

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE ENGENHARIA
ENGENHARIA ELÉTRICA – HABILITAÇÃO EM ROBÓTICA E AUTOMAÇÃO
INDUSTRIAL

Ana Carolina Caetano Teixeira

Controle de Prótese Robótica por Comandos de Voz utilizando Redes Neurais Artificiais

Juiz de Fora

2017

Ana Carolina Caetano Teixeira

Controle de Prótese Robótica por Comandos de Voz utilizando Redes Neurais Artificiais

Trabalho de conclusão de curso apresentado à Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito para obtenção do grau de Engenheira Eletricista.

Orientador: Leonardo Rocha Olivi

Juiz de Fora

2017

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Teixeira, Ana Carolina Caetano.

Controle de Prótese Robótica por Comandos de Voz utilizando Redes Neurais Artificiais / Ana Carolina Caetano Teixeira. -- 2017. 50 f.

Orientador: Leonardo Rocha Olivi

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia, 2017.

1. Prótese Robótica. 2. Reconhecimento Automático de Voz . 3. Redes Neurais Artificiais . I. Olivi, Leonardo Rocha, orient. II. Título.

Ana Carolina Caetano Teixeira

Controle de Prótese Robótica por Comandos de Voz utilizando Redes Neurais Artificiais

Trabalho de conclusão de curso apresentado à Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito para obtenção do grau de Engenheiro Eletricista.

Aprovada em (dia) de (mês) de (ano)

BANCA EXAMINADORA

Professor Leonardo Rocha Olivi
Universidade Federal de Juiz de Fora
Orientador

Professora Ana Sophia Cavalcanti Alves Vilas Boas
Universidade Federal de Juiz de Fora

Professor Exuperry Barros Costa
Universidade Federal de Juiz de Fora



ATA DE APRESENTAÇÃO DE TRABALHO FINAL DE CURSO

DATA DA DEFESA: 10 / 07 / 2017

CANDIDATO: ANA CAROLINA CAETANO TEIXEIRA

ORIENTADOR: PROF. LEONARDO ROCHA OLIVI

TÍTULO DO TRABALHO: **CONTROLE DE PRÓTESE ROBÓTICA POR COMANDOS DE VOZ UTILIZANDO REDES NEURAIS ARTIFICIAIS**

BANCA EXAMINADORA/INSTITUIÇÃO:

PREDIDENTE: PROF. LEONARDO ROCHA OLIVI

AVALIADOR 1: PROF^a. ANA SOPHIA CAVALCANTI ALVES VILAS BOAS

AVALIADOR 2: PROF. EXUPERRY BARROS COSTA

HORÁRIO E LOCAL: 10H00, ANFITEATRO I ED. ITAMAR FRANCO, FACULDADE DE ENGENHARIA , UFJF.

Nesta data, em sessão pública, após exposição oral de 27 minutos, o candidato foi arguido pelos membros da banca. Em decorrência desta arguição, a banca considerou o candidato:

APROVADO

REPROVADO

Na forma regulamentar foi lavrada a presente Ata que é abaixo assinada pelos membros da banca na ordem determinada e pelo candidato:

PREDIDENTE: Leonardo Rocha Olivi
AVALIADOR 1: Ana Sophia Cavalcanti Alves Vilas Boas
AVALIADOR 2: Exuperry Barros Costa
CANDIDATO: Ana Carolina e Teixeira

Juiz de Fora, 10 de julho de 2017.

Vistos: (coordenador do curso ou presidente da comissão de TFCs do curso).

Leonardo Rocha Olivi
Coordenador da Engenharia Elétrica
Robótica e Automação Industrial
Faculdade de Engenharia - UFJF

AGRADECIMENTOS

Foram muitas as pessoas que contribuíram direta e indiretamente para a execução deste trabalho, que não é apenas um trabalho, mas representa a conclusão de um ciclo.

Agradeço, primeiramente, ao meu Jesus, em quem me espelho a cada dia para me tornar uma pessoa melhor e que me ajuda sempre a dar o meu máximo em tudo que faço, me sustentando e me confortando com seu amor e sua presença.

Também agradeço aos meus pais, Gilda e José Luiz, que nunca mediram esforços e empenharam suas vidas para que eu pudesse chegar até aqui. Sempre me incentivaram e me compreenderam e, graças a eles, tive base e suporte suficientes para concluir esta etapa. Tudo que sou e tenho, eu devo à eles. Juntamente com eles, agradeço à toda minha família que esteve sempre torcendo e me incentivando.

Ao meu professor, orientador e coordenador, Leonardo Olivi, não existem palavras que possam expressar a minha gratidão. Sempre presente e mais do que excelente em tudo que faz. Foi um privilégio a oportunidade de trabalhar e aprender com ele, e, tenho certeza que sem a sua ajuda e suporte, não teria chegado até aqui. Agradeço por todos os ensinamentos, pela paciência, dedicação e por me mostrar um mundo novo na robótica pelo qual me apaixonei.

Ao meu namorado, David, que me ensina a levar a vida de forma mais leve, o que foi essencial ao longo deste processo. Agradeço pela paciência, por estar sempre presente, sempre acreditando em mim mais do que eu mesma e por sempre me abrir os olhos para perspectivas que nunca veria por mim mesma. Seu amor e carinho depositados em mim foram mais que essenciais para eu prosseguir na caminhada.

À todos os meus amigos de dentro e fora da graduação que sempre compreenderam minha ausência e sempre acreditaram em mim. Agradeço em especial ao Lucas e Gustavo que me apoiaram ao longo da graduação e me ajudaram nos momentos mais difíceis quando eu achava que nada iria dar certo. E, ainda, à Fabiane e Bruna, que chegaram nos últimos momentos e, quando percebi, já haviam se tornado essenciais. Agradeço por sempre me ouvirem, me consolarem e me ensinarem. Com vocês o processo se tornou mais colorido e divertido.

“Para as pessoas sem deficiência, a tecnologia torna as coisas mais fáceis. Para as pessoas com deficiência, a tecnologia torna as coisas possíveis”
(Radabaugh)

RESUMO

O presente trabalho visa a elaboração de um método para controle de uma prótese robótica através de comandos de voz. O método de controle se baseia na implementação de uma Rede Neural Artificial, a técnica mais utilizada na literatura como classificador de padrões. O trabalho está dividido em duas partes, a primeira parte é o desenvolvimento do algoritmo que gera características de comandos de voz e a aplicação de um algoritmo de Redes Neurais Artificiais para o reconhecimento dos comandos de voz, a segunda parte é aplicação do algoritmo em uma prótese real. O contexto das palavras é escolhido de modo a levar uma prótese robótica a executar diretamente comandos de voz dados pelo usuário, bem como para o refino de movimentos enviados à prótese por outras vias, como eletromiografia e eletroencefalografia. O trabalho se dividirá em: aquisição das amostras, filtragem dos sinais, treinamento da RNA e validação dos dados. Os resultados obtidos são satisfatórios, mostrando a eficácia do método utilizado.

Palavras-chave: Robótica Assistiva, Reconhecimento Automático de Voz, Redes Neurais Artificiais.

