



BRAVOZ – Desenvolvimento de um Sistema de Reconhecimento de Fala para o Português Brasileiro

Bolsista: Cristian Keil de Abreu

Orientador: André Gustavo Adami

CCTI – Centro de Computação e Tecnologia da Informação

UCS – Universidade de Caxias do Sul

{ckabreu, andre.adami}@ucs.br

Reconhecimento de Fala Contínua para Grandes Vocabulários

Definição

Transcrição ortográfica de um sinal de fala contínua (pausas curtas entre as palavras) com base em um grande vocabulário (mais do que 30.000 palavras).

Problema

Apesar de existir recursos para o desenvolvimento de sistemas para a língua inglesa, ainda não existe em domínio público um sistema de reconhecimento automático de voz para o Português Brasileiro,

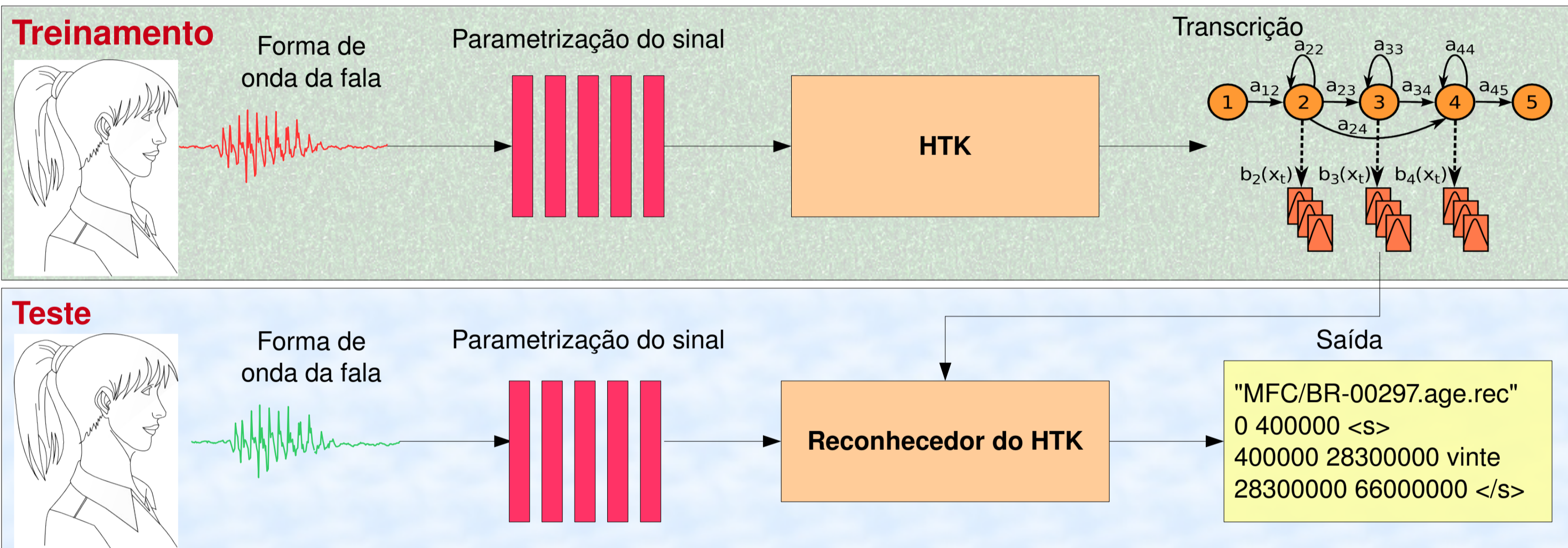
Aplicações

Discagem automática, atendimento automático, auxílio de pessoas portadoras de deficiência (comandos de voz).

Objetivos específicos

- ★ Desenvolver um sistema de referência de reconhecimento de fala contínua para o Português Brasileiro
- ★ Comparar o desempenho do sistema de reconhecimento utilizando diferentes tipos de alinhamentos fonéticos

HTK – Hidden Markov Model Toolkit



Hidden Markov Model (Modelo Oculto de Markov)

