



# **CLASSIFICADOR DE IDIOMAS**

**Victor De Bodt Sivieri  
Victor Velloce Ferreira  
Viviane Santos**

# INTRODUÇÃO: DESCRIÇÃO DO PROBLEMA

- Reconhecimento de idioma “falado”
- Classificação em 3 idiomas:
  - Alemão
  - Inglês
  - Português
- Baseado em processamento de áudio



# COLETA DE DADOS

- Os dados foram coletados da internet em formato mp3
- Foi coletado um total de 60 dados para cada idioma
  - 50% para o conjunto de treino
  - 50% para o conjunto de teste



# EXTRAÇÃO DE CARACTERÍSTICAS

- Entrada do extrator: vetor de amostras da captura do sinal sonoro
  - Trechos de áudio de 10s
  - Frequência de amostragem de 44.100Hz
  - Total de 441.000 amostras
- Foram utilizados dois conjuntos distintos de características para testar a eficácia



# EXTRAÇÃO DE CARACTERÍSTICAS

- Primeiro conjunto de características:
  - RMS[1]: usado como energia média do sinal

$$RMS = \sqrt{\frac{1}{T} \int_{-T/2}^{T/2} s^2(t) dt}$$

- Spectral Centroid[1]: análogo ao centro de massa para o espectro

$$C = \frac{\sum_{f=0}^M f \cdot |S(f)|^2}{\sum_{f=0}^M |S(f)|^2}$$



# EXTRAÇÃO DE CARACTERÍSTICAS

- Spectral Bandwidth[1]: variação do espectro em torno da centróide espectral

$$B^2 = \frac{\sum_{f=0}^M (f - c)^2 |S(f)|^2}{\sum_{f=0}^M |S(f)|^2}$$

- Band Energy Ratio[1]: razão da energia no primeiro quarto do espectro com relação ao espectro todo

$$BER = \frac{\sum_{f=0}^{M/4} |S(f)|^2}{\sum_{f=0}^M |S(f)|^2}$$



# EXTRAÇÃO DE CARACTERÍSTICAS

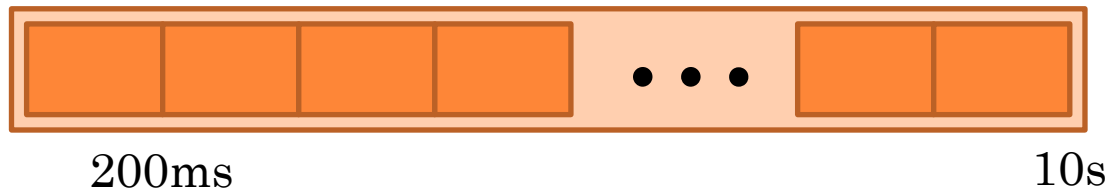
- 20 MFCCs (Mel-Frequency Cepstral Coefficients)[1]:  
espectro do espectro, muito utilizado em  
reconhecimento de voz

$$MFCC = | \mathcal{F} \{ \log[ mel(|S(f)|^2) ] \} |^2$$



# EXTRAÇÃO DE CARACTERÍSTICAS

- Segundo conjunto de características:
  - Análise espectral de 10s de áudio pode ser pouco eficaz
  - 10 MFCCs para cada trecho de 200ms





# EMBARQUE DA REDE: MOTIVAÇÃO

- O uso de telefones celulares para diversas aplicações é cada vez mais frequente
- Grande aumento na capacidade de processamento
- O reconhecimento de idioma é o primeiro passo para aplicações de tradução simultânea



# EMBARQUE DA REDE: PROCEDIMENTO

- Plataforma alvo: Android
- Software Development Kits(SDKs) facilitam o desenvolvimento de aplicativos
- Android SDK + Java
- Interação com o usuário através de uma interface gráfica
- Grava trecho de 10s por meio do microfone do celular, chama função que extrai características e chama função da rede neural

