



## FORMULÁRIO PARA RELATÓRIO FINAL

### 1. Identificação do Projeto

#### Título do Projeto PIBIC/PAIC

Criação de Uma Rede Neural Artificial Para o Reconhecimento de Vocábulo Através da Fala Humana.

#### Orientador

Jorge Yoshio Kanda

#### Aluno

Mathias Afonso Guedes de Menezes

### 2. Informações de Acesso ao Documento

#### 2.1 Este documento é confidencial?

SIM

NÃO

#### 2.2 Este trabalho ocasionará registro de patente?

SIM

NÃO

#### 2.3 Este trabalho pode ser liberado para reprodução?

SIM

NÃO

#### 2.4. Em caso de liberação parcial, quais dados podem ser liberados? Especifique.

### 3. Introdução

As Redes Neurais Artificiais (RNAs) são um conceito computacional que visa trabalhar no processamento de dados de maneira parecida ao cérebro humano. O cérebro assume o papel de um processador muito complexo realizando processos de maneira paralela. Sendo assim, ele organiza sua estrutura – milhões de neurônios - de forma que eles realizem o processamento necessário [1]. Isso é feito em uma velocidade extremamente alta devido à complexidade das conexões entre os neurônios e, atualmente, não existe qualquer computador no mundo capaz de realizar o que o cérebro humano faz. A ideia de construir uma RNA é de realizar o



processamento desejado tendo como princípio a organização dos neurônios do cérebro.

As RNAs são criadas a partir de algoritmos projetados para uma determinada finalidade. É impossível criar um algoritmo desses sem ter conhecimento de modelos matemáticos que simulem o processo de aprendizagem do cérebro humano. Basicamente, uma RNA se assemelha ao cérebro em dois pontos: o conhecimento é obtido através de etapas de aprendizagem e pesos sinápticos são usados para armazenar o conhecimento. Uma sinapse é o nome dado à conexão existente entre neurônios. Nas conexões são atribuídos valores, que são chamados de pesos sinápticos. Isso deixa claro que as RNAs têm em sua constituição uma série de neurônios artificiais - ou virtuais - que serão conectados entre si, formando uma rede de elementos de processamento [11].

Tendo uma RNA montada, uma série de valores podem ser aplicados sobre um neurônio, sendo que este está conectado a outros pela rede. Estes valores - ou entradas - são multiplicados no neurônio pelo valor do peso de sua sinapse [12]. Então, esses valores são somados. E se esta soma ultrapassar um valor limite estabelecido, um sinal é propagado pela saída – axônio - deste neurônio. Em seguida, essa mesma etapa se realiza com os demais neurônios da rede. Isso quer dizer que os neurônios vão enfrentar algum tipo de *ativação*, dependendo das entradas e dos pesos sinápticos.

Existem várias formas de se desenvolver uma RNA. Ela deve ser montada de acordo com o problema a ser resolvido. Em sua arquitetura são determinados os números de camadas usadas - as camadas são formadas por neurônios, a quantidade de neurônios em cada camada, o tipo de sinapse utilizado, etc. [1, 12].

O processo de aprendizagem das RNAs é realizado quando ocorrem várias modificações significantes nas sinapses dos neurônios. Essas mudanças ocorrem de acordo com a ativação dos neurônios [13]. Se determinadas conexões são mais usadas, estas são reforçadas enquanto que as demais são enfraquecidas. É por isso que quando uma RNA é implantada para uma determinada aplicação, é necessário um tempo para que esta seja treinada.

Existem, basicamente, 3 tipos de aprendizado nas RNAs:

1. Supervisionado: neste tipo, a RNA recebe um conjunto de entradas padronizados e seus correspondentes padrões de saída, onde ocorrem ajustes nos pesos sinápticos até que o erro entre os padrões de saída gerados pela rede tenha um valor desejado;



2. Não-supervisionado: neste tipo, a RNA trabalha os dados de forma a determinar algumas propriedades dos conjuntos de dados. A partir destas propriedades é que o aprendizado é constituído;
3. Híbrido: neste tipo ocorre uma "mistura" dos tipos supervisionado e não-supervisionado. Assim, uma camada pode trabalhar com um tipo enquanto outra camada trabalha com o outro tipo.

