

Manual de uso do sistema com Inteligência Artificial para avaliação de fluência em leitura

Apresentação

O sistema desenvolvido é composto dos módulos seguintes:

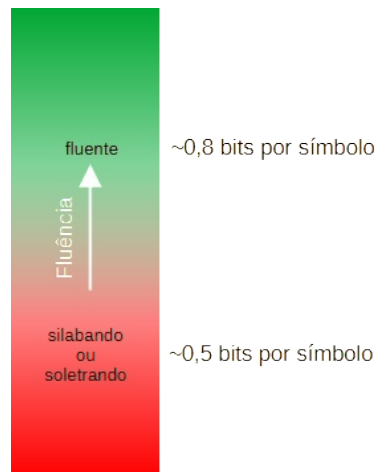
- a) Leitura e filtragem dos arquivos no formato PCM wav: Nesse módulo, é feita uma primeira análise do perfil de potência do sinal, o que permite, entre outras, a eliminação de intervalos muito longos de silêncio por parte do orador, e a segmentação da oração em intervalos de fala e de silêncio.
- b) Análise espectral dos segmentos de sinal de fala: para cada segmento de sinal de 100 ms, a avanços de 10 ms, é convertido 3 filtros de predição linear que são usados para a construção de 3 perfis espectrais consecutivos de 50 frequências cada, que são codificados como um único vetor de 150 dimensões normalizado.
- c) Reconhecimento de fonemas em Português brasileiro: os vetores espectrais normalizados, representando segmentos curtos de fala, são sistematicamente processados por uma primeira Rede Neural Artificial (há duas, no sistema todo) para a rotulação do segmento de fala com um fonema do Português Brasileiro. Esses fonemas são encadeados numa longa sequência de símbolos (frase).
- d) Estimação de taxa de bits por símbolo: A frase de símbolos é analisada através da medida de complexidade de Lempel-Ziv para uma estimativa de quantos bits de informação (Informação de Shannon) há, efetivamente, em cada símbolo/fonema emitido pelo orador. Espera-se que um leitor fluente seja capaz de emitir mais símbolos novos (inovação) por duração de fala que um leitor iniciante, o que torna a taxa de bits por símbolo uma medida fundamental deste sistema.
- e) Estimação de relação média entre *bursts* de fala e intervalos de silêncio: usando o perfil de potência do sinal tratado em (a), pares consecutivos de silêncios e falas contínuas são comparados como um marcador relevante da fluência do orador.
- f) Estimador de quantidade de letras do alfabeto pronunciadas: para melhor distinguir leituras silabadas de soletradas, esse último módulo detecta e conta letras do alfabeto soletradas isoladamente.

Medidas relevantes

Na tarefa de classificação da fluência dos leitores registrados em áudio, duas medidas se mostraram bons descritores, a saber:

- a) a taxa de bits por símbolo (T1) e
- b) a taxa de letras por segmentos sonoros (T2)

A primeira, T1, como ilustrado na figura seguinte, é maior para leitores fluentes, pois isso reflete a taxa de novos fonemas proferidos por segmentos sonoros. Essa taxa foi estimada através do método de Lempel-Ziv, que computa a complexidade de uma sequência de observações simbólicas.

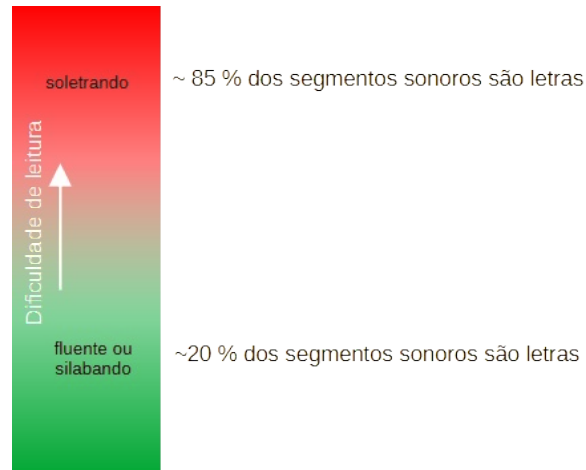


Escala de fluência em bits por símbolo (fonema).

No entanto, essa taxa, sozinha, não é capaz de diferenciar leituras do tipo soletrada de leituras silabadas, na medida em que cada letra soletrada se confunde com sílabas pronunciadas. Por exemplo, as letras “B”, “C”, “V” etc. são pronunciadas como bê, cê e vê, que se confundem com sílabas em uma leitura do tipo silabada.

Para diferenciar a classe soletrada da classe silabada, o sistema incorporou um detector de sons de letras do alfabeto brasileiro, que provê a medida T2, que mede a proporção de segmentos sonoros isolados reconhecidos como letras pronunciadas.