ABSTRACT

The present work elaborates a method to control a robotic prosthesis through voice commands. The control method is based on the implementation of an Artificial Neural Network, a widely known technique present in the literature as a standard classifier. The work is divided in two parts: the first is the development of the algorithm that generates features of voice commands, and the second is the application of an algorithm of Artificial Neural Networks for the recognition of voice commands. The words are chosen in a context of assistive robotics, for a robotic prosthesis to directly execute voice commands given by the user, as well as a refining method of movements sent to the prosthesis by other routes, such as electromyography and electroencephalography. The work will be divided into: sample acquisition, signal filtering, RNA training and data validation. The results obtained are satisfactory, showing the efficacy of the method used.

Keywords: Assistive Robotic, Automatic Voice Recognition, Artificial Neural Network.

LISTA DE ILUSTRAÇÕES

Figura 1 - Etapas do RAV	17
Figura 2 - Exemplo de sinal acústico da palavra "Olá"	19
Figura 3 - Exemplo de sinal acústico da palavra "Adeus"	19
Figura 4 - Modelo de um neurônio artificial	22
Figura 5 - Modelo PMC	24
Figura 6 - Resposta ideal do filtro Butterworth.....	27
Figura 7 - Organização das etapas do trabalho.....	28
Figura 8 - Etapas da RNA.....	29
Figura 9 - Fase de Treinamento da RNA.....	30
Figura 10 - Fase de Validação da RNA	33
Figura 11 - Treinamento LPC 3ª ordem	35
Figura 12 - Treinamento LPC 5ª ordem	36
Figura 13 - Treinamento LPC 7ª ordem	37
Figura 14 - Treinamento LPC 10ª ordem	38
Figura 15 - Validação 1 LPC de 3ª ordem.....	40
Figura 16 - Validação 1 LPC de 5ª ordem.....	40
Figura 17 - Validação 1 LPC de 7ª ordem.....	41
Figura 18 - Validação 1 LPC de 10ª ordem.....	41
Figura 20 - Validação 2 LPC de 3ª ordem.....	43
Figura 21 - Validação 2 LPC de 5ª ordem.....	43
Figura 22 - Validação 2 LPC de 7ª ordem.....	44
Figura 23 - Validação 2 LPC de 10ª ordem.....	44

LISTA DE TABELAS

Tabela 4.1 Acertos do treinamento de 3 ^a ordem	35
Tabela 4.2 Acertos do treinamento de 5 ^a ordem	36
Tabela 4.3 Acertos do treinamento de 7 ^a ordem	37
Tabela 4.3 Acertos do treinamento de 10 ^a ordem	38
Tabela 4.5 Resultados quantitativos para a validação 1	42
Tabela 4.6 Resultados quantitativos para a validação 2	45

LISTA DE ABREVIATURAS E SIGLAS

DTW	Dynamic Time Wrapping
EQM	Erro Quadrático Médio
IBGE	Instituto Brasileiro de Geografia e Estatística
LM	Levenberg-Marquardt
LPC	Linear Predictive Coding
MFCC	Mel Frequency Cepstral Coefficient
PMC	Perceptron de Múltiplas Camadas
RASTA	Relative Spectra Processing
RNA	Rede Neural Artificial
RB	Regularização Bayesiana
UFJF	Universidade Federal de Juiz de Fora
ZCPA	Zero crossings with Peak Amplitudes

SUMÁRIO

1 INTRODUÇÃO	14
2 REVISÃO BIBLIOGRÁFICA	16
2.1 Reconhecimento de Voz.....	16
2.2 Redes Neurais Artificiais.....	21
2.3 Filtragem.....	26
2.4 Organização do trabalho.....	28
3 REDES NEURAS ARTIFICIAIS.....	29
3.1 Treinamento.....	29
3.2 Validação.....	32
4 RESULTADOS.....	34
4.1 Resultados da fase de treinamento da RNA	34
4.2 Resultados da fase de validação	39
5 CONCLUSÕES E TRABALHOS FUTUROS.....	46
5.1 Conclusões.....	46
5.2 Trabalhos Futuros	47

1 INTRODUÇÃO

Robótica é a ciência e a técnica da concepção, construção e utilização de robôs. Robôs são dispositivos eletromecânicos capazes de realizar atividades que os humanos não querem, não podem ou não conseguem praticar. Tem uma ampla utilização principalmente na indústria, porém, não se restringem a esse uso, sendo aplicados nas mais diversas áreas como robótica assistiva, robótica de vigilância e robótica de salvamento [1,2,3].

Os robôs são capazes de realizar atividades pré-programadas e tem, necessariamente, a possibilidade de serem reprogramados para outras atividades [4]. Este trabalho terá enfoque em uma das aplicações da robótica assistiva, a qual tem como foco auxiliar pessoas com algum tipo de limitação ou necessidade especial como cadeiras de rodas, próteses biomecânicas, dentre outras que auxiliem as pessoas a recuperarem sua mobilidade.

Atualmente no Brasil, segundo o Censo de 2010 do IBGE, mais de 45 milhões de pessoas possuem algum tipo de deficiência, o que representa mais de 20% da população. Em relação às deficiências motoras, elas podem ser classificadas em três grupos: ‘Deficiência motora – não consegue de modo algum (734.421 pessoas)’; ‘Deficiência motora – grande dificuldade (3.698.929 pessoas)’; ‘Deficiência motora – alguma dificuldade (8.832.249 pessoas)’. Dentro deste grupo, existem pessoas que sofreram algum tipo de amputação devido a acidente, doença ou outros fatores [5].

Além das limitações físicas, as pessoas com algum tipo de dificuldade motora têm que lidar diariamente com diversos tipos de riscos de acidentes de trânsito, acidentes de trabalho, problemas congênitos, além da possibilidade de desenvolvimento de problemas psicológicos. Nos dias atuais, o desenvolvimento de próteses biomecânicas tem sido vastamente pesquisado, principalmente em países como Estados Unidos e Japão.

Até o início dos anos 2000, as próteses eram uma tecnologia cara e inacessível, seu custo poderia chegar a algumas dezenas de milhares de dólares. Porém, ao longo das últimas décadas, outras opções vem surgindo, como as próteses feitas a partir de impressões 3D, o que reduz drasticamente seu custo e torna a tecnologia mais acessível [6].

Ao se criar uma prótese, diversas técnicas podem ser implementadas para facilitar a utilização do usuário. Uma destas técnicas é o reconhecimento automático de voz, que tornou possível ao computador ouvir a comandos da voz humana e interpretá-los. Reconhecimento de voz é o processo de converter um dado sinal de entrada em uma sequência de palavras através de um algoritmo implementado [7].

Ao longo dos anos, diversas técnicas foram desenvolvidas para se realizar o reconhecimento de voz, como modelos baseados em árvores de decisão [8], técnicas baseadas em máxima entropia [9], e, finalmente, redes neurais artificiais, técnica escolhida para ser utilizada neste trabalho, que vem sendo muito utilizada devido sua alta robustez a dados ruidosos, fácil adaptação a novos tipos de dados, inúmeras possibilidades de aplicação como classificação de padrões e previsão de dados, como demonstrado por BENGIO et al. (2003) que aplicou uma RNA em uma sequência de amostras de palavras e demonstrou que a rede poderia aprender qual seria a palavra seguinte [10].