As RNAs podem ser aplicadas para resolver uma grande quantidade de problemas. Um bom exemplo de aplicação são softwares de reconhecimento de voz, que precisam aprender a conhecer a voz de determinadas pessoas [3]. RNAs também são usadas em robôs que desarmam bombas. Existem até alguns softwares que aprendem a identificar *Web Spams* [18] em e-mails e apagá-los - e conseguem uma margem aceitável de acertos. Mas no geral as RNAs são usadas principalmente em aplicações mais complexas, como em usinas, mercado financeiro etc. [3].

É importante ressaltar que quando uma RNA é implementada, isso não significa que ela terá 100% de acerto nas situações em que trabalha. Este é um assunto que há tempos vem sendo estudado, mas que ainda sofre pesquisas e certamente terá ainda muitas inovações. Em nossa pesquisa, esperamos contribuir com a área da computação no que tange às redes neurais artificiais, pois foi essa a inspiração que idealizou este assunto, mas ainda há muito trabalho a ser feito. Dessa forma, a nossa pesquisa visa criar uma RNA para solucionar a tarefa de reconhecimento do vocábulo através da fala humana [14, 15, 16]. A utilização da estrutura de neurônios citados acima resultará em um sistema robusto com baixa taxa de erros de reconhecimento. A robustez dos algoritmos é resultado da característica intrínseca de adaptabilidade das RNAs, quanto maior o uso maior a robustez. Portanto, para garantir tal robustez do sistema composto pelos algoritmos, o projeto também trabalhará com um conjunto pré-definido de palavras que está sendo utilizado para a realização de comandos de voz para esse sistema.

#### 4. Justificativa

Assim que um computador conseguir entender e identificar a voz humana e obedecer a comandos verbais, o homem terá revolucionado mais uma vez a informática. Criando uma ferramenta computacional com tal funcionalidade, o usuário pode emitir um comando de voz para o computador, smartphone e ditar um texto, iniciar uma aplicação, tirar uma foto, registrar lembretes fazer ligações, etc. O usuário pode até mesmo fazer restrições de uso a determinadas informações e através da sua voz, ter acesso a elas (Reconhecimento Automático da Fala – RAF) [14, 16].



A tecnologia de reconhecimento de vocábulo visa facilitar o uso de computadores e smartphones, como também auxiliar pessoas com problemas de dicção, distúrbio da linguagem ou audição [8, 9]. Um modelo computacional de reconhecimento de voz pode ter diferentes objetivos, tais como:

**Identificação de palavras:** a mesma palavra pronunciada por pessoas diferentes deve ser considerada como igual, ou seja, os padrões a serem reconhecidos são as palavras [10, 4]. Sendo esta aplicadas em sistemas em que diferentes usuários dão os mesmos comandos vocais, como por exemplo o acionamento de equipamentos e o auxílio para navegação em softwares para microcomputadores [5].

**Identificação de pessoas:** há o reconhecimento da pessoa através da pronúncia de alguma palavra, sendo esta aplicação utilizada em sistemas de segurança, como por exemplo, o controle de acesso em ambientes ou sites da internet [3].

**Identificação de pessoa/palavra:** Uma amostra só é reconhecida quando uma determinada pessoa pronuncia uma determinada palavra. É a implementação mais simples, sendo aplicada de maneira menos dinâmica nas duas versões anteriores.

Em Brandão [2] foi implementado um sistema de Reconhecimento Automático de Voz (RAV) com vocabulário e número de locutores restritos. Técnicas de processamento digital de sinais e RNAs foram utilizadas. A filtragem do sinal de voz foi realizada por *software* através da aplicação de filtros digitais a fim de minimizar o efeito de borda devido ao truncamento do sinal na etapa de aquisição e os ruídos de fundo inerentes ao sinal. Rotinas para detecção de início e fim de cada Comando de Voz (CVZ) foram implementadas com a finalidade de diminuir o tamanho do vetor o qual contém as amostras do sinal de voz.