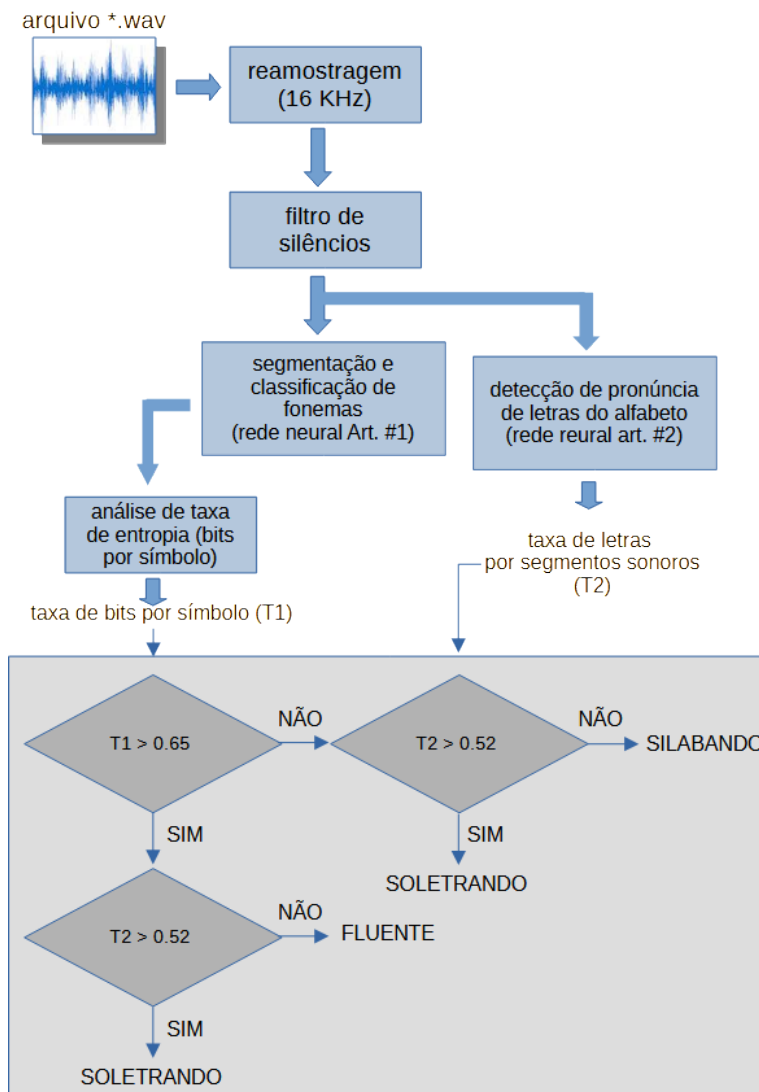
De acordo com os arquivos de áudio disponibilizados para execução do projeto do sistema, essa taxa se distribui entre as classes de acordo com a figura seguinte.



Escala de dificuldade de leitura em percentual de letras pronunciadas por segmentos sonoros.

Apresentação geral da solução desenvolvida

Nenhuma das duas medidas, T1 ou T2, é suficiente para uma classificação satisfatória das 3 classes-alvos, pois a primeira confunde as classes soletrando e silabando, ao passo que a segunda confunde fluente com silabando. Assim, um esquema de classificação foi concebido e ajustado como descrito na figura seguinte.



Esquema geral do sistema

Instruções de uso da rotina computacional em Python

Instruções de uso da rotina computacional em Julia

A implementação em linguagem Julia é composta pela rotina **tutorVirtual.jl** e pelos arquivos de parâmetros seguintes:

"W1.csv"
 "B1.csv"
 "W2.csv"
 "B2.csv"
 "W1Letras.csv"
 "B1Letras.csv"
 "W2Letras.csv"
 "B2Letras.csv"
 "ConjuntoFonemas.csv"

A rotina **tutorVirtual.jl** deve ser executada em computador com Julia Versão 1.7.1 instalada (disponível em <https://julialang.org/>), na mesma pasta em que se encontram os arquivos de parâmetros descritos acima.

Os arquivos de áudio a serem processados/analísados devem ser colocados em uma pasta-alvo. Por exemplo, supondo que a pasta-alvo onde se encontram os áudios (arquivos wav) é a C:\dados\audios, então a chamada ao **tutorVirtual.jl** deve ser:

```
C:\> julia tutorVirtual.jl "C:/dados/audios"
```

Importante: Nos argumentos da chamada, atenção ao uso do “/” em lugar do “\”. Isto é, usa-se: "C:/dados/audios" em lugar de "C:\dados\audios"

Na mesma pasta onde estão os arquivos wav, arquivos com os mesmos nomes, mas com extensão “.rel” serão gerados. Esses arquivos são, na verdade, arquivos do tipo “txt” contendo, em ASCII, os relatórios de avaliações de cada arquivo de áudio correspondente. Por exemplo, para um dado áudio “amostra#5.wav,” será gerado, na mesma pasta, o arquivo amostra#5.rel, cujo conteúdo é ilustrado abaixo.



Precauções

Os programas descritos neste manual foram ajustados para processar arquivos de áudio em formato PCM com extensão “wav” sem compressão, e com registros de vozes humanas em tarefa de leitura de textos. Outros formatos e/ou tipos de conteúdo de áudio não devem ser usados, pois as medidas apresentadas não se aplicam, portanto serão medidas espúrias, desprovidas de significados uteis.

Níveis de ruído de fundo muito elevados também podem tornar não confiáveis as medidas e classificações dos áudios, portanto recomenda-se o cuidado da realização dos registros em ambientes silenciosos. Ademais, no caso de registros do tipo soletrando, recomenda-se que as letras soletradas sejam intercaladas por pausas claras, para tornar a avaliação automática mais segura.