Ao lidar com sinais, é sempre necessário o processamento do sinal e a utilização de filtros. Os sinais, principalmente de áudio, contem grande quantidade de ruídos e frequências indesejadas. Através da filtragem é possível reduzir drasticamente a influência dos ruídos e das frequências irrelevantes. Para tal, existem diversas técnicas e vários tipos de filtros, para diversos casos, pois sua eficiência depende do tipo e intensidade do ruído e da técnica escolhida [11]. BRANDÃO et al. (2005) utiliza 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 [12].

O presente trabalho visa o estudo e a implementação de uma técnica para o funcionamento de uma prótese biomecânica por meio do reconhecimento de voz, utilizando redes neurais artificiais. Por meio desta técnica, pretende-se facilitar e otimizar o uso das próteses, reduzindo, desta forma, as limitações impostas pela deficiência.

A organização do trabalho se dá em cinco capítulos, sendo o primeiro esta introdução. No segundo capítulo é apresentada a revisão bibliográfica descrevendo os temas apontados no trabalho através da literatura já existente. No terceiro capítulo é apresentado o desenvolvimento da rede neural artificial implementada para o problema em questão. No quarto capítulo são mostrados os resultados dos experimentos realizados. Finalmente, o quinto capítulo contém a conclusão do trabalho e possíveis melhorias em trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

O objetivo deste capítulo é apresentar uma revisão de trabalhos acadêmicos já existentes sobre os temas presentes neste trabalho. Serão abordados os assuntos: Reconhecimento de Voz, Redes Neurais Artificiais e Filtragem de Sinais. Ao final do capítulo, será exposto, ainda, uma explicação do problema proposto e como pretende-se resolvê-lo.

2.1 Reconhecimento de Voz

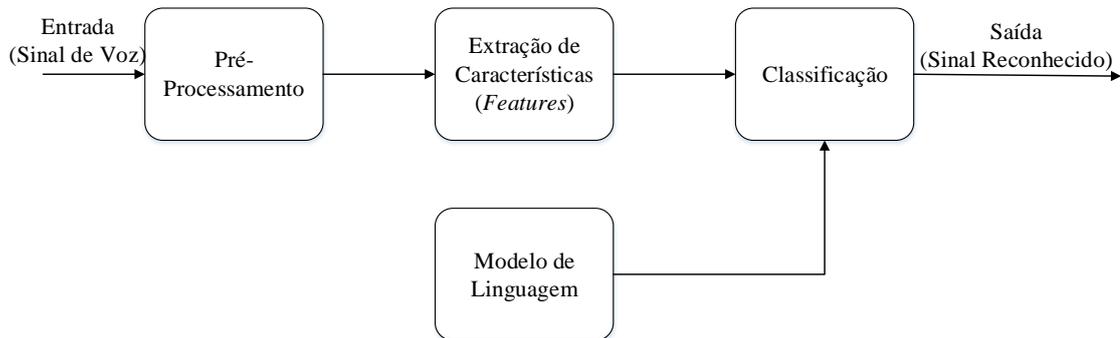
Em 2004, MADAN et al. (2004) [4] discorreu sobre o reconhecimento automático de voz (RAV) e comparou alguns métodos de extração e classificação. Em seu trabalho os autores ressaltam a importância deste tema ao enfatizar que o RAV pode dar suporte a diversas aplicações que requerem a interação homem-máquina.

Sobre a classificação dos sistemas de reconhecimento de voz, a autora propõe quatro categorias baseadas no tipo de fala que o sistema é capaz de reconhecer, sendo elas:

- Palavras Isoladas: são sistemas que aceitam uma palavra por vez e requerem um espaçamento entre cada entrada de áudio.
- Palavras Conectadas: estes sistemas permitem palavras sendo ditas juntas com um mínimo de espaço entre elas.
- Fala Contínua: sistemas permitem quase uma fala natural, são difíceis de serem implementados e requerem um esforço maior para reconhecer os limites entre as palavras.
- Fala espontânea: sistemas de reconhecimento da fala espontânea e natural devem ser capazes de reconhecer diferentes características deste tipo de fala, como palavras que são concatenadas e até ditas de maneira confusa. A fala natural inclui má pronúncia, neologismos, etc.

A figura 1 apresenta um modelo básico de um sistema de reconhecimento de voz, contendo as etapas: pré-processamento, extração das características do sinal, classificação e modelo de linguagem. O pré-processamento adequa o sinal antes de se extrair qualquer informação; a etapa de extração, extrai as informações necessárias para as etapas de modelagem e classificação; a etapa de classificação utiliza as características extraídas do sinal e compara com o modelo de linguagem proposto.

Figura 1 - Etapas do RAV



Fonte: MADAN et al. (2004) [4]

Em relação a extração das características do sinal, a autora propõe duas técnicas amplamente utilizadas, e, resalta ainda que as características devem ser medidas facilmente, consistentes com o tempo e resistentes à ruídos e ao meio. As duas técnicas propostas no trabalho são: *Linear Predictive Coding (LPC)* e *Mel Frequency Cepstral Coefficients (MFCC)*.

A primeira técnica tem como principal objetivo a aproximação de uma dada amostra de voz por uma combinação linear de amostras anteriores, comportando-se como um filtro de resposta finita ao impulso (*Finite Impulse Response, FIR*), este método será melhor explicado mais à frente. Já a segunda técnica é uma conversão das frequências do sinal para unidades *mel* (do inglês *melody*), as quais destacam a resposta do ouvido humano aos sons, onde as frequências são interpretadas de maneira não linear através do espectro de áudio. Desta forma, o objetivo desta técnica é distorcer a frequência de forma que ela obedeça a relação espacial da distribuição das células ciliares da orelha humana, facilitando a análise espectral para sons emitidos por humanos.

Por fim, a autora discorre sobre os sistemas de classificação para reconhecimento de voz. Ela destaca as vantagens e desvantagens de três deles:

- Modelos escondidos de Markov: modelam a distribuição temporal de sinais de voz, são simples de se implementar, podem modelar sistemas discretos ou contínuos e suportam entradas de tamanhos variados, porém, assumem que, provavelmente, para um determinado estado existir ele depende do estado anterior.
- Redes Neurais Artificiais: possuem habilidade de auto-organização e auto-aprendizado, são robustas a ruídos e distorções, e facilmente ajustáveis a novos ambientes, adequadas para reconhecimento de padrões. Sua desvantagem pode

decorrer da necessidade de treinamento de conjuntos massivos de dados, o que pode vir a ser um processo moroso.

- Máquina Vetor de Suporte: não possui problemas com mínimos locais e super treinamento, capaz de lidar com vetores de entrada de altas dimensões. Suas desvantagens são a necessidade de uma entrada com tamanho fixo, o alto custo computacional à medida que o número de classes cresce.

O trabalho acima explicado fornece uma visão abrangente e boas comparações entre as técnicas mais utilizadas no RAV.

