt; AmE</i>	
<i>time</i>	[ˈtʰaɪm]	[ˈtaɪ.mɪ]	
<i>him</i>	[ˈhɪm]	[ˈhĩ]	*[ˈhĩ.mɪ]
<i>nice</i>	[ˈnaɪs]	[ˈnaɪ.sɪ]	
<i>mass</i>	[ˈmæs]	[ˈmɛs]	*[ˈmɛ.sɪ]

**Quadro 1.** Exemplo de influência de padrões de escrita na fala do aprendiz.

## 2.1. Aquisição de Segunda Língua (ASL)



	FONES CONSONANTAIS DO PORTUGUÊS BRASILEIRO													
	Bilabial		Labiodental	Dental	Alveolar		Alveopalatal	Palatal	Velar	Glotal				
Oclusiva	p	b			t	d			k	g				
Africada					tʃ	dʒ								
Nasal		m				n		ɲ						
Vibrante														
Tepe						r								
Fricativa			f	v		s	z	ʃ	ʒ		x	ɣ	h	ɦ
Aproximante									j		w			
Lateral						l			ʎ					

	FONES CONSONANTAIS DO INGLÊS												
	Bilabial		Labiodental	Dental	Alveolar		Alveopalatal	Palatal	Velar	Glotal			
Oclusiva	p	b			t	d			k	g			
Africada					tʃ	dʒ							
Nasal		m				n			ŋ				
Vibrante													
Tepe													
Fricativa			f	v	θ	ð	s	z	ʃ	ʒ			h
Aproximante							r		j		w		
Lateral							l						

**Quadro 2.** Inventário fonético consonantal do PB e do AmE.

(CRISTÓFARO-SILVA, 2005; OGDEN, 2012)

## 2.1. Aquisição de Segunda Língua (ASL)



VOGAIS DO PORTUGUÊS BRASILEIRO						
	Anterior		Central		Posterior	
	Não-arr.	Arr.	Não-arr.	Arr.	Não-arr.	Arr.
<b>Alta</b>	i ɪ ĩ					u ʊ ũ
<b>Média-alta</b>	e ě					o õ
<b>Média-baixa</b>	ɛ					ɔ
<b>Baixa</b>			a ə ã			

VOGAIS DO INGLÊS AMERICANO						
	Anterior		Central		Posterior	
	Não-arr.	Arr.	Não-arr.	Arr.	Não-arr.	Arr.
<b>Alta</b>	i: ɪ					u: ʊ
<b>Média-alta</b>						
<b>Média</b>			ə			
<b>Média-baixa</b>	ɛ		ɜ:		ʌ	ɔ:
<b>Baixa</b>	æ		ɑ:			

**Quadro 3.** Inventário fonético vocálico do PB e o AmE.

(CRISTÓFARO-SILVA, 2005; OGDEN, 2012)

## 2.1. Aquisição de Segunda Língua (ASL)

É **extensa a literatura** existente para o ensino da pronúncia do inglês, em suas diversas variantes:

- Halliday (1970);
- Jones (1976);
- O'Connor (1980);
- Clifford (1985);
- Kreidler (1989);
- Ladefoged (1993);
- Dalton & Seidlhofer (1994);
- Gilbert (2000);
- Kenworthy (2000);
- Staun (2010);
- Ogden (2012);
- *etc.*



**A GRANDE MAIORIA DOS  
MANUAIS DE ENSINO DE  
PRONÚNCIA DE INGLÊS  
DESCONSIDERA A LÍNGUA  
NATIVA DO APRENDIZ NO  
PROCESSO DE ASL.**

(HALLIDAY, 1970; JONES, 1976; O'CONNOR, 1980; CLIFFORD, 1985; KREIDLER, 1989; LADEFOGED, 1993; DALTON & SEIDLHOFER, 1994; GILBERT, 2000; KENWORTHY, 2000; STAUN, 2010; OGDEN, 2012)

## 2.1. Aquisição de Segunda Língua (ASL)

Na classificação dos erros do *Listener*, utilizaremos como base os trabalhos de **Godoy et al. (2006)**, **Zimmer et al. (2009)** e **Cristófar-Silva (2012)**.