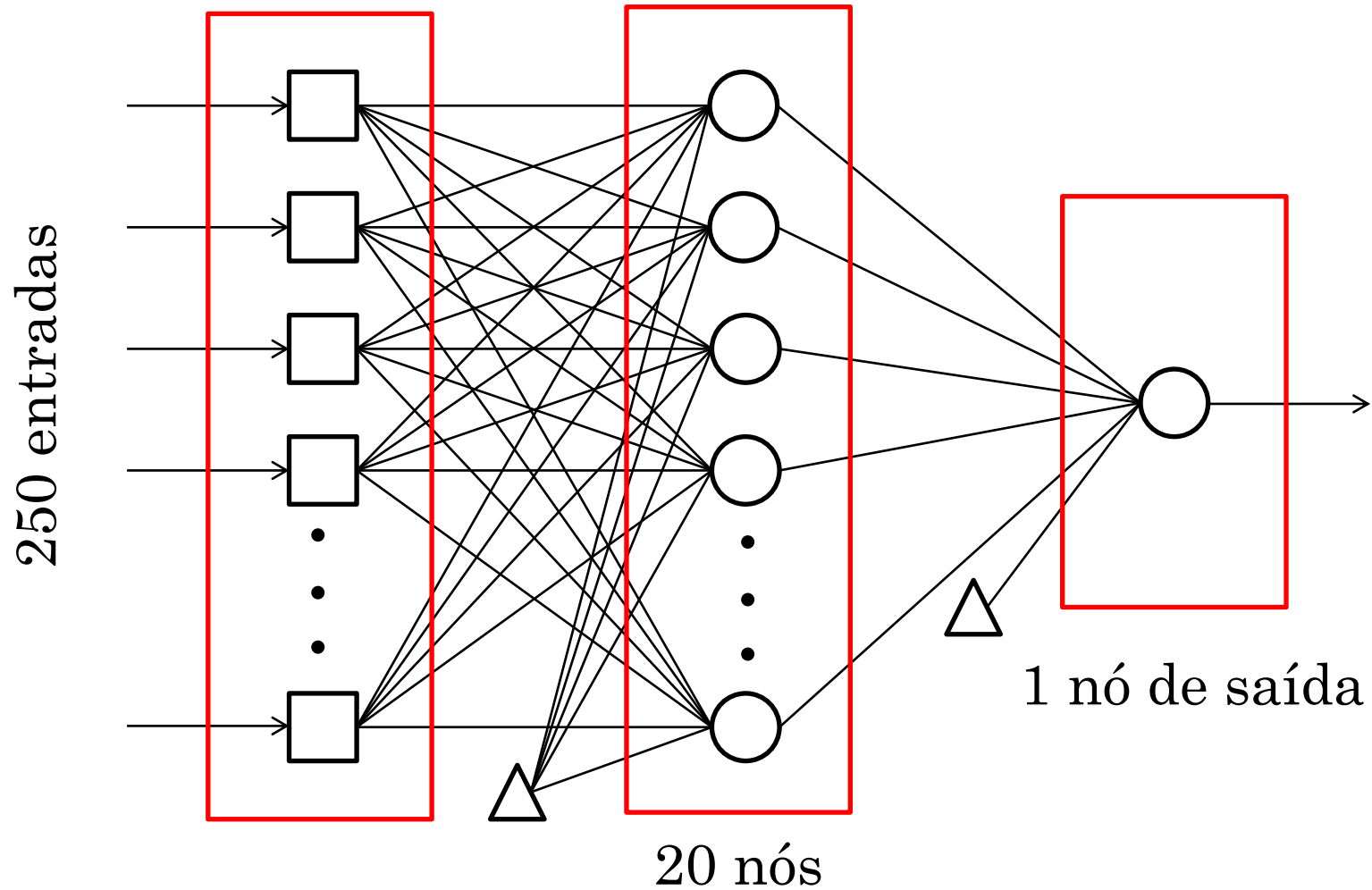


# RNA

- Três RNAs: uma para cada idioma
- 250 características
  - A partir de pré-processamento