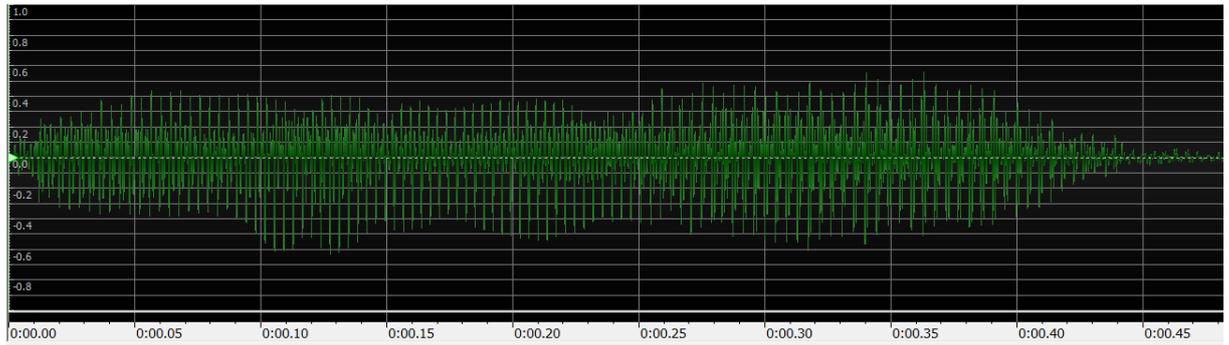
Além das duas técnicas de extração já citadas anteriormente, KURZEKAR et al. (2014) [13], compara outras três técnicas que podem ser utilizadas para caracterizar um sinal:

- *Dynamic Time Wrapping* (DTW): seu algoritmo é baseado em programação dinâmica. Ele é utilizado para medir a semelhança entre duas séries temporais, que podem variar no tempo ou na velocidade. Uma das séries é alongada ou encolhida ao longo do eixo do tempo e, através da envoltória não linear, é possível determinar regiões correspondentes e semelhantes entre as duas séries.
- *Relative Spectra Processing* (RASTA): esta técnica se baseia na filtragem linear dos ruídos do sinal. Os valores espectrais do sinal de entrada são comprimidos, filtrados e então expandidos novamente.
- *Zero Crossings with Peak Amplitudes* (ZCPA): método similar a MFCC, pois também é baseado no sistema auditivo humano. Ele utiliza um intervalo do sinal entre dois cruzamentos de zero nas amplitudes para representar a informação de frequência do sinal e o valor de amplitude para representar a informação de intensidade, estas duas informações são combinadas para fornecer a saída.

Por meio de comparação entre as taxas de eficácia de cada técnica, o autor conclui que LPC, que é uma das técnicas mais utilizadas, é a que fornece os parâmetros de fala mais acurados. Ele demonstra, ainda, que testes realizados com palavras em inglês fornecem taxas mais altas de acerto (acima de 90%) quando comparado com outros idiomas, como espanhol e chinês, por exemplo (entre 80-90%).

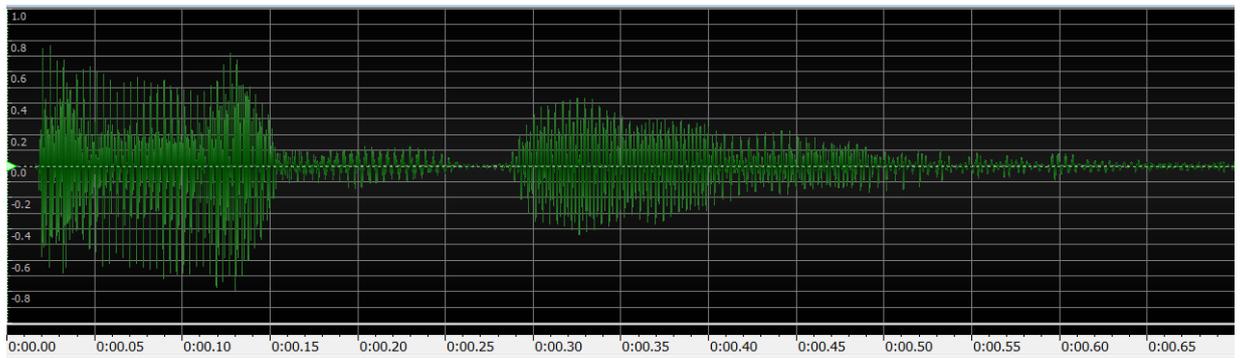
Além das técnicas aqui já apresentadas, existe, ainda, outra forma de caracterização de sinais, por meio de seus parâmetros medidos. Do ponto de vista matemático, a voz é um sinal acústico. Um sinal contém informações sobre um fenômeno ou sobre um acontecimento e é uma função de uma ou mais variáveis. Nas figuras 3 e 4 pode-se observar a diferença entre os sinais gerados ao se pronunciar “Olá” e “Adeus”, respectivamente.

Figura 2 - Exemplo de sinal acústico da palavra "Olá"



Fonte: Autora

Figura 3 - Exemplo de sinal acústico da palavra "Adeus"



Fonte: Autora

Cada sinal possui diferentes características de acordo com o tipo de informação que carrega. São muitos os parâmetros que podem ser calculados e utilizados para se classificar um determinado sinal. Para este trabalho, foram selecionados os mais comumente utilizados. Seja um sinal aleatório $x(t)$ que possui n termos e varia em função do tempo. Calculam-se [14,15,16,17]:

- Média: é o valor médio de todos os seus termos. É dado pela soma de cada um deles individualmente dividido pelo número total de termos.

$$X = \frac{\sum_{i=0}^n x_i(t)}{n} \quad (1)$$

- Desvio Padrão: o desvio padrão representa o quão distante uma distribuição está de ser uniforme. Ele é calculado baseado na diferença entre cada termo da amostra e a sua média aritmética. É dado por:

$$\sigma = \sqrt{\frac{\sum_{i=0}^n (x_i(t) - X)^2}{n - 1}} \quad (2)$$

- RMS: é o valor quadrático médio (*root mean square*) de um sinal aleatório e é dado por:

$$V_{ef} = \sqrt{\frac{\sum_{i=0}^n x_i(t)^2}{n}} \quad (3)$$

- Energia: a energia de um sinal descrito no tempo é definida como:

$$E = \int_{-\infty}^{\infty} |x(t)|^2 dt \quad (4)$$

- Potência: a potência de um sinal descrito no tempo é definida como:

$$P = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 dt \quad (5)$$

- Frequência mais baixa: é a frequência mais baixa presente no sinal, dada em Hz.
- Frequência mais alta: é a frequência mais alta presente no sinal, dada em Hz.
- Banda Total Ocupada: é a largura de banda que contem 99% da potência total integrada do espectro transmitido, centrada na frequência do canal atribuído.

Apesar destes parâmetros serem uma boa forma de caracterização, será demonstrado nos próximos capítulos que para o problema em questão é necessária a utilização de uma técnica mais robusta. Para tal, foi escolhida a técnica LPC que já foi citada e, em português, pode ser traduzida como Predição de Coeficientes Lineares. Segundo KESARKAR et al. (2003) [27], as amostras de fala podem ser aproximadas como combinações lineares das amostras passadas através da minimização da soma dos quadrados das diferenças (em um intervalo finito) entre a amostra atual e as preditas linearmente. Por meio deste método, os coeficientes são determinados e o sinal é modelado como a saída de um sistema linear e variante no tempo. A predição linear é um método que fornece robustez e confiança por ser acurado ao estimar os parâmetros que caracterizam o sistema linear variante representando o trato vocal.