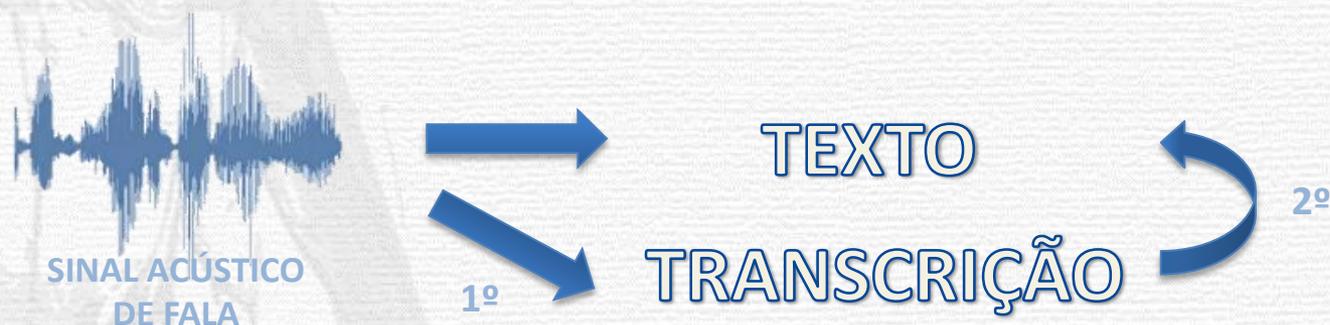
Tais trabalhos constituem o que há de mais recente no ensino de pronúncia de inglês para falantes do PB, por estabelecerem uma metodologia de ensino de pronúncia que considera aspectos de transferência de L1 para L2.

Em outras palavras, tais trabalhos propõem um método de ensino de pronúncia direcionado ao falante de PB aprendiz de inglês.

## 2.2. Reconhecimento de pronúncia

Um **reconhecedor de pronúncia** nada mais é do que um **reconhecedor de fala voltado a uma tarefa específica**, qual seja: compreender e analisar a pronúncia de um aprendiz.

“ O propósito de um reconhecedor de fala é transformar, de forma eficiente e precisa, o sinal acústico da fala em sua contraparte textual. (RABINER & SCHAFER, 2007)



**Figura 7.** Esquema simplificado de um sistema de reconhecimento de fala.

## 2.2. Reconhecimento de pronúncia



**O PROBLEMA:** Se o estado da arte em reconhecimento de fala já apresenta nível razoável de confusão para dados de fala de nativos, como reconhecer (e processar) a fala de não-nativos?

Há diversas formas de se possibilitar um eficiente reconhecimento automático de fala de não-nativos, por exemplo, através da utilização de:

- (i) adaptação ao falante;
- (ii) interpolação acústica;
- (iii) modelos bilíngues;
- (iv) modelos combinados, ou de interlíngua.

TÉCNICA ESCOLHIDA PARA  
ESTE PROJETO

## 2.2. Reconhecimento de pronúncia



### Julius

*Julius* é um motor de reconhecimento de fala de alta performance e de código aberto para a construção de sistemas de reconhecimento de fala. Ele incorpora grande parte das técnicas do estado da arte em reconhecimento de fala e executa Reconhecimento de Fala Contínuo com Grande Vocabulário (LVCSR).

O motor de reconhecimento vem sendo desenvolvido há mais de 20 anos por diversos laboratórios do Japão:

- Kawahara Lab., Kyoto University (1991-2013); 
- Information-technology Promotion Agency, Japan (1997-2000);  独立行政法人 情報処理推進機構  
Information-technology Promotion Agency, Japan
- Shikano Lab., Nara Institute of Science and Technology (2000-2005);  Shikano Lab.  
Speech and Acoustics Processing Laboratory
- Julius project team, Nagoya Institute of Technology (2005-2013).

