



# Distvox – Reconhecimento Distribuído de fala para o Português Brasileiro

Bolsista: Janaina Graciele Steffens - BIC/UCS

Orientador: Ricardo Vargas Dorneles

Centro de Computação e Tecnologia da Informação – Universidade de Caxias do Sul  
{jgsteffe, rvdornel}@ucs.br

- ➔ **Objetivo:** Desenvolver um sistema distribuído de reconhecimento de fala contínua para ser utilizado em grandes bases de dados.
- ➔ **Ferramenta utilizada:** Toolkit HTK

## Metodologia

A gravação das diversas fontes de áudio vem se tornando uma alternativa interessante de consulta futura. Entretanto, a posterior busca em bases de dados de áudio é um trabalho complexo e para facilitar o processo de recuperação destes dados pode-se utilizar um sistema de reconhecimento de fala. Para maximizar o uso de um grande número de máquinas, o projeto Distvox visa desenvolver um sistema distribuído de reconhecimento de fala contínua para ser usado em grandes bases de dados, utilizando em o seu desenvolvimento o toolkit HTK.

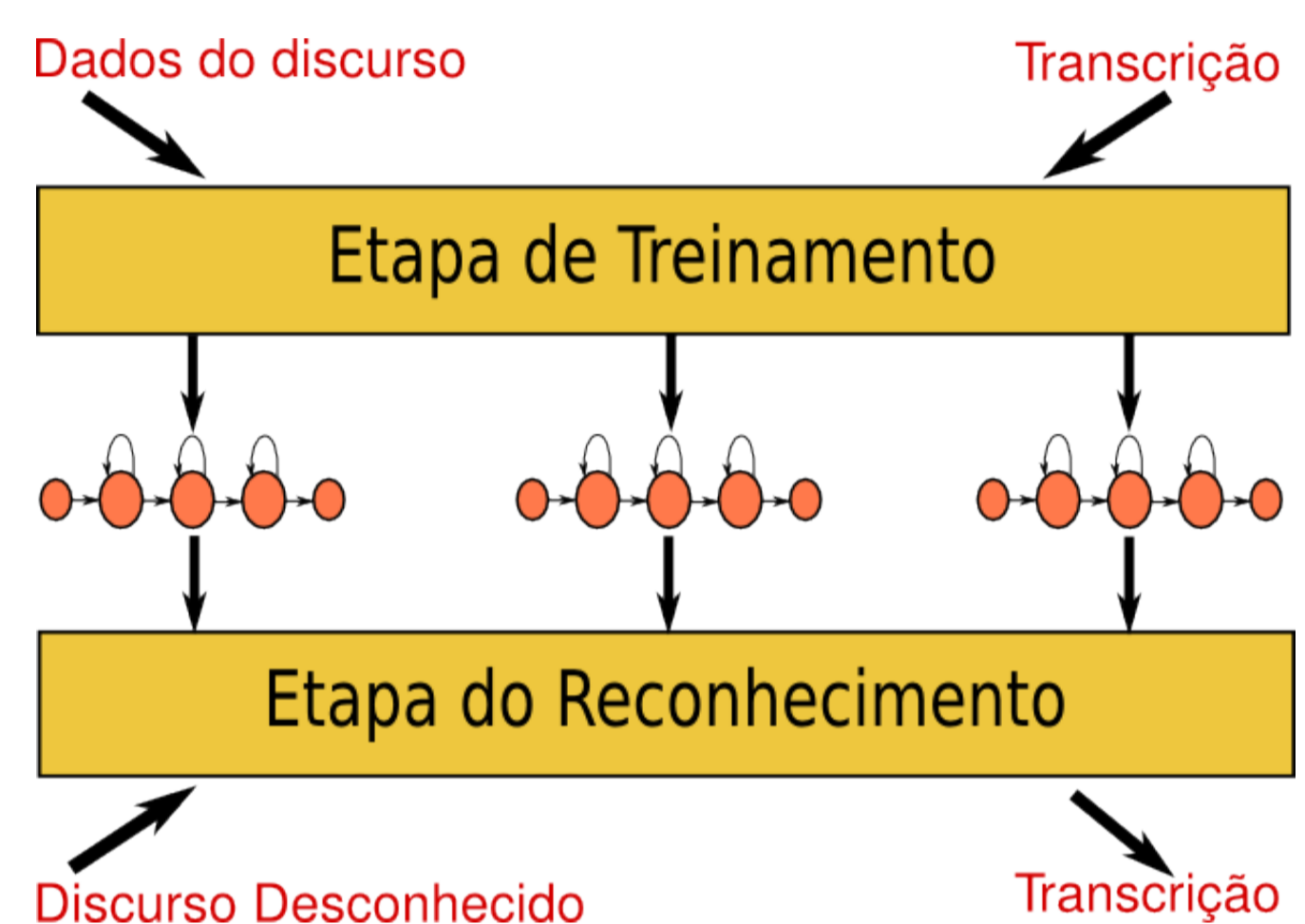
### ➔ Toolkit HTK

Possui recursos para construir e manipular modelos ocultos de Markov (HMMs).

**Utilização:** Usado principalmente para reconhecimento da fala, embora tenha sido utilizado para muitas outras aplicações, incluindo reconhecimento de caracteres e sequenciamento de DNA.

### O HTK pode ser dividido em duas etapas principais:

- Etapa de treinamento: Usado para estimar os parâmetros de um conjunto de HMMs através de declarações.
- Etapa do reconhecimento: Trata do reconhecimento do sistema através da transcrição de declarações desconhecidas.