SHRAWANKAR et al. [26] descreve a metodologia da técnica como semelhante à da voz humana. A minimização da soma do quadrado das diferenças entre o sinal original e o sinal estimado em um tempo finito é usado para gerar um conjunto de coeficientes. Estes coeficientes preditores são estimados a cada *frame*, que duram normalmente 20 ms. A função de transferência do filtro digital variante no tempo é dada pela equação 6, onde G é o ganho, a_k são os coeficientes preditores e p é a ordem definida para o LPC.

$$H(Z) = \frac{G}{1 - \sum_{k=1}^p a_k Z^{-k}} \quad (6)$$

Seja a amostra atual x_i do sinal de áudio estimada a partir das p amostras passadas, o seu valor previsto será [13]:

$$x'_i = -a_2x_{i-1} - a_3x_{i-2} - \dots - a_{p+1}x_{i-p} \quad (7)$$

em que x'_i é a amostra prevista e $\{1, a_2, \dots, a_{p+1}\}$ são os $(p + 1)$ coeficientes.

2.2 Redes Neurais Artificiais

2.2.1 Introdução

Uma Rede Neural Artificial se baseia em técnicas computacionais para implementar um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento por meio de experiência [18]. Como descrito pelo Professor Ivan Nunes da Silva em seu livro “Redes Neurais Artificiais – Curso Prático” [19], em 1949 foi proposto por Hebb o primeiro método de treinamento para redes neurais artificiais, que se denominou “regra de aprendizado de Hebb”, sendo esta baseada em hipóteses e observações de caráter fisiológicos.

Já nos anos 50, dentre diversos trabalhos desenvolvidos, o trabalho de Frank Rosenblatt se destaca ao desenvolver o primeiro neurocomputador e idealizar o modelo básico do Perceptron. Este modelo despertou interesse devido sua capacidade de reconhecer padrões simples. Em 1960, Widrow e Hoff desenvolveram a rede Adaline (*Adaptive Linear Element*) e, posteriormente, propôs-se a Madaline, a Adaline múltipla, também conhecida como algoritmo de aprendizado pelo Método dos Mínimos Quadrados (*Least Mean Square*, LMS).

Diante destes trabalhos pioneiros, muitos pesquisadores se interessaram em pesquisar sobre este assunto. Porém, em 1969, Minsky e Papert demonstram em seu livro a limitação das redes neurais artificiais constituídas de apenas uma camada (Perceptron e Adaline) em resolverem problemas de classificação como a simples função lógica Ou Exclusivo (*Exclusive Or*, XOR). Com o XOR eles demonstraram que os neurônios artificiais não eram capazes de solucionar problemas que não fossem linearmente separáveis.

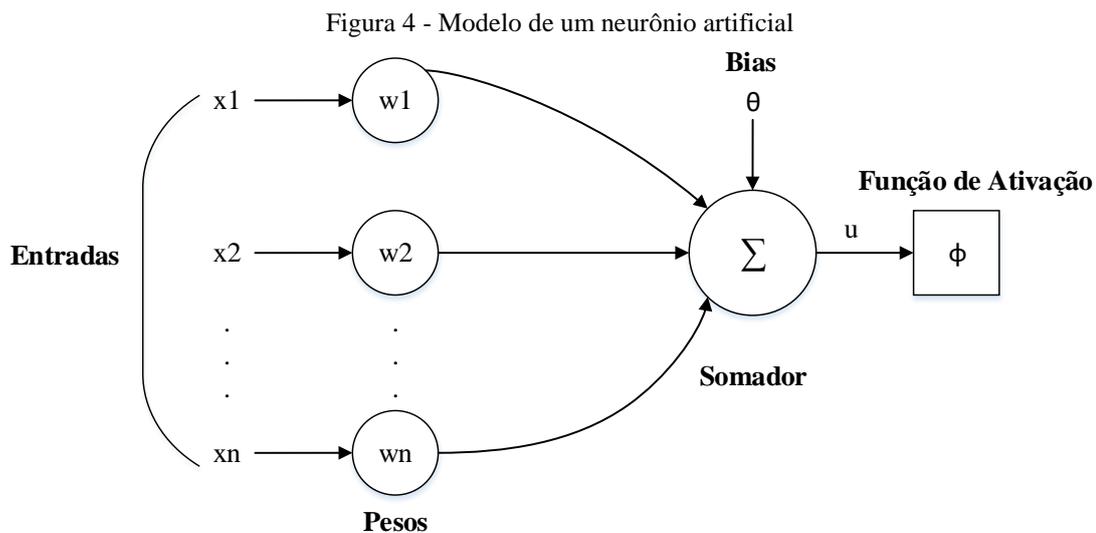
Devido a esta publicação, houve um período com pouquíssimas pesquisas sendo realizadas na área. Somente em 1989, Rumelhart, Hilton e Williams publicam um livro intitulado “*Parallel Distributed Processing*”, onde eles desenvolvem um algoritmo, chamado *Backpropagation*, que possibilita o ajuste de pesos em uma rede com mais de uma camada e soluciona o problema do XOR.

Nos últimos anos, as redes neurais tem sido aplicadas em diversas áreas do conhecimento e contribuições tem sido feitas para aumentar a eficiência e abrangência das redes neurais artificiais.

2.2.2 O Neurônio Artificial

Como já dito anteriormente, as RNAs foram desenvolvidas a partir de modelos conhecidos de sistemas biológicos e do cérebro humano. Estes modelos se baseiam na análise da geração e propagação de impulsos elétricos pela membrana celular dos neurônios, como descrito em 1952 por Hodgkin e Huxley.

Os neurônios artificiais são não-lineares, fornecem saídas contínuas e produzem uma resposta de acordo com as entradas coletadas dos sinais externos, esta resposta é dependente da sua função de ativação. O modelo de um neurônio artificial pode ser visualizado na figura 4.



Fonte: GSIGMA [20]

O modelo do neurônio artificial tem equivalência com o neurônio biológico na sua forma de funcionamento. As ponderações exercidas pelas junções sinápticas do modelo biológico equivalem ao conjunto de pesos sinápticos $\{\omega_1, \omega_2, \dots, \omega_n\}$. De forma semelhante, uma entrada é executada por meio de sua multiplicação pelos respectivos pesos sinápticos, ponderando-se todas as informações externas que chegam ao neurônio. Assim define-se como u , a saída do corpo celular artificial, que é a soma ponderada de suas entradas.