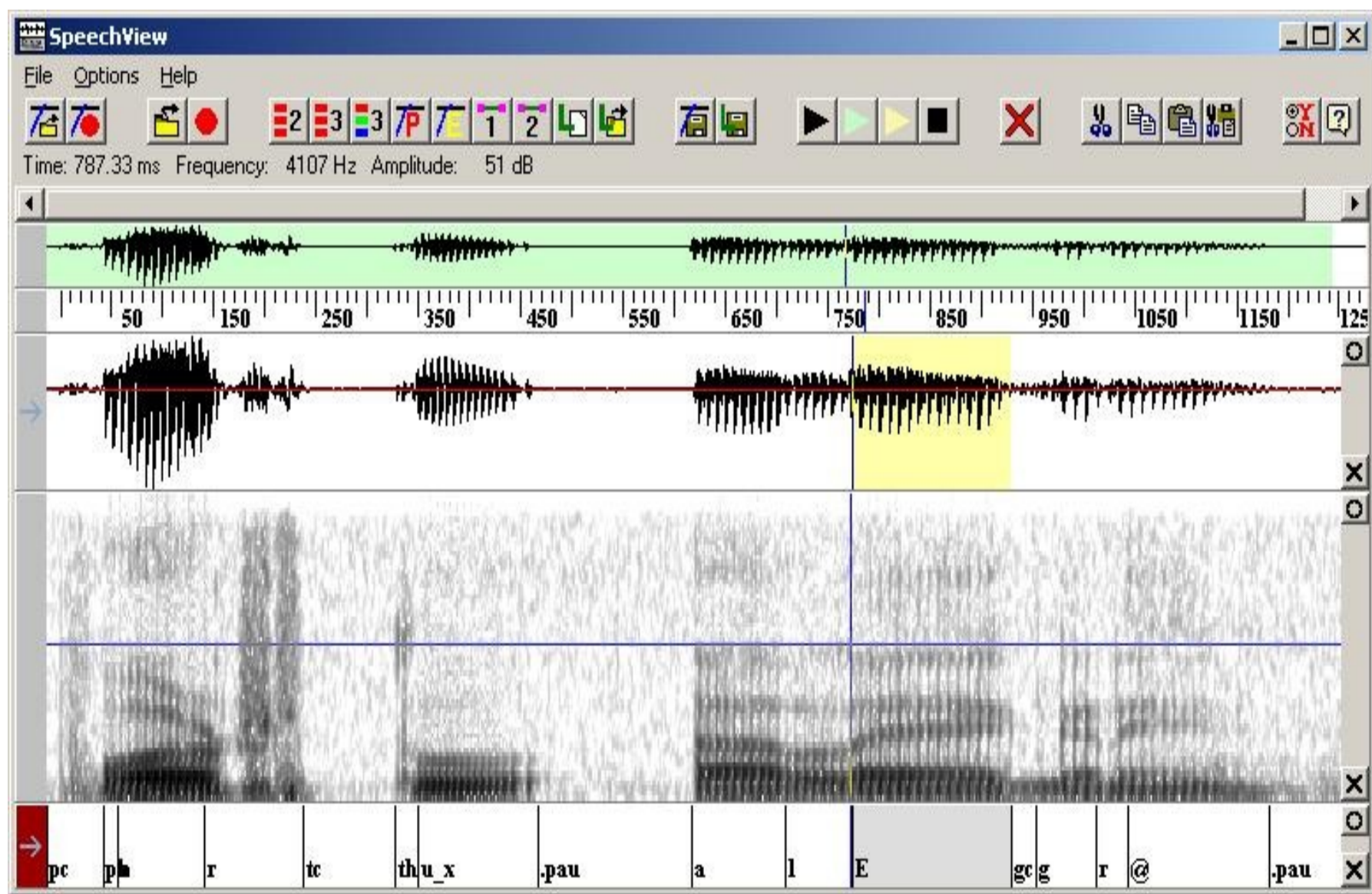
Objetivo: Modelagem de qualquer série temporal a partir de parâmetros observáveis buscando determinar parâmetros ocultos.

Aplicação: reconhecimento de padrões temporais da fala.

Esquema básico de etapas do HTK da Universidade de Cambridge (<http://htk.eng.cam.ac.uk>)