(LEE & KAWAHARA, 2009; LEE et al., 2001; KAWAHARA et al., 2000)

## 2.2. Reconhecimento de pronúncia

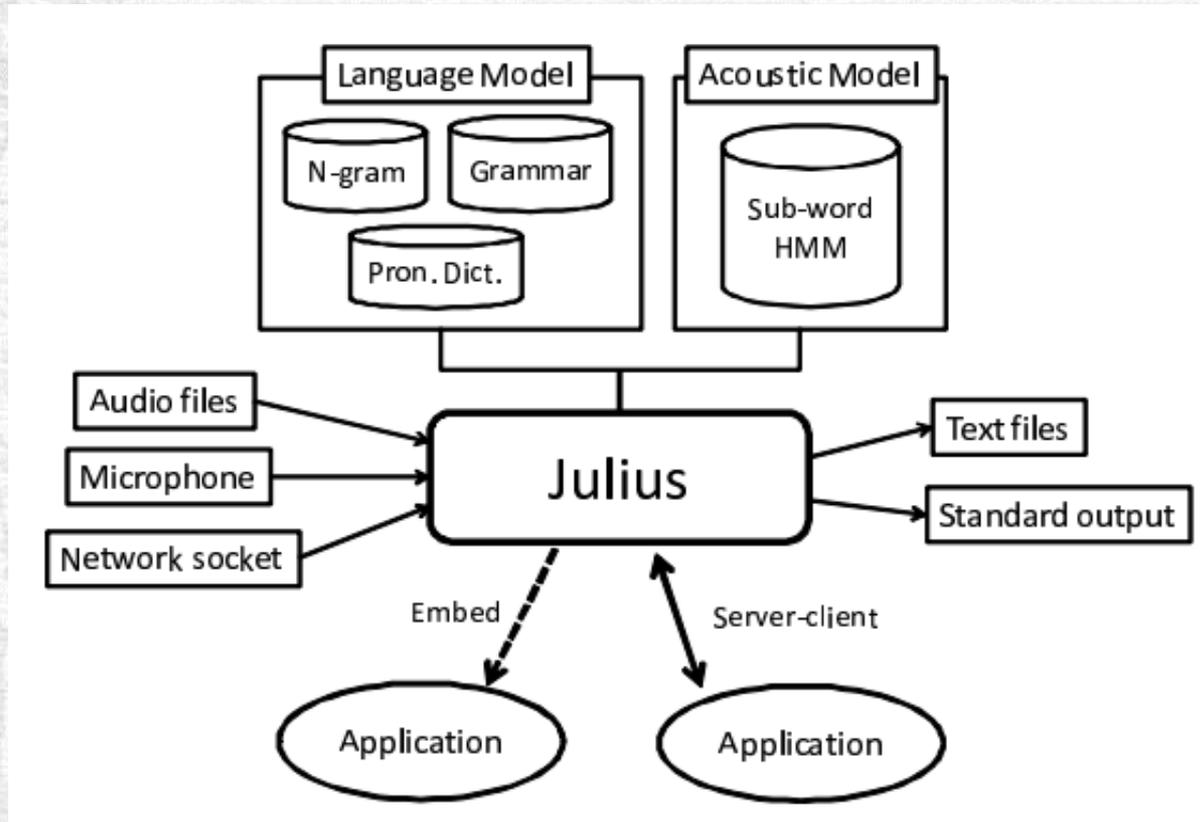


### Características:

- open-source;
- reconhecimento em tempo real, de alta precisão, via estratégias *2-pass*;
- suporte a *LMs* na forma de N-gramas, regras e palavras isoladas;
- suporte a *LMs* em formato *ARPA* e *AMs* em *HTK ASCII* e *HMMDefs*.
- altamente configurável: pode-se modificar diversos parâmetros de busca e alternar-se o algoritmo de decodificação (*1-best/word-pair approx.*, *word trellis/word graph intermediates*, etc.);
- reconhecimento *on-the-fly* via microfone ou *stream* de internet;
- rejeição de entrada via *GMM*;
- saídas do tipo N-best, grafo de palavras, val. de confiança e matrizes de confusão;
- alinhamento forçado em palavras, fonemas e janelas de áudio;
- modelo de *LM* configurável pelo usuário;
- reconhecimento otimizado para palavras isoladas;
- *etc.*