Pode-se observar na Figura 4 os sete elementos básicos de um neurônio artificial, sendo eles:

- 1) Sinais de entrada $\{x_1, x_2, \dots, x_n\}$: São os sinais ou medidas advindas do mundo externo e representam os valores das variáveis de determinada aplicação. Normalmente são normalizados visando aumentar a eficiência computacional dos algoritmos.
- 2) Pesos Sinápticos $\{\omega_1, \omega_2, \dots, \omega_n\}$: São os valores que irão ponderar as entradas da rede, permitindo-se qualificar a relevância de cada uma delas em relação à funcionalidade do respectivo neurônio. O conhecimento da rede está, de fato, nos pesos sinápticos.
- 3) Combinador Linear (Σ): Tem a função de agregar todos os sinais de entrada, já ponderados, a fim de produzir um valor de potencial de ativação.
- 4) Limiar de Ativação (bias - θ): É o valor que especifica qual será o patamar apropriado para que o valor gerado pelo combinador linear possa gerar um valor de disparo para a saída do neurônio.
- 5) Potencial de ativação $\{u\}$: É o resultado produzido pela diferença entre o valor do combinador linear e do limiar de ativação.
- 6) Função de ativação $\{\phi\}$: Tem a função de limitar a saída do neurônio dentro de um intervalo de valores.
- 7) Sinal de saída $\{y\}$: É o valor final produzido por um neurônio dado um conjunto de sinais de entradas. Este valor pode ser usado como entrada para outros neurônios que estiverem conectados à ele.

Segundo McCulloch e Pitts o resultado produzido por um neurônio pode ser resumido por:

$$u = \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n - \theta = \sum_{i=1}^n \omega_i x_i - \theta \quad (6)$$

$$y = g(u) \quad (7)$$

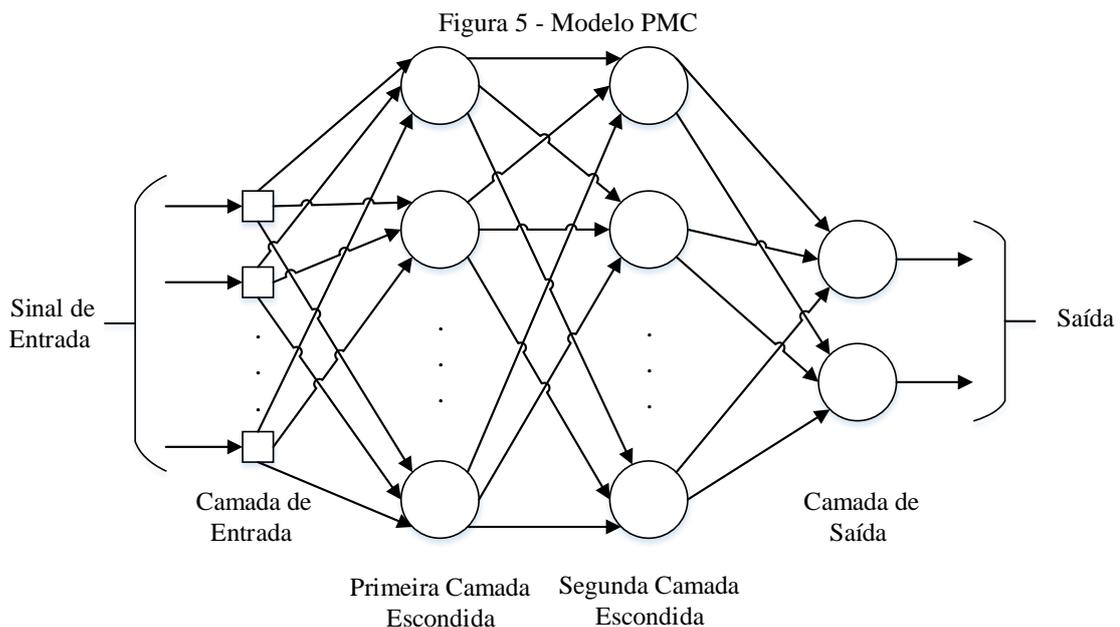
2.2.3 Redes Perceptron Multicamadas

As redes Perceptron de Múltiplas Camadas (PMC), em inglês *Multilayer Perceptron* (MLP), caracterizam-se pela presença de, pelo menos, uma camada intermediária (ou, escondida) entre a cama de entrada e a cama de saída. Elas podem ser utilizadas nas mais diversas áreas do conhecimento com as mais variadas funções, como aproximação universal de funções, reconhecimento de padrões, previsão de séries temporais, etc.

Na figura 5, está representado um modelo de uma rede com duas camadas escondidas, o número de camadas e neurônios em cada uma delas deve ser decidido pelo projetista de acordo com a aplicação. Uma das diferenças da PMC para as redes Perceptron e Adaline, é a possibilidade de mais de um neurônio na saída, aumentando, assim, as aplicações possíveis.

O treinamento da rede se dá por um método conhecido como *backpropagation*. Este método é composto pelas fases *forward* e *backward*. A primeira fase aplicada é a “propagação adiante” (*forward*), onde as entradas inseridas na rede se propagam entre as camadas, uma a uma, até a produção das respectivas saídas, portanto a função desta fase é gerar uma resposta considerando as entradas e os respectivos pesos sinápticos, os quais permanecem inalterados.

Em seguida, as respostas geradas são comparadas com as respostas desejadas, pois este é um processo supervisionado de aprendizado. De acordo com os erros obtidos, é aplicada a fase de “propagação reversa” (*backward*), onde os pesos sinápticos são ajustados. Desta forma, diminui-se, a cada iteração a soma dos erros entre a resposta obtida e a resposta gerada.



Fonte: RESEARCHGATE [21]

2.2.4 Redes Neurais Artificiais aplicadas ao Reconhecimento de Voz

Como já mencionado anteriormente, as redes neurais podem ser aplicadas nas mais diversas áreas de conhecimento e, o reconhecimento automático de voz é umas dessas áreas que tem despertado interesse dos pesquisadores. Em 1995, TEBELSKIS et al. (1995) [22] demonstrou que devido a capacidade de aprender funções complexas, a tolerância a ruídos e por suportar o paralelismo, as redes neurais são uma boa opção quando usadas cuidadosamente para o RAV. Ele demonstrou ainda que elas possuem maior eficácia quando comparadas com outros métodos, como o Modelo Escondido de Markov, por exemplo.

Em 2005, BRANDÃO et al. (2005) [12], aplica dois métodos de RNA ao reconhecimento de voz, O método de aceleração de treinamento de Levenberg-Marquardt (LM) e de Regularização Bayesiana (RB). O primeiro algoritmo apresenta vantagem em sua agilidade ao convergir para o resultado, porém, requer um alto esforço computacional na fase de treinamento. O segundo algoritmo é uma modificação do primeiro, e busca melhorar a desvantagem citada do LM.

O autor realiza o pré-processamento dos sinais e utiliza o método de LPC para gerar as entradas da rede. Os resultados obtidos são satisfatórios (erro máximo de 3% para LM e 2,67% para RB) e foi concluído ainda que existe um valor ótimo para os parâmetros da rede (número de camadas escondidas e número de neurônios por camada), antes ou após este valor, a rede apresenta resultados menos acurados.

Durante o trabalho descrito, muita ênfase foi dada à parte de processamento do sinal, que é, de fato, muito importante, porém uma ênfase maior poderia ter sido dada à etapa de classificação, para a obtenção de resultados mais acurados e esclarecedores.

Em 2016, GUPTA et al. (2016) [23], demonstra a robustez das redes neurais ao submeter uma rede a dados com relação sinal-ruído de -3 à -5dB. O autor compara alguns algoritmos de redes neurais para diferentes níveis de ruído e obtém, em algumas situações, resultados excelentes com acurácia superior a 99%.