O trabalho de Dias [7] apresenta a constituição de um sistema básico de reconhecimento de voz em língua portuguesa, utilizando como ferramenta as RNAs. Na execução do projeto foi utilizado o Mapa de Kohonen, por se constituir em treinamento não supervisionado. Como bases de dados para treinamento das RNA's foram selecionadas 20 (vinte) vocábulos que poderão ser reconhecidos pelo sistema. O pré-processamento de sinais foi realizado utilizando-se filtros e a Transformada Rápida de Fourier.

A ferramenta implementada por Souza [19] tinha como objetivo reconhecer vogais isoladas através de algoritmos de RNAs. No trabalho realizado foram utilizadas 15 amostras de cada vogal coletadas de 6 faladores. Essas amostras foram processadas e como resultados foram extraídas 20 características, as quais foram introduzidas na rede neural. A rede apresentou um acerto médio de 70%, com voz de



pessoas que não participaram do treinamento.

O sistema de reconhecimento de voz por redes neurais também permite que o usuário tenha maior mobilidade, pois fica livre para realizar outras tarefas, como por exemplo, realizar pesquisa ou ler um texto enquanto está ditando. O reconhecimento de voz permite uma taxa de escrita que calcula a média em torno de 45 a 65 palavras por minuto, sendo que alguns usuários alcançam pontuações tão altas quanto 90 palavras por minuto. Um usuário que dê a entrada de dados pode digitar 80 palavras por minuto [6], mas não pode fazer isso o dia todo sem pausar para caminhar e descansar, então as taxas de digitação e reconhecimento de voz podem ser quase as mesmas. Além disso, o reconhecimento de voz quase sempre acha a palavra exata, quando sua soletração é certa. O usuário também teria proteção contra danos físicos, já que por não digitar ou utilizar um mouse, pelo menos não muito frequentemente, o usuário será menos propenso a sofrer Lesão por Esforço Repetitivo (LER) ou Distúrbios Osteomusculares Relacionados ao Trabalho (DORT).

A tecnologia de reconhecimento de vocábulo permite a pessoas com limitações físicas o acesso ao computador, e serve para aqueles usuários que não podem digitar ou tenham problemas ao utilizar um teclado, sendo assim uma importante ferramenta na área de inclusão digital.

## **5. Objetivos**

Para melhor compreensão dos objetivos, estes foram divididos em geral e específicos.

### **5.1. Objetivo geral:**

A presente pesquisa tem como objetivo identificar comandos de voz de palavras isoladas independente do locutor que as pronuncia, para aquisição, filtragem e extração de características do sinal de voz e RNAs para realizar a tarefa de reconhecimento de vocábulo.

### **5.2. Objetivos específicos:**

- 5.2.1. Estudar o processo de formação da fala;
- 5.2.2. Identificar as características do sinal de voz a serem utilizadas para facilitar o processo de reconhecimento das palavras ditadas;
- 5.2.3. Estudar RNAs e, posteriormente, aprofundar na área dos problemas envolvendo reconhecimento de sinais de voz;
- 5.2.4. Estudar métodos de reconhecimento por comparação de padrões;

5.2.5. Verificar a possibilidade de implementação de um sistema de Reconhecimento Automático de Palavras em tempo real;

## 6. Metodologia

Durante os primeiros meses dessa pesquisa, realizou-se o levantamento bibliográfico sobre RNAs e extração de dados da fala, para se obter o conhecimento do tema deste trabalho. Foram pesquisados também os equipamentos e softwares necessários para execução dos algoritmos desenvolvidos. Primeiro foi preciso processar e tratar a fala utilizando uma placa Arduino Uno R3 e uma *Shield*, combinando esses dois componentes foi possível realizar a extração de dados da voz. Em seguida foi pré-definido uma lista de palavras que foram utilizadas para treinar a RNA. Após essa etapa a RNA é responsável por comparar a palavra ditada com as amostras usadas na fase de treinamento. Onde a palavra ditada foi validada e exibida ao locutor.