# RNA - TOPOLOGIA



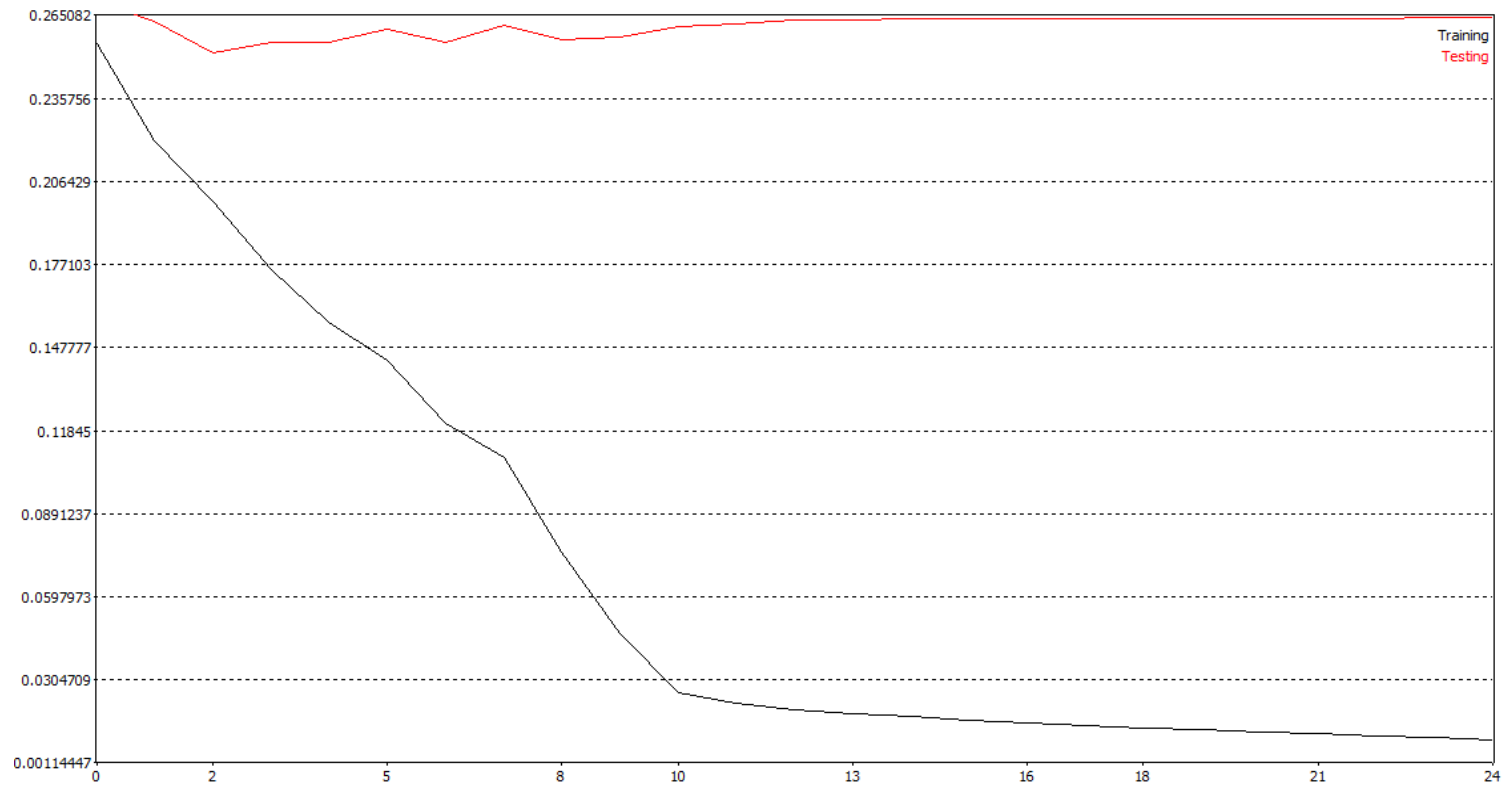
# TESTES

- De forma a aumentar a taxa de acerto, algumas variações foram feitas:
  - Número de nós
  - Número de características a partir de pré-processamentos
  - Número de exemplares