Com base na literatura já existente, as redes neurais foram escolhidas para este trabalho devido às várias vantagens citadas. Como rede, escolheu-se o algoritmo de PMC devido à sua versatilidade de implementação e expansão da rede, o qual será implementado com técnicas de aceleração de treinamento a serem descritas em momento futuro neste trabalho.

2.3 Filtragem

Ao se lidar com sinais, sejam eles digitais ou analógicos, é de suma importância a filtragem do sinal para a redução da influência de ruídos sobre o sinal de voz. Os sinais podem conter ruídos oriundos do aparelho de medição, do ambiente de origem, do gerador do sinal, dentre outros. Ao se desenvolver uma aplicação com um sinal, estes ruídos interferem negativamente, pois acrescentam frequências indesejadas ao espectro do sinal, por isso é necessário eliminar, ou ao menos amenizar a influência destas frequências. Denomina-se este processo por filtragem.

Filtros são sistemas lineares invariantes no tempo capazes de modificar as características dos sinais conectados em sua entrada, de modo que, apenas uma parcela específica dos componentes de frequência do sinal chega à saída do filtro [24]. Os filtros podem ser divididos em filtros digitais e filtros analógicos. Neste trabalho, o enfoque será nos filtros digitais, devido à natureza dos sinais que serão filtrados.

O filtro digital é um algoritmo implementado em *hardware* ou *software* que opera sobre um sinal $x[n]$ presente na sua entrada e gera um sinal $y[n]$ filtrado na saída, pode ser classificado de acordo com o comprimento da sua sequência de resposta ao impulso como:

- Filtro de Resposta Finita ao Impulso (*Finite Impulse Response*, FIR), cuja saída é dada por:

$$y[n] = \sum_{k=0}^{N-1} h[k] * x[n - k] \quad (8)$$

- Filtro de Resposta Infinita ao Impulso (*Infinite Impulse Response*, IIR), cuja saída é dada por:

$$y[n] = \sum_{k=0}^{\infty} h[k] * x[n - k] \quad (9)$$

Porém, como esta é uma resposta impossível de ser implementada na prática, é usada uma implementação recursiva da forma [24]:

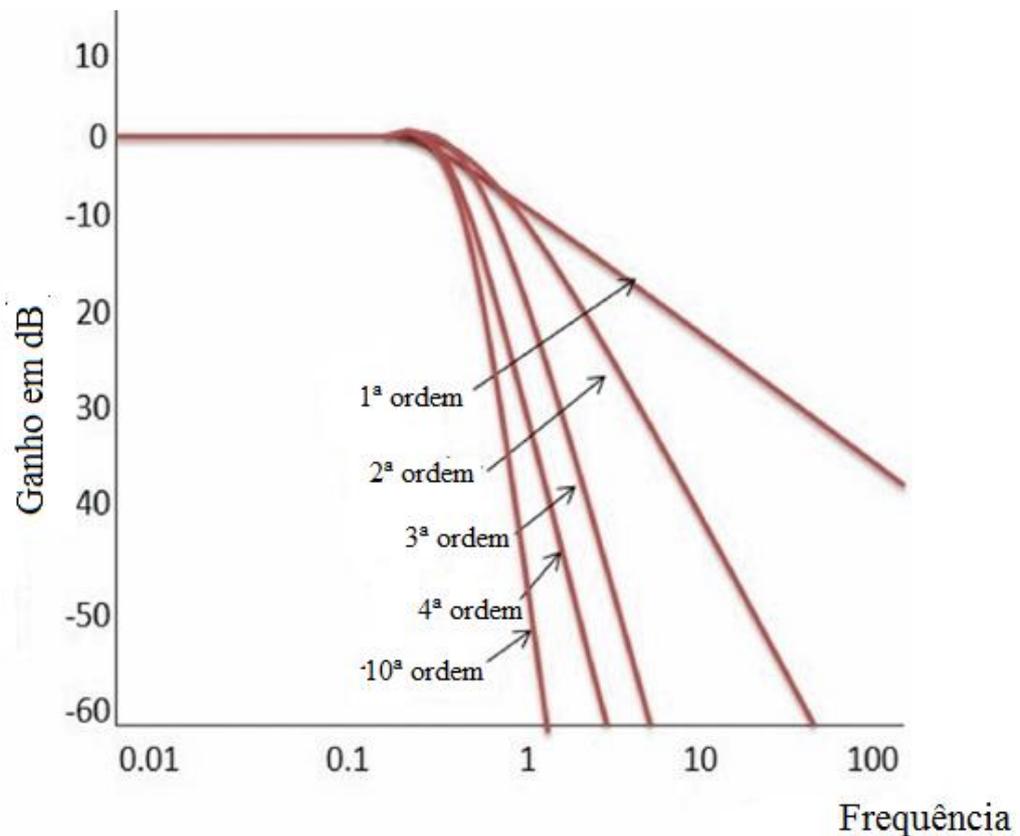
$$y[n] = \sum_{k=0}^N b_k * x[n - k] - \sum_{k=0}^M a_k * y[n - k] \quad (10)$$

A principal diferença entre estes dois grupos de filtros é que o FIR produz uma resposta sem distorção na fase do sinal e são sempre estáveis, pois não são recursivos. Já os IIR geram uma resposta filtrada defasada em relação ao sinal filtrado, porém necessitam de menos coeficientes para atender a mesma aplicação. A escolha do tipo de filtro a ser utilizado depende do tipo de aplicação e projeto.

Existem diversos tipos de filtros digitais que podem ser implementados, entre os mais utilizados está o Butterworth que realiza operações específicas no sinal original e modifica o seu conteúdo. Ele é descrito pela sua ordem e frequência de corte e pode ser implementado como passa-baixa, passa-alta, passa-faixa ou rejeita-faixa, como descrito por SANDRU et al. (2016) [25]. Suas variações de resposta estão mostradas na figura 6.

O software Matlab possui uma toolbox, utilizada neste trabalho, que contém as funções dos filtros acima citados implementadas.

Figura 6 - Resposta ideal do filtro Butterworth



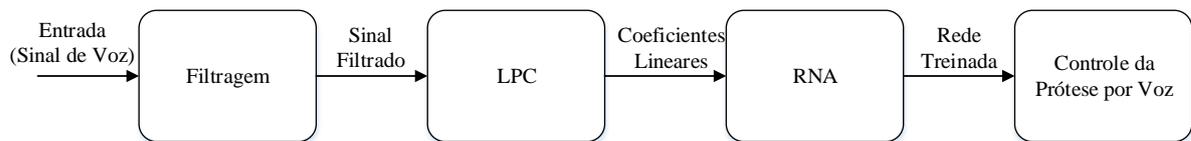
Fonte: SANDRU (2016) [25]

2.4 Organização do trabalho

Diante da teoria apresentada, com este trabalho pretende-se coletar amostras de palavras pré-determinadas de acordo com o contexto do trabalho, filtrá-las para amenizar os ruídos de modo que o sinal filtrado não seja defasado em relação ao sinal original, pois esta é uma característica importante em sinais variantes no tempo. Após a filtragem, será utilizado o método LPC para cálculo dos coeficientes lineares e caracterização de cada amostra, estes coeficientes serão utilizados como entrada para a rede neural que, após o treinamento, será capaz de classificar cada amostra em seu respectivo grupo.