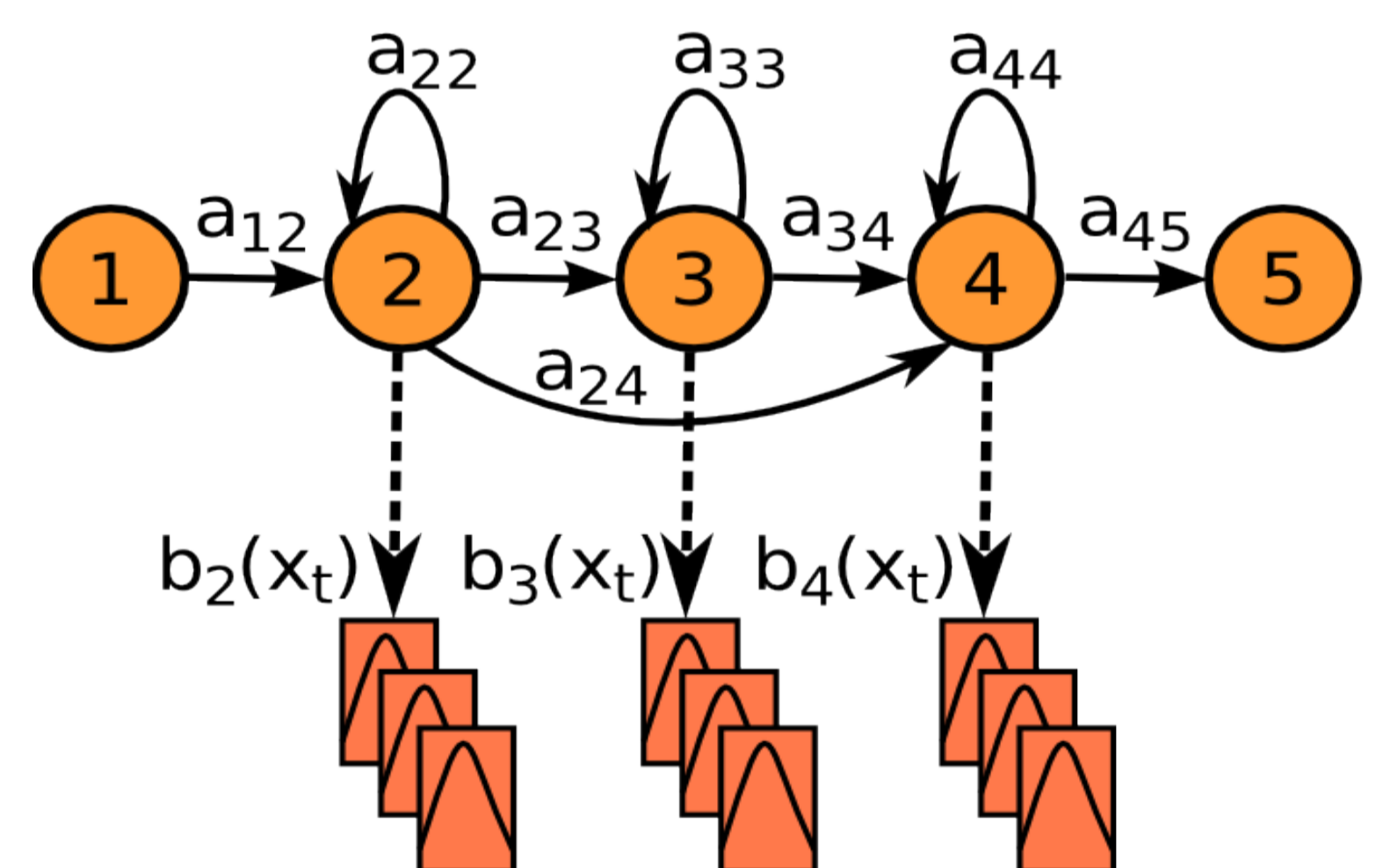
### ➔ HMMs – Modelos Ocultos de Markov

É um método estatístico para caracterizar um conjunto de dados observados de uma série temporal discreta. Pode ser visto como um processo duplamente estocástico onde o processo estocástico (sequência de estados, 1, 2, 3, 4 e 5) não é diretamente observável.

Em reconhecimento de fala os HMMs são utilizados para modelar classes acústicas que representam algum segmento menor do que uma palavra (como os fonemas).

Cada unidade acústica pode ser modelada utilizando um HMM com três estados emissores. Durante a etapa de reconhecimento os modelos são unidos para formar uma palavra ou até mesmo uma frase a fim de facilitar a união. Também são adicionados a unidade dois estados (um no início e outro no fim) não emissores.

- $x_t$  representa um dado vetor acústico emitido em uma unidade de tempo  $t$ ;
- $A = \{ a_{ij} \}$  matriz de todas as probabilidades de transições;
- $B = \{ b_i \}$  matriz de todas as probabilidades de emissão de saída;
- $\Omega = \{1, 2, \dots, n\}$  conjunto de estados.



## Resultados:

A pesquisa encontra-se em andamento, como principais resultados obtivemos até então:

- Toolkit instalado e configurado em ambiente Linux.
- Ferramentas estudadas e executadas.

## Considerações Finais:

- Desenvolvimento de um sistema de referência para o PB. O sistema de referência será desenvolvido utilizando o HTK.
- Desenvolvimento de um sistema distribuído de reconhecimento para o português brasileiro.
- Desenvolvimento e validação do sistema distribuído de fala no GRIDUCS.