Para chegar ao objetivo desse trabalho foram listadas quatro etapas fundamentais explicadas detalhadamente nas sessões a seguir:

### 6.1. Processamento e tratamento da fala

Nessa etapa, o sinal analógico de voz foi representado sob a forma digital e processado de maneira que eliminou ou atenuou ruídos no canal, variações na amplitude ou, ainda, estresses do locutor (estado emocional) [8]. Para reconhecer a palavra isolada, como foi o caso desse trabalho, foi necessário determinar o início e o fim de uma amostra de voz a fim de separar ruídos de fundo do sinal. Para tornar o sinal mais puro possível, para que nas próximas sessões pudesse extrair as características, filtros digitais foram aplicados com objetivo de excluir frequências irrelevantes.

### 6.2. Padrões de referência pré-definidos

Os padrões foram estabelecidos no momento em que foram definidas as palavras reconhecidas. Estes padrões foram utilizados para treinar o sistema e as demais amostras de palavras para fazer o teste e a validação do mesmo.

### 6.3. Comparação dos padrões

Após a etapa de treinamento, outro conjunto de amostras distintas dos padrões de referência foi inserido no sistema para que se pudesse analisar a eficiência do reconhecedor de vocábulo. As amostras desconhecidas foram comparadas com os padrões de referência e o



reconhecedor de vocábulo apresentou um resultado que possibilitou a verificação ou não da generalização do sistema de reconhecimento [17, 20].

#### **6.4. Saída e validação da palavra ditada**

Essa etapa consistiu em fazer os ajustes finais de amplitude da saída da estrutura neural avaliada e assim concebida a afirmação de que a palavra ditada por comando de voz foi identificada sem que houvesse dupla interpretação das demais estruturas envolvidas.

### **7. Resultados e Discussão**

A aplicação das RNAs atinge todas as áreas no que tange a sociedade, auxiliando o ser humano no desenvolvimento da economia, da indústria, da saúde e da educação. Durante as últimas décadas muitas pesquisas foram realizadas objetivando o desenvolvimento e aprimoramento da Inteligência Artificial em relação as RNAs. Porém ainda há um vasto campo de conhecimento a ser investigado e expandido. Para criar e manipular uma RNA, são necessárias muitas técnicas.

O objetivo principal dessa pesquisa é reconhecer o vocábulo por meio da fala humana. A partir dessa ideia pesquisas envolvendo RNAs podem promover grandes inovações científicas e tecnológicas na utilização de aplicações como: reconhecimento de vocábulo, reconhecimento do locutor, ferramenta de apoio ao aprendizado infantil ou de pessoas com problemas de dicção, ferramenta importante na área de inclusão digital, ferramenta de apoio à segurança pública, dentre outras mais aplicações.

Durante o andamento do trabalho foi feito levantamento bibliográfico com a finalidade de compreender a robustez de uma rede neural artificial em processar dados e classificar informações. Foi realizado o estudo de como extrair dados da fala humana através da utilização de algoritmos e mecanismos que permitem tal feito. Também foi implementada uma rede neural artificial na linguagem de programação C++, linguagem utilizada pela IDE Arduino – Programa utilizado para executar os algoritmos implementados. Esta rede neural operou de forma a retornar a palavra ditada pelo locutor. Utilizando uma placa de Arduino e uma *shield* foi possível extrair dados da voz e eliminar ruídos que atrapalham na validação da palavra.

Foram pré-definidas cinquenta (50) palavras comumente utilizadas nas áreas que tangem a informática, palavras tanto em língua inglesa quanto em língua portuguesa. As suas principais características foram extraídas e transformadas em pesos sinápticos para poder operar dentro da RNA.



Nos meses onde foram feitos os testes realizados para validar a eficácia e eficiência do sistema reconhecedor de vocábulo, foram utilizados três (3) locutores, cuja função era ditar as palavras – as cinquenta palavras pré-definidas. Houve uma variância significativa entre um locutor e outro, pois, uma determinada palavra só era reconhecida caso um determinado locutor a ditasse. Por fim, trinta e duas (32) palavras das cinquenta (50) foram reconhecidas ao serem ditadas pelos três (3) locutores, totalizando uma acurácia de 64% da RNA em reconhecer vocábulos.