(LEE & KAWAHARA, 2009; LEE et al., 2001; KAWAHARA et al., 2000)

## 2.2. Reconhecimento de pronúncia



**Figura 8.** Arquitetura do motor de reconhecimento de fala *Julius*.

## 2.3. Linguística de *Corpus*

A Linguística de Corpus é **um método de investigação linguística**, de base empirista, que propõe o estudo da língua a partir de **exemplos reais de uso**.

“

Na linguística, um *corpus* é uma coleção de textos (um “corpo” da língua) armazenado em um banco de dados eletrônico. Comumente, *corpora* são grandes coleções de textos legíveis, em formato legível por computadores, os quais contêm milhares ou milhões de palavras (BAKER et al. 2006).

Tipos de *corpora* (KENNEDY, 1998):

- **gerais/de referência** vs. **especializados**;
- **históricos** vs. **da língua atual**;
- **regionais** vs. **multidialectais**;
- **de aprendizes** vs. **de nativos**;
- **multilíngues** vs. **monolíngues**;
- **falado** vs. **escrito** vs. **transcrito**.

## 3.1. Levantamento dos erros de pronúncia



Com base nos trabalhos de **Godoy et al. (2006)**, **Zimmer et al. (2009)** e **Cristófaró-Silva (2012)**. Dez tipos de erros de pronúncia foram propostos para o reconhecedor de pronúncia.

---

### DESVIOS DE PRONÚNCIA A SER ANALISADOS PELO *LISTENER*

---

1. epêntese de [i, ɪ]
  2. vocalização da lateral alveolar vozeada [l] em final de sílaba
  3. apagamento de consoantes nasais em final de sílaba
  5. articulação da retroflexa alveolar vozeada [r]
  6. articulação da consoante nasal velar [ŋ]
  7. articulação das consoantes fricativas dentais [θ] e [ð]
  8. aspiração em oclusivas em início de sílaba
  9. pronúncia de verbos regulares na 3ª pessoa do singular do presente do indicativo
  10. pronúncia de verbos regulares no simple past ou past participle
- 

**Quadro 4.** Desvios de pronúncia a ser analisados pelo *Listener*.

### 3.1. Levantamento dos erros de pronúncia

FORMA ORTOGRÁFICA	FORMA FONÉTICA <i>AmE</i>	FORMA FONÉTICA <i>PB&gt;&gt;AmE</i>
<i>time</i>	['tʰaɪm]	['taɪ.mɪ]
<i>dog</i>	['dɔːg]	['dɔ̃.gɪ]
<i>book</i>	['bʊk]	['bu.kɪ]
<i>start</i>	['staːrt]	[is.'tʰaɪ.tʃɪ]
<i>sprite</i>	['sprɑɪt]	[is.'praɪ.tʃɪ]

**Quadro 5.** Epêntese de [i, ɪ].

FORMA ORTOGRÁFICA	FORMA FONÉTICA <i>AmE</i>	FORMA FONÉTICA <i>PB&gt;&gt;AmE</i>
<i>pool</i>	['pʰuːl]	['puʊ]
<i>all</i>	['ɑːl]	['ɔ̃]
<i>cold</i>	['kʰoʊld]	['kou.dʒɪ]
<i>pile</i>	['pʰaɪl]	['paɪ.ou]
<i>myself</i>	[maɪ'self]	[maɪ'seʊ.fi]

