

Geração de prosódia para o português brasileiro em sistemas text-to-speech

Felipe Cortez de Sá

Universidade Federal do Rio Grande do Norte

Junho de 2018

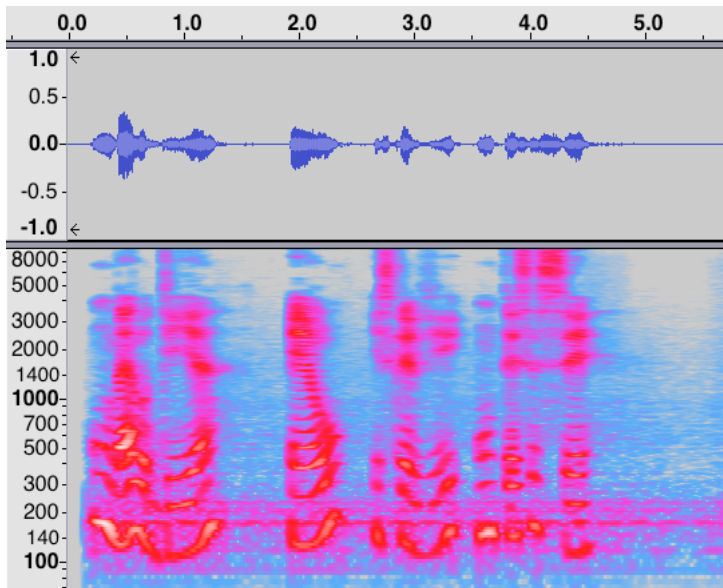
- 1 Introdução
- 2 Prosódia
- 3 Fundamentação teórica
- 4 Implementação
- 5 Perguntas

It would be a considerable invention indeed, that of a machine able to mimic speech, with its sounds and articulations. I think it is not impossible.” (Leonhard Euler, 1761)

- *Voice User Interfaces*
 - *Apple - Siri*
 - *Google Assistant*
 - *Microsoft - Cortana*
 - *Amazon - Alexa*
- Acessibilidade
- Ensino de linguagens
- Estudo de linguística
- Prosódia afetiva

- pros (verso) - odé (canto)
- Suprasegmental
- Frequência
- Duração
- Intensidade

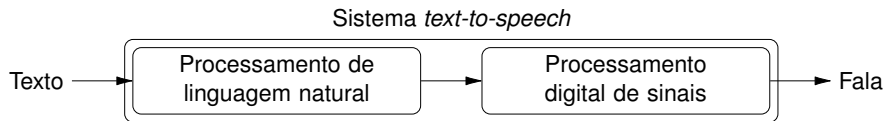
Prosódia



Função prosódica

- Suprasegmental
- Afetiva
- Aumentativa

Sistemas text-to-speech



■ Front end

- Normalização de texto
- Conversão grafema-fone
- Geração de prosódia

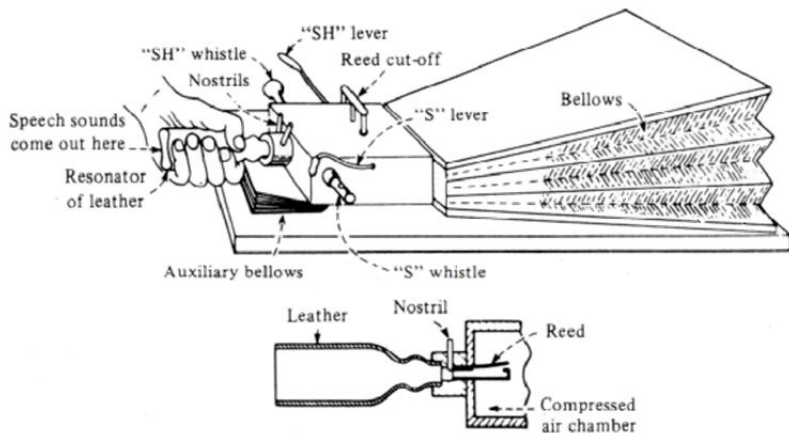
■ Back end

- Síntese articulatória
- Síntese por formantes
- Síntese concatenativa
- Síntese por Hidden Markov Models e Deep Neural Networks