Embora os sistemas de reconhecimento de vocábulo sejam bastante reconhecidos e até comercializados, além de serem utilizados em diversos campos de pesquisa – como já foi citado antes – procurou-se implementar um sistema reconhecedor de vocábulo baseando-se em uma RNA *Perceptron* de multicamadas, com pouco processamento, tendo em vista que o algoritmo seria compilado dentro de uma placa Arduino. Portanto, como o desempenho da RNA perante as palavras ditadas não foi tão satisfatório – tendo uma acurácia de 64% em relação ao total de palavras utilizadas nessa pesquisa - o sistema de reconhecimento não pode ser utilizado.

O presente trabalho terá continuidade, onde será possível melhorar o desempenho da RNA, estudando novas redes que aprimorem o sistema de reconhecimento de vocábulo. Tornando assim o sistema mais robusto e possibilitando a expansão da quantidade de palavras reconhecidas.

## 8. Referências

- [1] Braga, A. P., Carvalho, A. C. P. L. F., Ludemir, T. B., *Redes Neurais Artificiais: Teoria e Aplicações*, Rio de Janeiro: LTC, 2000.
- [2] Brandão A. S., *Redes Neurais Artificiais Aplicadas ao Reconhecimento de Comandos de Voz*. UFV, Viçosa - Minas Gerais, 2005.
- [3] Caricati, A. M., Weigang, L. *Reconhecimento de locutores em língua portuguesa com modelos de redes neurais e gaussianos*. UNB, 2001.
- [4] Chou, W., Juang, B.H., *Pattern recognition in speech and language processing*. CRC Press, 2003.
- [5] Chu, W. C., *Speech coding algorithms*. Wiley-Interscience, 2003.
- [6] Davis, H., Biddulph, R., Balashek, S., *Automatic recognition of spoken digits*. The Journal of the Acoustical Society of America, 1952.
- [7] Dias, M. V. S., *Reconhecimento de Vocábulos Utilizando Redes Neurais*. UniCEUB, Brasília – DF, 2008.
- [8] Furui, S. *Digital speech processing, synthesis and recognition*. Marcel Dekker, Inc., 1989.



# UNIVERSIDADE FEDERAL DO AMAZONAS

## RELATÓRIO FINAL PIBIC/PAIC 2015-2016



UFAM

- [9] Furui, S. *Speech Recognition - Past, Present and Future*. NTT Review, 1995.
- [10] Gold, B., Morgan, N., *Speech and Audio Signal Processing*, New York, John Wiley & Sons, Inc., 2000.
- [11] Haykin, Simon. *Redes Neurais – princípios e prática*. Porto Alegre: Bookman, 1999.
- [12] Kovacs, Zsolt Laszlo. *Redes Neurais Artificiais: Fundamentos e Aplicações*. São Paulo: Livraria da Física, 1996.
- [13] Ludemir, Teresa B; Carvalho, André Carlos P.L.F; Braga, Antônio P. *Redes Neurais Artificiais: Teoria e Aplicações*. Rio de Janeiro: LTC, 2000.
- [14] Martins, J. A. *Avaliação de diferentes técnicas para reconhecimento de fala*. Tese de doutorado. UNICAMP, SP, 1997.
- [15] Moraes, E. S. *Reconhecimento automático de fala contínua empregando Modelo Híbridos ANN + HMM*. Dissertação de mestrado. UNICAMP, SP, 1995.
- [16] Paula, M. B. *Reconhecimento de palavras faladas utilizando redes neurais artificiais*. Monografia de final de curso. UFPEL, 2000.
- [17] Rabiner, L. R., Schafer, R.W., *Digital Processing of Speech Recognition*, Englewood Cliffs, Prentice-Hall, 1978.
- [18] Silva M. R., Almeida A. R., Yamakami A., *Redes Neurais Artificiais para Detecção de Web Spams*. VIII Simpósio Brasileiro de Sistemas de Informação (SBSI - 2012).
- [19] Souza, de N., Souza F. W. G., Conceição M. L., Silva J. A. dos S., *Utilização de Redes Neurais Artificiais no Reconhecimento de Vogais*, XIV CONEMI, Salvador – BA, 2014.
- [20] Ynoguti, C. A. *Reconhecimento de Fala Contínua Usando Modelos Ocultos de Markov*. Tese de Doutorado. UNICAMP, SP, 1999.