Treinamento: Diferentes tipos de alinhamento fonético

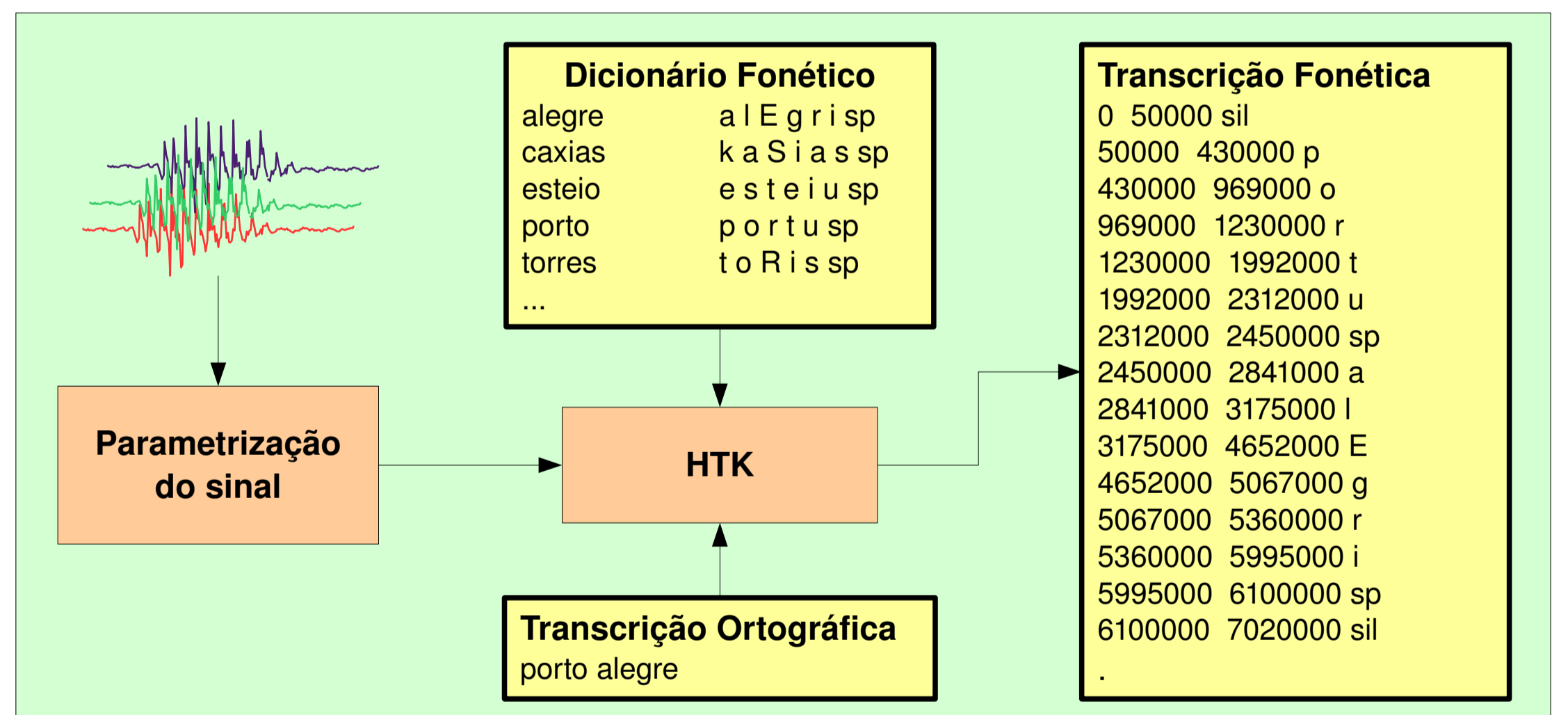
Alinhamento Manual: produzido por um foneticista.



Tela de uma ferramenta utilizada por foneticistas.

Alinhamento Automático: produzido pelo HTK. Os parâmetros de um conjunto de HMMs são estimados com base nos fonemas disponíveis em transcrições fonéticas (intervalos de tempo onde cada fonema é produzido) a fim de maximizar a taxa de palavras corretamente reconhecidas.

A ferramenta utiliza: **Dicionário fonético, Transcrição Ortográfica, Sinal parametrizado.**



Representação geral do alinhamento fonético produzido pelo HTK.

Resultados

Dados

Base: Spoltech (coletado em diversas regiões do Brasil)
Transcrição fonética: SAMPA (Alfabeto fonético dos métodos de avaliação da fala)
Número de arquivos: 263 arquivos - total de 632 palavras (41 distintas)

Sistema

Características: Coeficientes espectrais com base no modelo de audição
Modelo Fonema: HMM com 5 estados
Modelo Linguagem: bi-gramas
Avaliação: 166 arquivos para treinamento
 97 arquivos para teste
Desempenho: Palavras_Corretas / Total_Palavras

Transcrição Fonética	Número de Acertos	Taxa de Acerto
Automático	560	88,6%
Manual	553	87,5%

Considerações Finais

O resultado do reconhecimento mostra que a diferença de desempenho não é estatisticamente significativa apesar da precisão do alinhamento manual. Uma das razões para tal resultado é que o alinhamento executado pelo HTK visa a generalização enquanto que o alinhamento manual adiciona as idiossincrasias do locutor que afetam a generalização dos modelos.