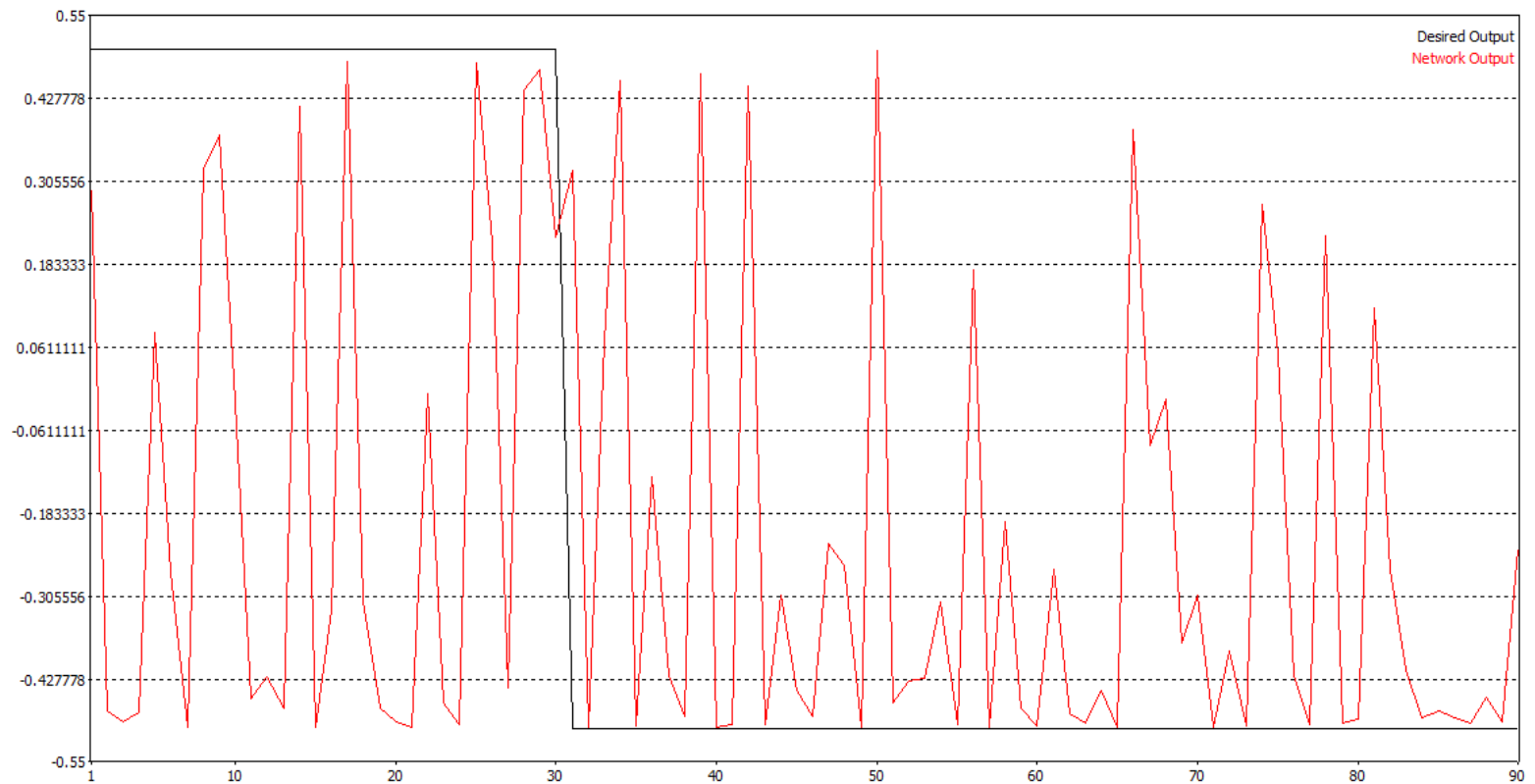
# RESULTADOS / RMS ERROR

## ○ Alemão



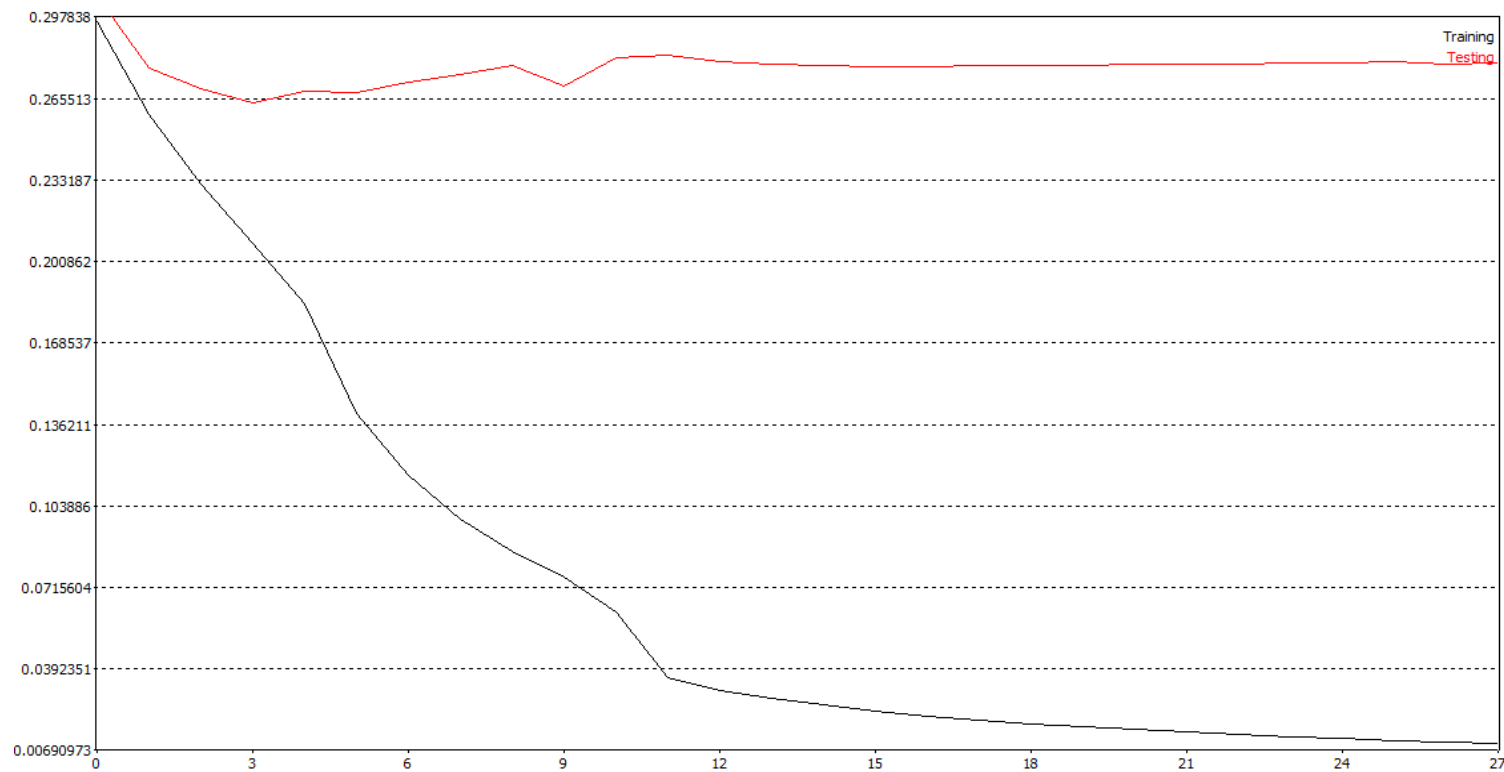
# RESULTADOS / TESTE

## ○ Alemão