**Quadro 6.** Vocalização da lateral alveolar vozeada [l] em final de sílaba.

(GODOY et al., 2006; ZIMMER et al., 2009; CRISTÓFARO-SILVA, 2012)

### 3.1. Levantamento dos erros de pronúncia

FORMA ORTOGRÁFICA	FORMA FONÉTICA <i>AmE</i>	FORMA FONÉTICA <i>PB&gt;&gt;AmE</i>
<i>him</i>	['hɪm]	['hĩ]
<i>soon</i>	['su:n]	['sũ]
<i>moon</i>	['mu:n]	['mũ]
<i>on</i>	['ɑ:n]	['õ]
<i>then</i>	['ðen]	['dẽ]

**Quadro 7.** Apagamento de consoantes nasais em final de sílaba.

FORMA ORTOGRÁFICA	FORMA FONÉTICA <i>AmE</i>	FORMA FONÉTICA <i>PB&gt;&gt;AmE</i>
<i>art</i>	['ɑ:ɹt]	['aɹt]
<i>car</i>	['kɑ:ɹ]	['kaɹ]
<i>rent</i>	['ɹent]	['hẽ.tɹ]
<i>fruit</i>	['fru:ɹt]	['fruɪ.tɹ]
<i>fry</i>	['fɹaɪ]	['fraɪ]

**Quadro 8.** Articulação da retroflexa alveolar vozeada [ɹ].

(GODOY et al., 2006; ZIMMER et al., 2009; CRISTÓFARO-SILVA, 2012)

### 3.1. Levantamento dos erros de pronúncia

FORMA ORTOGRÁFICA	FORMA FONÉTICA <i>AmE</i>	FORMA FONÉTICA <i>PB&gt;&gt;AmE</i>
<i>king</i>	['kɪŋ]	['kĩ.gɪ]
<i>reading</i>	['ri:.dɪŋ]	['ri.dĩ]
<i>writing</i>	['raɪ.tʃɪŋ]	['raɪ.tʃĩ]
<i>singer</i>	['sɪŋ.ə]	['sĩ.gə]
<i>finger</i>	['fɪŋ.gə]	['fĩ.gə]

**Quadro 9.** Articulação da consoante nasal velar [ŋ].

FORMA ORTOGRÁFICA	FORMA FONÉTICA <i>AmE</i>	FORMA FONÉTICA <i>PB&gt;&gt;AmE</i>
<i>thank</i>	['θæŋk]	['fě.kɪ]
<i>thought</i>	['θɑ:t]	['tɔ.tʃɪ]
<i>fought</i>	['fɑ:t]	['fɔ.tʃɪ]
<i>then</i>	['ðen]	['dě]
<i>this</i>	['ðɪs]	['dis]

**Quadro 10.** Articulação das consoantes fricativas dentais [θ] e [ð].

(GODOY et al., 2006; ZIMMER et al., 2009; CRISTÓFARO-SILVA, 2012)

## 3.1. Levantamento dos erros de pronúncia

FORMA ORTOGRÁFICA	FORMA FONÉTICA AmE	FORMA FONÉTICA PB>>AmE
<i>think</i>	['θɪŋk]	['fĩ.kɪ]
<i>thinks</i>	['θɪŋks]	['fĩ.kɪs]
<i>kiss</i>	['kɪs]	['kɪs]
<i>kisses</i>	['kɪs.ɪz]	['ki.sɪd]
<i>fly</i>	['flaɪ]	['flaɪ]
<i>flies</i>	['flaɪz]	['flaɪs]
<i>live</i>	['lɪv]	['li.vɪ]
<i>lives</i>	['lɪvz]	['li.vɪd]
<i>witch</i>	['wɪtʃ]	['wɪtʃ]
<i>witches</i>	['wɪtʃ.ɪz]	['wi.tʃɪs]
<i>bus</i>	['bʌs]	['bʌs]
<i>buses</i>	['bʌs.ɪz]	['bʌ.sɪs]

**Quadro 11.** Articulação das consoantes fricativas dentais [θ] e [ð].

...

