



Conversor Grafema-Fonema e Alinhador Forçado para Português Brasileiro

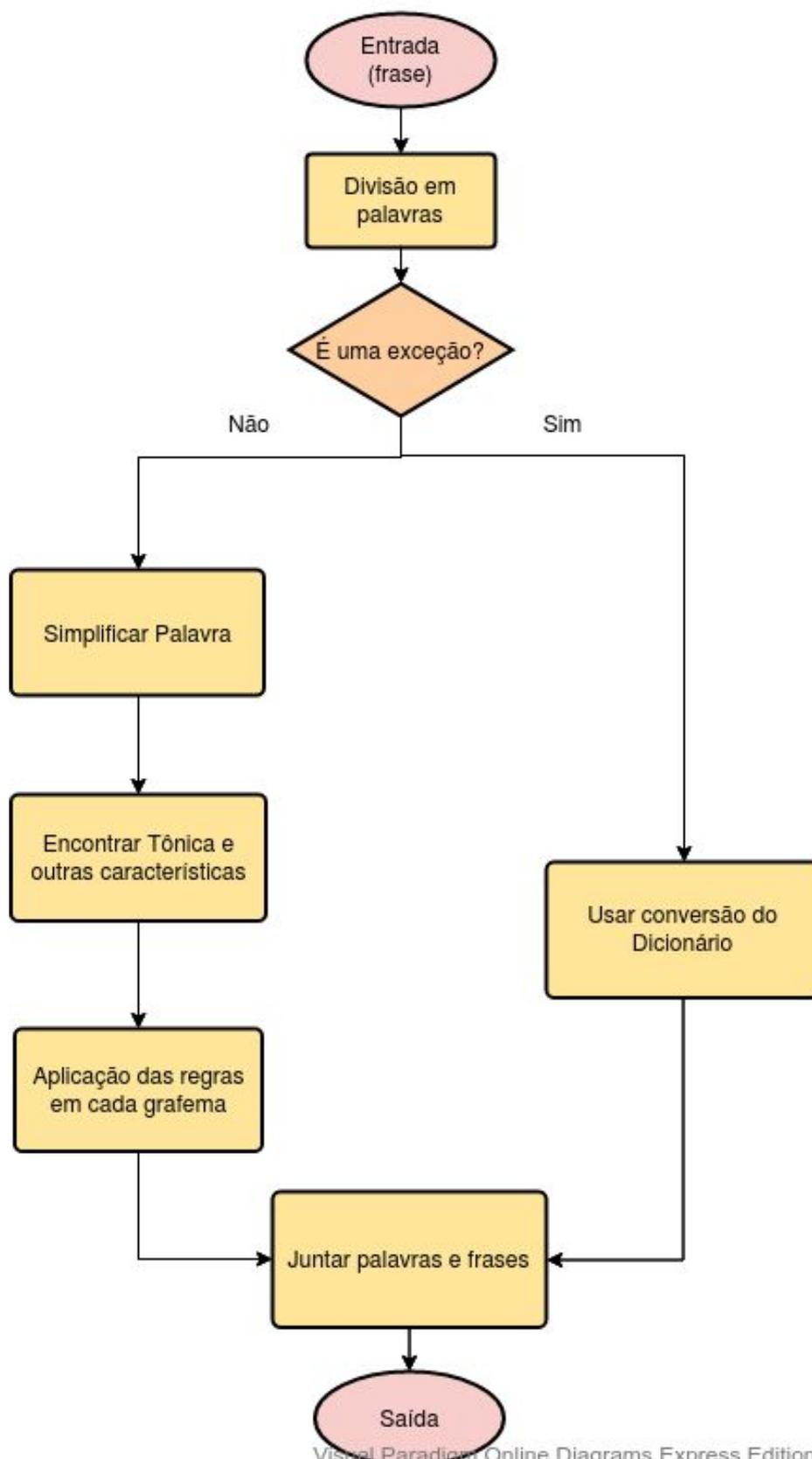
Um alinhador forçado é uma ferramenta que recebe como entrada um áudio com sua transcrição ortográfica e retorna o seu alinhamento, ou seja, os intervalos nos quais cada fonema (ou ainda sílaba fonética ou palavra) é pronunciado no áudio. Isso é uma ferramenta muito útil tanto para o estudo fonético quanto para outras tarefas como a criação de datasets de treinamento para modelos de closed captioning, por exemplo, acelerando tarefas muito laboriosas que teriam de ser feitas manualmente. Modelos que realizam a tarefa de alinhamento forçado já foram treinados para diversas línguas, porém, até o momento, nenhum havia sido feito de forma satisfatória para o português brasileiro, de modo que esse foi o objetivo desse projeto.

O alinhador forçado pode ser dividido em duas componentes principais: o Conversor Grafema-Fonema e o Alinhador. O primeiro é responsável por receber o texto de entrada em sua representação em grafemas e convertê-lo para seu equivalente fonético, enquanto que o alinhador é responsável pelo alinhamento fonético de fonemas, sílabas ou palavras obtidos.