O desafio da geração de prosódia

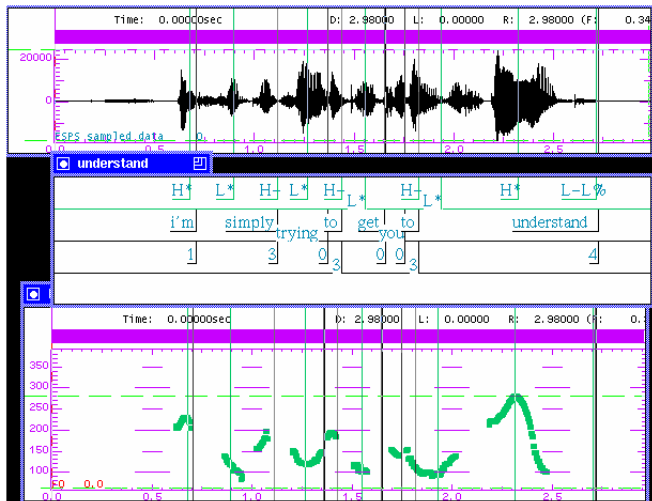
- Elemento extra-textual

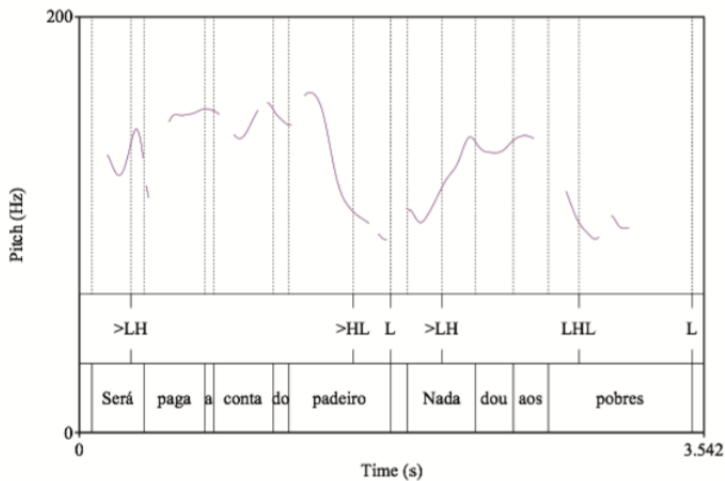
Máquina de Wolfgang von Kempelen (1769)



- Aiuruetê (1997)
- espeak-ng
- Couto et al (MaryTTS)
- LianeTTS

- ToBI – Tone Breaks and Indices
- DaTo – Dynamic Tones
- INTSINT – International Transcription System for Intonation





Ele FOI lá HOje?

[↑ > ↑↑ ↓]

- SSML – Speech Synthesis Markup Language
- EmotionML
- Anotações manuais

```
< speak >
```

```
  Siga < emphasis level = " strong " > aquele < / emphasis > carro.
```

```
< / speak >
```

```
<emotionml version="1.0" xmlns="http://www.w3.org/2009/10/emotionml">  
  <emotion category-set="http://www.w3.org/TR/emotion-voc/xml#everyday-categories">  
    <emotion>  
      <category name="happy" />  
      Que bom te ver!  
    </emotion>  
  </emotionml>
```

(Utterance Words

(The

(boy ((accent L*))

saw

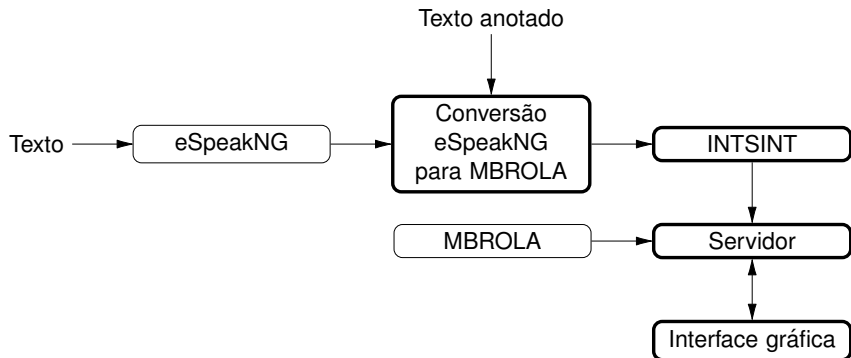
the

(girl ((accent H*) (tone L-)))

with

the

(telescope ((accent H*) (tone H-H%))))))



Regra	Cálculo
Top	$\text{key} \times \sqrt{2^{range}}$
Middle	key
Bottom	$\text{key} / \sqrt{2^{range}}$
Higher	$\sqrt{P_{i-1} \times T}$
Same	P_{i-1}
Lower	$\sqrt{P_{i-1} \times B}$
Upstepped	$\sqrt{P_{i-1} \times \sqrt{P_{i-1} \times T}}$
Downstepped	$\sqrt{P_{i-1} \times \sqrt{P_{i-1} \times B}}$

_ 150 50 150
t 70 50 125
e 125 50 75
c 70 50 125
e 125 50 75
c 70 50 125
e 116 20 232 80 300
_ 150 50 150

Demonstração!

- Adicionar suporte a outros modelos de análise entoacional
- Usar *Natural Language Understanding* para estimar prosódia
- Gerar prosódia a partir de marcação SSML
- Criação de corpus anotado com prosódia para o português brasileiro

*If you understood everything I said, you'd be
me* (Miles Davis)