## 3.2. Modelagem do reconhecedor



### ABORDAGEM DE INTERLÍNGUA

A **abordagem de interlíngua** propõe um design de reconhecedor que **busca adequar-se à fala de aprendizes**.

“

De início, alimenta-se o modelo acústico com dados de fala tanto de nativos, quanto de não-nativos. Em seguida, reformula-se o dicionário de pronúncia para conter, além da pronúncia padrão, as pronúncias desviantes que os aprendizes produzirão (NERI et al., 2003).

Adicionalmente, propomos a utilização de um **modelo de língua que seja mais condizente** com as estruturas sintáticas produzidas pelo aprendiz.

## 3.2. Modelagem do reconhecedor

### MODELO ACÚSTICO (AM)

- O modelo acústico proposto será elaborado via *HMM* (através do *HTK Hidden Markov Model Toolkit*)
- Definido para **trifones** (um trifone **corresponde a um fone da forma  $L-F+R$** , em que  $F$  é o fone em questão,  $L$  o à esquerda e  $R$  o à direita);
- O modelo acústico **será compilado a partir de dois corpora de fala**, um de falante nativos do inglês: *TIMIT Acoustic-Phonetic Continuous Speech Corpus*, outro de falantes nativos do PB, aprendizes de inglês como L2: *COBAI – Corpus Oral Brasileiro de Aprendizes de Inglês*.

## 3.2. Modelagem do reconhecedor



### MODELO ACÚSTICO (AM)

O corpus **TIMIT foi elaborado**, conjuntamente, pelo Instituto de Tecnologia de Massachusetts (MIT), SRI Internacional e Texas Instruments Inc. (TI) **com o propósito fornecer dados para a realização de estudos de fonética acústica do inglês**, bem como para **o desenvolvimento de sistemas automáticos de reconhecimento de fala**.

#### Características:

- Contém gravações de cerca de **630 falantes**;
- Cobre os **8 dialetos majoritários** do inglês americano;
- Compõe-se de leituras de **10 sentenças artificiais**, de **contexto foneticamente rico**;
- Está **transcrito ortográfica e foneticamente**;
- Possui **alinhamento temporal** entre o áudio e a transcrição;
- Foi **gravado em cabine acústica** a 16kHz com 16 bits por amostra.



## 3.2. Modelagem do reconhecedor

### DICIONÁRIO DE PRONÚNCIA

O dicionário de pronúncia a ser utilizado integra o *VoxForge Speech Corpus*.

#### Características:

- *Open source* - o VoxForge é um repositório gratuito e *open source* de recursos para reconhecimento de fala;
- Baseado no *CMU Pronouncing Dictionary*;
- Constituído por **131.411 palavras** do inglês americano transcritas foneticamente em formato *ARPAbet*.

Ao dicionário serão **acrescidas as hipóteses de pronúncia dos aprendizes**, por meio de **regras transformacionais** e (quiçá!) **aprendizado de máquina**.

## 3.2. Modelagem do reconhecedor

### DICIONÁRIO DE PRONÚNCIA

ID DA PALAVRA	TRANSCRIÇÃO	FONÉTICA (ARPAbet)
	[dʒiza'pih]	
DISANO	[dʒiza'piɹ]	d ix s aa n ow
DISANTI	[dʒiza'piɹ]	d ix s ae n t iy
DISANTIS	[diza'piɹ]	d ix s aa n t ix s
DISANTO	[dʒiza'piɹ]	d ix s ae n t ow
DISAPPEAR	[dʒisa'piɹ]	d ih s ax p ih r
DISAPPEAR(2)	[dʒisa'piɹ]	d ih s ax p iy r
DISAPPEARANCE	[dizə'piɹ]	d ih s ax p ih r ax n s
DISAPPEARANCE(2)	[dʒisa'piɹ]	d ih s ax p iy r ax n s
DISAPPEARANCES	[disə'piɹ]	d ih s ax p ih r ax n s ix z
DISAPPEARANCES(2)	[disə'piɹ]	d ih s ax p iy r ax n s ix z

**Quadro 1.** Exemplo de entradas no dicionário de pronúncia do *VoxForge Speech Corpus*, com adição de pronúncias dos aprendizes.