# RESULTADOS / RMS ERROR

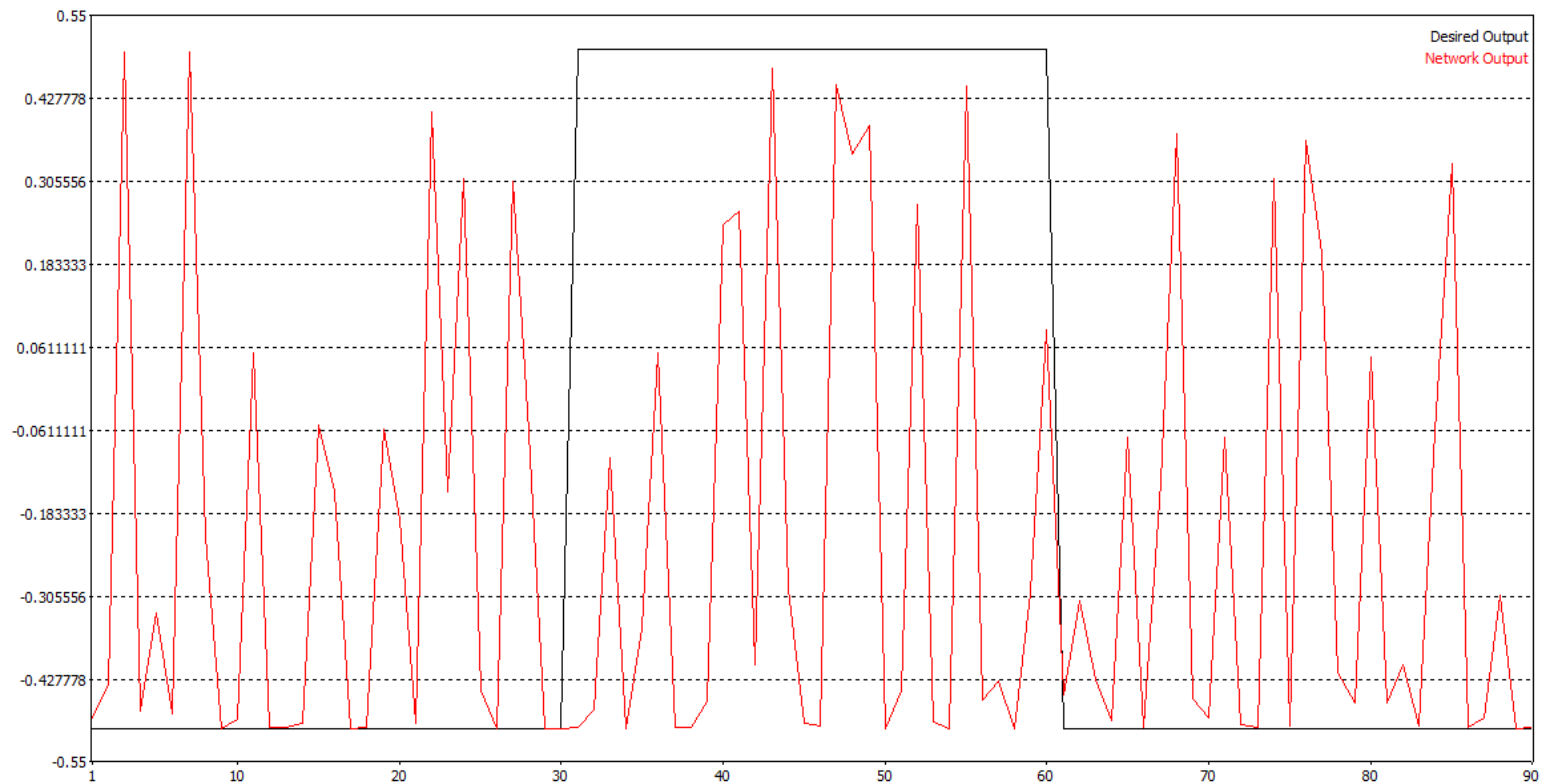
## ○ Inglês





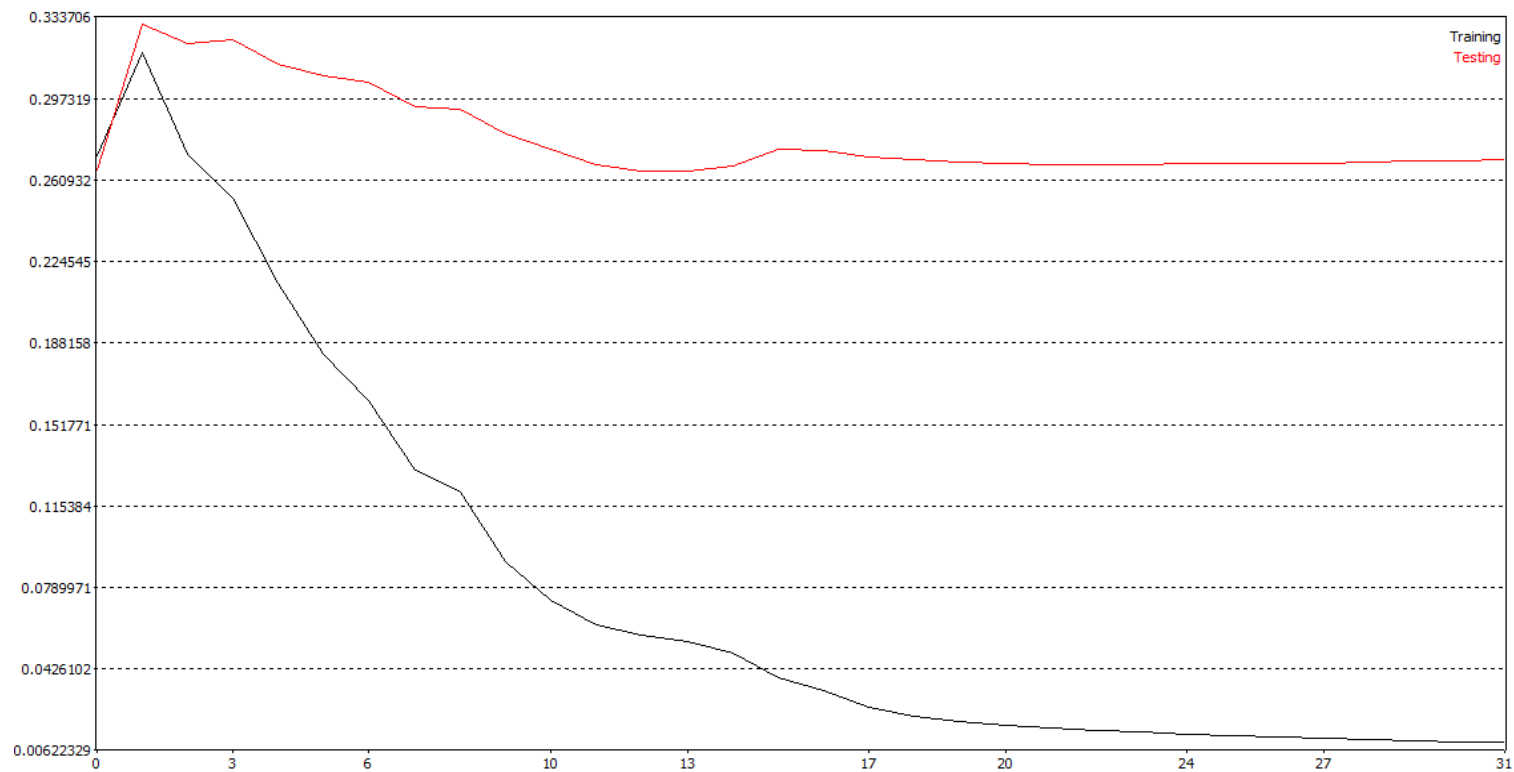
# RESULTADOS / TESTE

## ○ Inglês