O Conversor-Grafema foi implementado usando regras fonéticas que foram obtidas a partir de uma análise de frequência de casos. Essas regras podem ser divididas em algumas categorias principais, as conversões independentes (que só dependem do grafema independentemente de contexto, uma conversão direta), conversões dependentes de contexto interno (que dependem da posição do grafema na palavra ou dos grafemas que antecedem ou sucedem na própria palavra) e conversões que dependem de contexto mais específico (classe gramatical ou sequência de duas palavras). Os casos que não se adequam a nenhuma regra são adicionados a um dicionário de exceções que pode ser acessado durante a conversão. Abaixo um esquema mostra de forma simplificada o processo de conversão de uma frase:



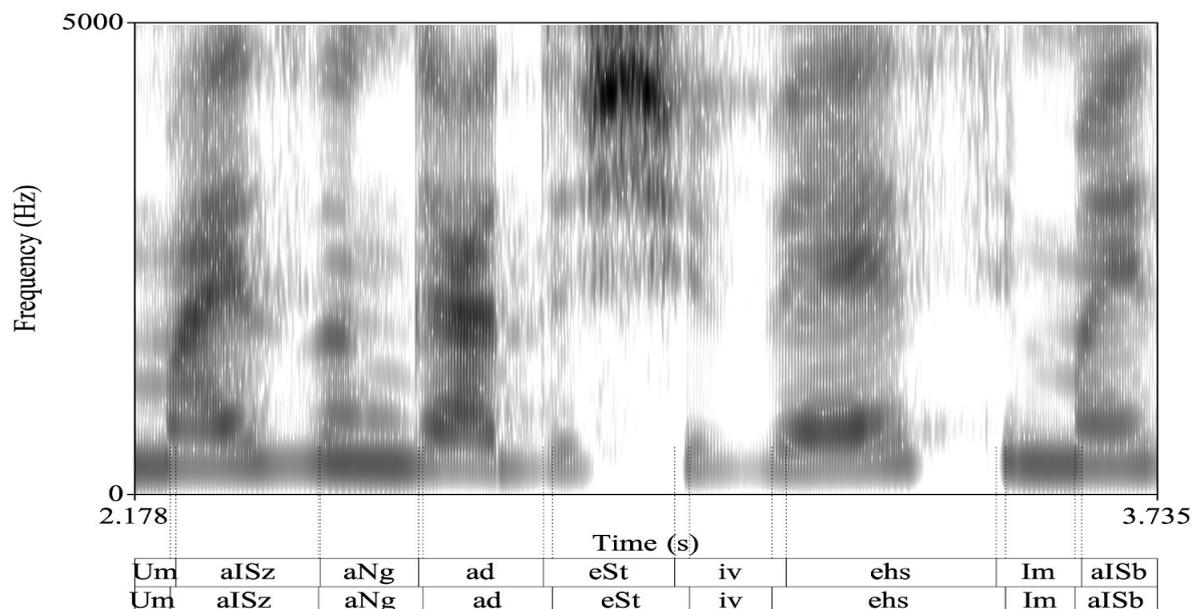
Processo de conversão de frases





O conversor foi validado usando textos contendo ao todo 1500 fonemas em frases com a maior diversidade de casos possíveis. Nesses testes o alinhador obteve uma taxa de acertos de mais de 99%, sendo o maior caso de erros a distinção entre os fonemas /e/, /ɛ/ e /o/, /ɔ/, já que não foi encontrada nenhuma regra geral que conseguisse distinguir esses casos.

Devido à quantidade de dados disponíveis para o treinamento, foi optado pelo uso de HMMs (Hidden Markov Models). Os Modelos Ocultos de Markov são um modelo estatístico que pode ser usado para descrever a evolução de eventos observáveis que dependem de parâmetros internos, de modo que ele é amplamente utilizado para a tarefa de alinhamento fonético. Por conta disso, diversos toolkits e ferramentas para facilitar o seu treinamento e uso já foram implementadas, sendo as usadas aqui o HTK e o Kaldi (toolkits para HMMs) e o Prosodylab-Aligner e o Montreal Forced Aligner (interfaces que automatizam o uso do HTK e do Kaldi). Para o treinamento foram utilizados áudios de 7 falantes distintos, contendo ao todo 1492 sílabas fonéticas. Uma das diferenças entre os modelos que usam redes neurais, por exemplo, em comparação com Modelos Ocultos de Markov é que as HMMs tendem a ter uma generalização pior, ou seja, têm uma performance pior com falantes muito diferentes dos presentes no treinamento. Por conta dessa característica os modelos foram validados usando falantes que estavam e que não estavam no treinamento separadamente. Abaixo podemos ver um exemplo de conversão realizada pelo modelo para um falante que estava no treinamento, sendo o texto em cima o alinhamento do alinhador e abaixo o alinhamento manual:



Como podemos ver, o modelo conseguiu obter alinhamento muito precisos quando comparados com os alinhamentos manuais. O melhor modelo treinado alcançou 87,8% de erros abaixo de 25 ms e uma média de 16 ms para os erros.

Os resultados tanto para o conversor quanto para o alinhador foram satisfatórios e já auxiliam nesses processos, pois eles permitem que o foneticista apenas corrija os erros obtidos pelo alinhador e pelo conversor, acelerando drasticamente essas tarefas. Por fim, para facilitar o uso desses softwares uma plataforma simples que pode ser acessada em <<https://conversoralinhador.herokuapp.com/>> ou executada localmente em <<https://github.com/jkruse27/Alinha-PB>> foi criada. Ele permite a realização de conversões de frases, além do alinhamento forçado utilizando tanto o toolkit HTK quanto o Kaldi, e fornecendo a opção de uso de modelos personalizados criados pelo próprio usuário.