## 3.2. Modelagem do reconhecedor



### MODELO DE LÍNGUA (LM)

Há diversos modelos de língua para o inglês:

- *Gigaword*;
- *CSR LM1*;
- *HUB4*;
- *etc.*;
- bem como diversos *corpora* disponíveis para treino de modelos.

Todavia, a grande maioria desses modelos foi gerada a partir de *corpora* de artigos de jornal e é sabido que textos jornalísticos *tendem a possuir estrutura sintática e vocabulário complexos* (CANNING, 2002; CANNING & TAIT, 1999).

## 3.2. Modelagem do reconhecedor

### MODELO DE LÍNGUA (LM)

Propomos a utilização da *Simple English Wikipedia* como base para treino do Modelo de Língua (LM) de trigramas a ser utilizado no *Listener*.



Articles in the Simple English Wikipedia use fewer words and easier grammar than the English Wikipedia.[...] Some examples of people who use Simple English Wikipedia: students, children, adults who might find it hard to learn or read and people who are learning English. – *Simple English Wikipedia*



### Características:

- **Inglês simplificado** (vocabulário mais simples, sentenças curtas, estruturas sintáticas menos complexas, etc.);
- **101.299 artigos** (v. de 24 de junho 2013);
- **Crowd-sourcing** (artigos oriundos de múltiplas fontes).

# 4. Avaliação

A **eficiência** do reconhecedor de pronúncia será **mensurada de modo intrínseco/in vitro**, através das medidas:

- **Word Error Rate (WER)**

$$WER = \frac{S + D + I}{N}$$

- **Character Error Rate (WER)**

$$CER = \frac{C}{N}$$

- **Matrizes de confusão de fonos e palavras**

	$\phi_1$	$\phi_2$	$\phi_3$	...	$\phi_n$
$\phi_1$	$n_{REC}(\phi_1, \phi_1)$	$n_{REC}(\phi_1, \phi_2)$	$n_{REC}(\phi_1, \phi_3)$		$n_{REC}(\phi_1, \phi_n)$
$\phi_2$	$n_{REC}(\phi_2, \phi_1)$	$n_{REC}(\phi_2, \phi_2)$	$n_{REC}(\phi_2, \phi_3)$		$n_{REC}(\phi_2, \phi_n)$
$\phi_3$	$n_{REC}(\phi_3, \phi_1)$	$n_{REC}(\phi_3, \phi_2)$	$n_{REC}(\phi_3, \phi_3)$		$n_{REC}(\phi_3, \phi_n)$
...					
$\phi_n$	$n_{REC}(\phi_n, \phi_1)$	$n_{REC}(\phi_n, \phi_2)$	$n_{REC}(\phi_n, \phi_3)$		$n_{REC}(\phi_n, \phi_n)$

Tais medidas serão analisadas **a partir dos dados de ambos os corpora empregados**, por meio de **cross validation**.