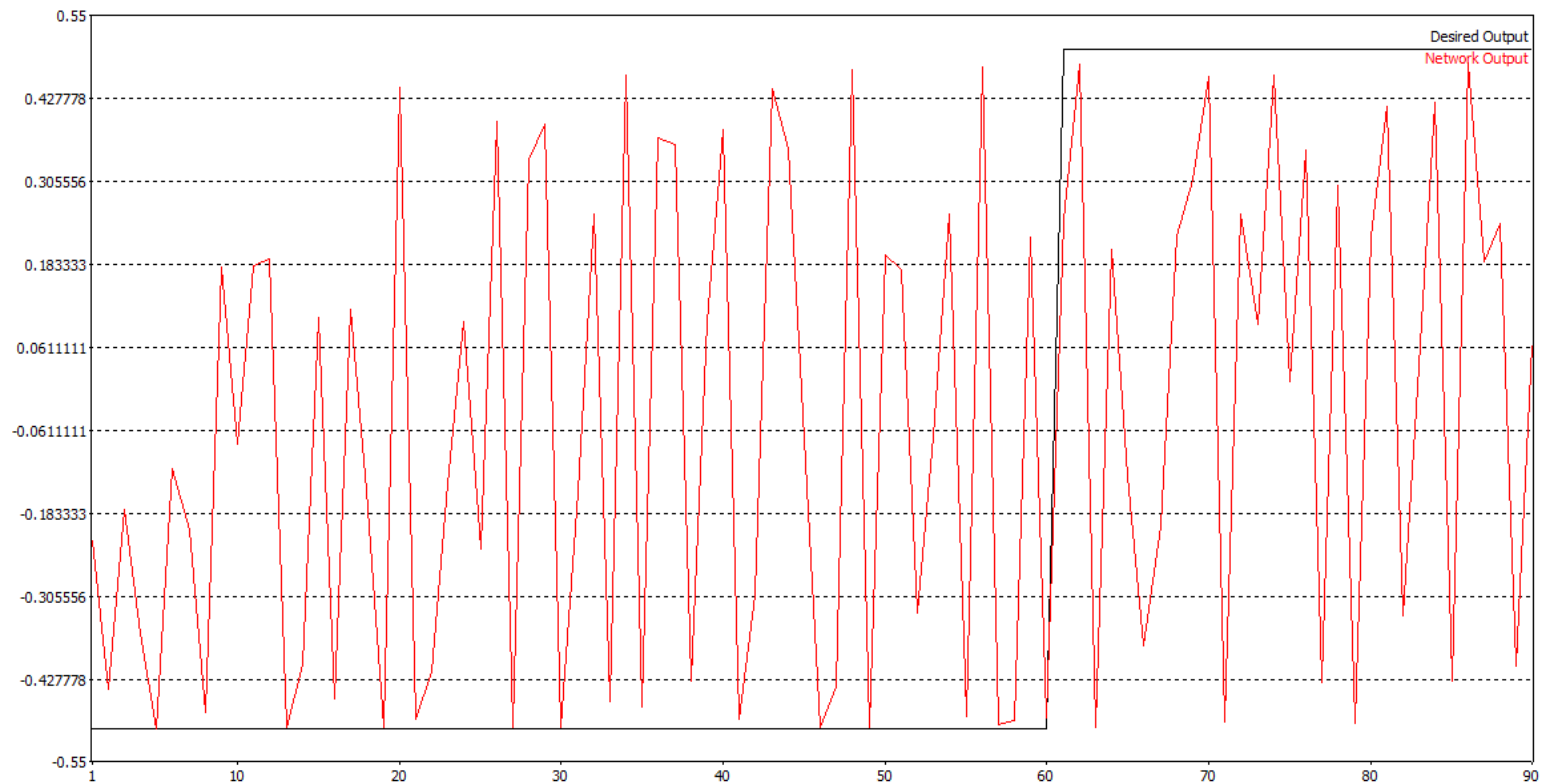
# RESULTADOS / RMS ERROR

## ○ Português



# RESULTADOS / TESTE

## ○ Português



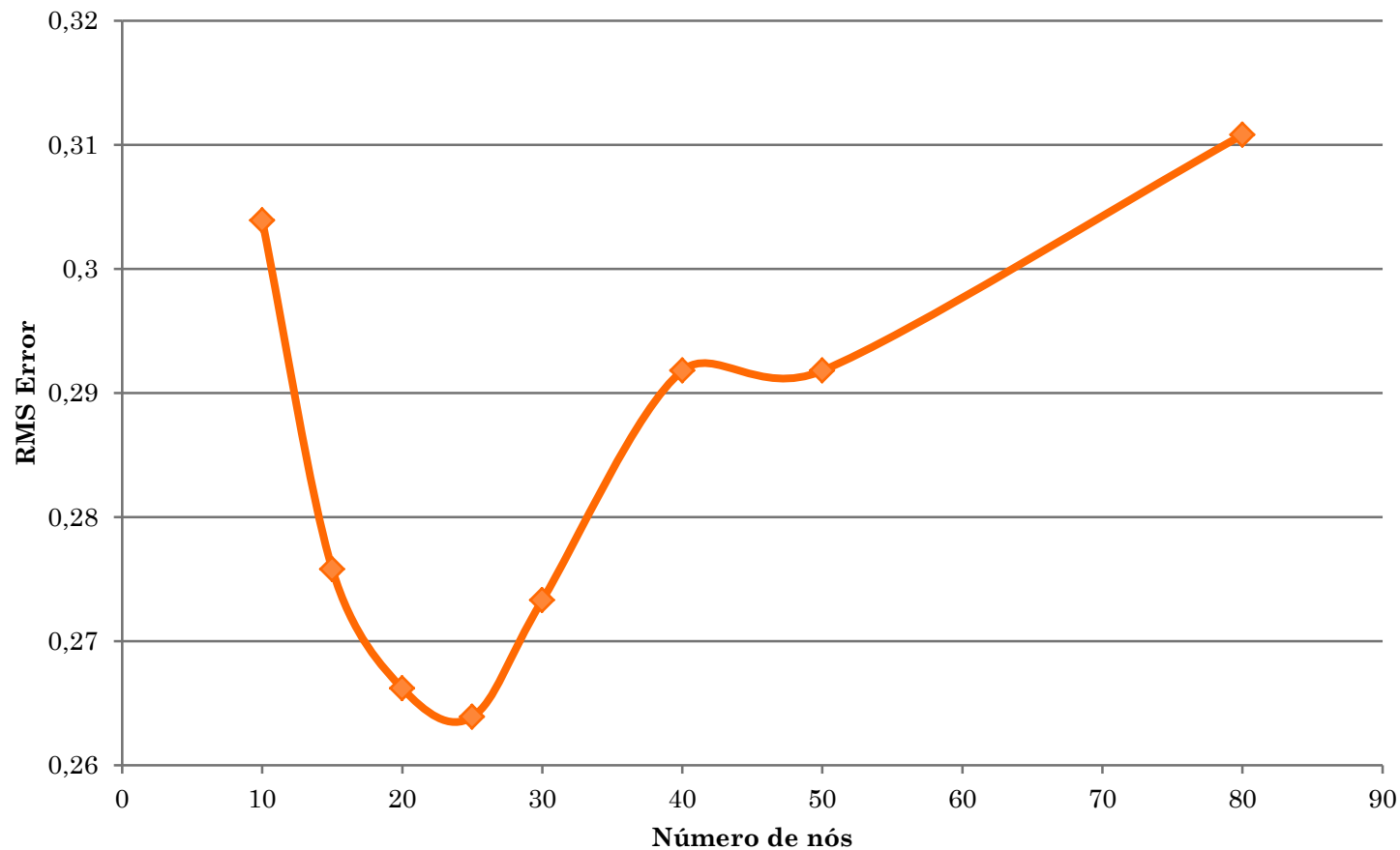
# RESULTADOS / MEDIDAS DE QUALIDADE

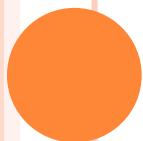
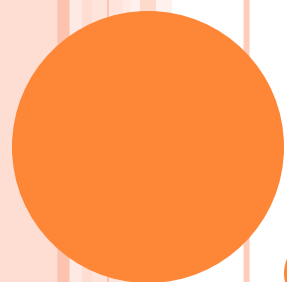
- Taxa de acerto
  - Alemão: 68,9%
  - Inglês: 54,4%
  - Português: 62,2%
- RMS
  - Alemão: 0,2662
  - Inglês: 0,3065
  - Português: 0,2699



# RESULTADOS

## ○ RMS Error (Teste) x número de nós





**MUITO OBRIGADO**