# 5. Considerações Finais

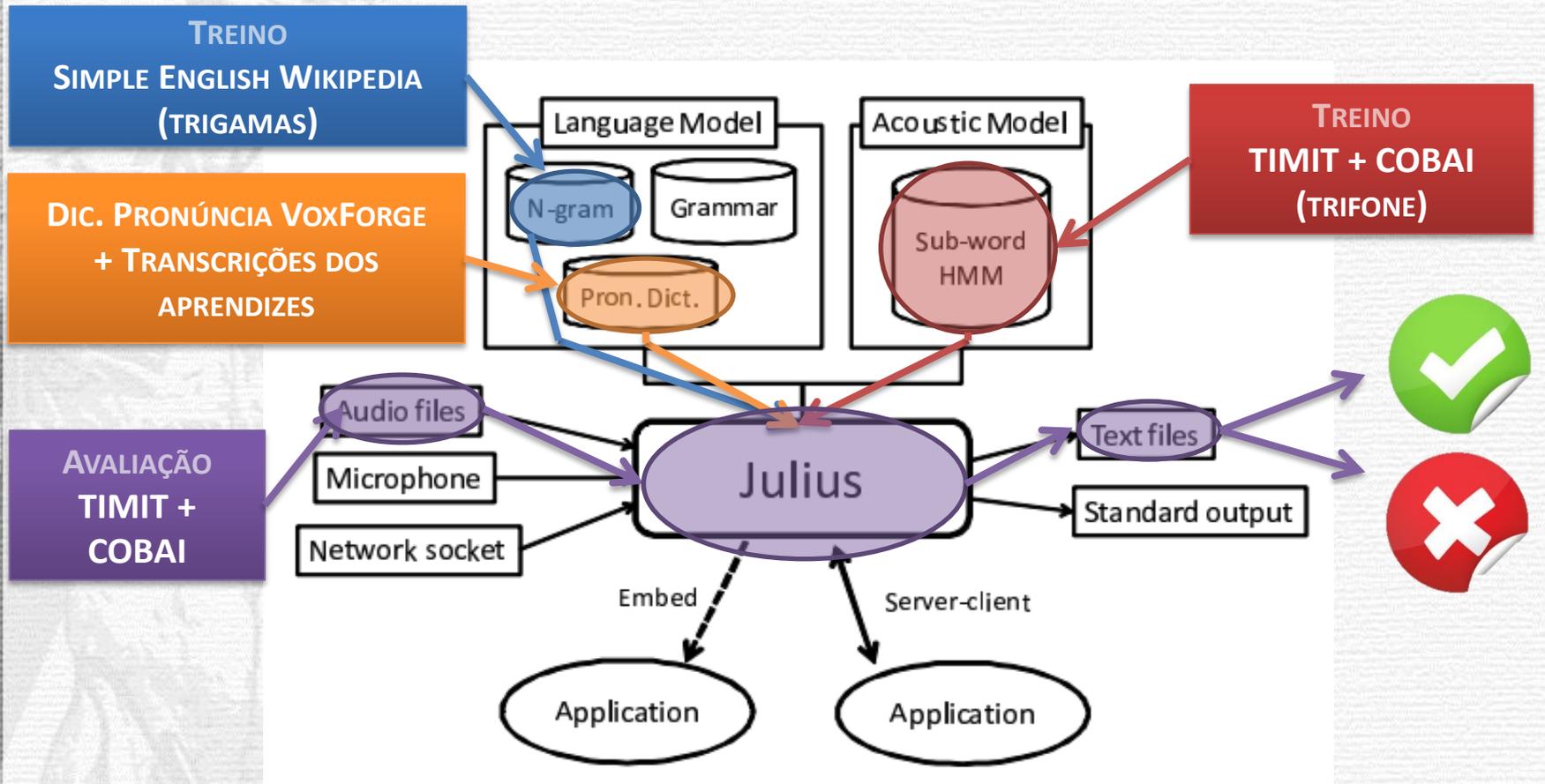


Figura 9. Esquema de desenvolvimento do *Listener*.



**Obrigado!**

**Gustavo Augusto de Mendonça Almeida (USP)**

[gustavoauma@gmail.com](mailto:gustavoauma@gmail.com)

**Orientadora: Profa. Dra. Sandra Maria Aluisio (USP)**

[sandra@icmc.usp.br](mailto:sandra@icmc.usp.br)

**Co-orientador: Prof. Dr. Aldebaro Klautau Jr. (UFPA)**

[aldebaro.klautau@gmail.com](mailto:aldebaro.klautau@gmail.com)