Por fim, deseja-se utilizar uma prótese já existente para teste do algoritmo implementado, à cada palavra treinada na RNA, será associado um movimento da prótese e, desta forma, poderá ser realizado o controle da prótese através do comando de voz. O esquema de todas as etapas descritas pode ser observado na figura abaixo:

Figura 7 - Organização das etapas do trabalho

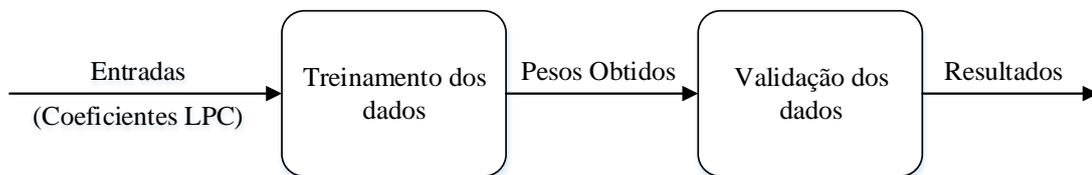


Fonte: Autora

3 REDES NEURAIAS ARTIFICIAIS

Neste capítulo serão apresentados os detalhes da RNA implementada para o reconhecimento de voz. Este processo pode ser dividido em duas partes: o treinamento da rede e a validação dos dados utilizando os pesos obtidos na primeira etapa, conforme pode ser observado na figura 8.

Figura 8 - Etapas da RNA



Fonte: Autora

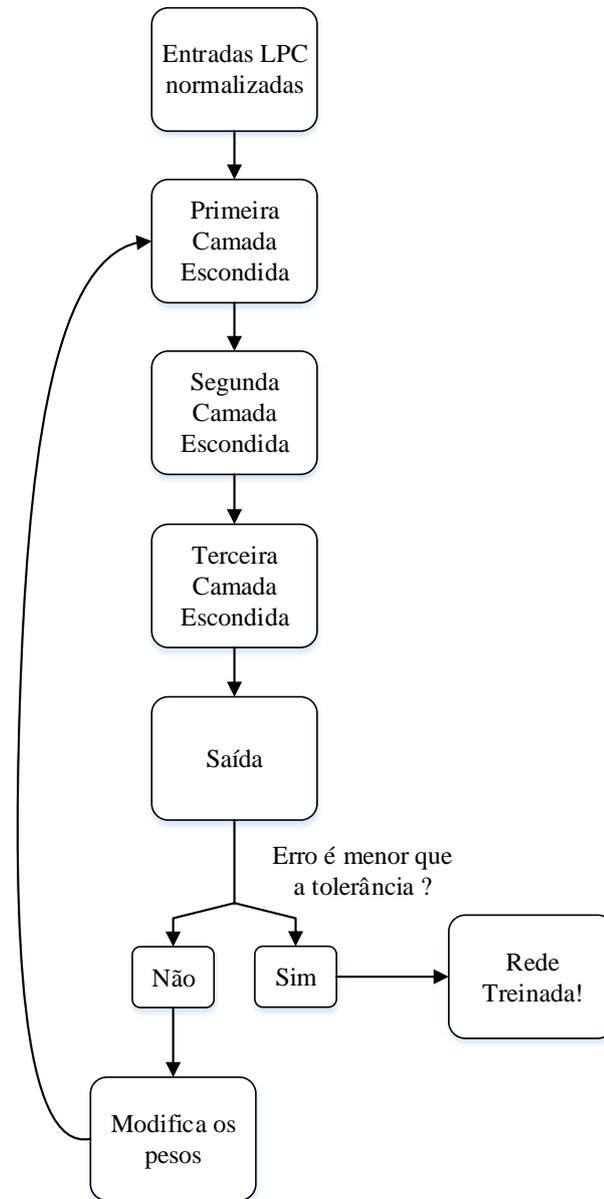
3.1 Treinamento

Nesta seção será detalhada a primeira etapa após o processamento do sinal de áudio, que é o treinamento dos dados. Nesta etapa, os coeficientes encontrados através do método LPC são utilizados como entrada da rede e o objetivo é que ela aprenda a reconhecer o padrão nos coeficientes de cada grupo de palavras e saiba separar estes grupos.

Um esquema do processo de treinamento está retratado na Figura 9. O primeiro passo é a normalização das entradas para que estas estejam dentro do intervalo desejado. E então, os parâmetros principais da rede podem ser decididos com base em experimentação do projetista. Dois parâmetros importantes são o número de camadas escondidas e o número de neurônios em cada camada.

Para o caso específico deste trabalho, não foi obtido sucesso com uma ou duas camadas escondidas, pois a rede não convergiu, portanto foram necessárias três camadas escondidas. O número de neurônios também foi estabelecido, por meio de experimentação, como 20, 10, 10 e 2 para cada uma das respectivas camadas, sendo que o número de neurônios na saída é determinado pelo número de grupos que se deseja classificar, é possível separar dois grupos com cada neurônio.

Figura 9 - Fase de Treinamento da RNA



Fonte: Autora

Com os parâmetros devidamente ajustados, inicia-se a fase de *backpropagation*, onde a rede será treinada de fato. Os pesos de cada camada são inicializados aleatoriamente e uma parcela (80%) dos dados de entrada são selecionados também de forma aleatória para o treinamento da rede.

Com os pesos existentes, obtidos de maneira aleatória, calcula-se a saída de rede (fase *forward*) e esta é comparada com o valor final desejado, se o erro entre elas for menor do que o estipulado, significa que a rede convergiu, senão, os pesos devem ser ajustados (fase *backward*) e o processo reiniciado.

O método utilizado para o ajuste dos pesos, é o do gradiente descendente. O gradiente é um vetor matemático que aponta na direção de máximo crescimento de uma função. Desta maneira, pode-se aplicá-lo à função quadrática de erro e ajustar os pesos da Rede Neural na direção oposta à do gradiente, garantindo, assim, que o erro seja ajustado para o seu valor mínimo (o mais próximo possível de zero). Outra ferramenta utilizada no ajuste dos pesos é o “momentum”, método que se baseia na teoria de momento da física com o objetivo de aumentar a rapidez com a qual o erro evolui para seu valor mínimo.

O equacionamento utilizado para cálculo e ajuste dos pesos será detalhado a seguir. Sejam as entradas x da rede, o limiar de ativação x_0 e os pesos w_1 e w_{01} da primeira camada, calculam-se o potencial de ativação (u_1) e a saída da primeira camada (y_1) como nas equações 11 e 12, respectivamente.

$$u_1 = w_1 * x + w_{01} * x_0 \quad (11)$$

$$y_1 = tgh(incl * u_1) \quad (12)$$

A função de ativação usada foi a tangente hiperbólica, e sua inclinação é modificada pelo parâmetro “*incl*”. Calculada a saída da primeira camada, esta é utilizada como entrada da segunda camada, e a saída da segunda como entrada da terceira, portanto, obtém-se:

$$u_2 = w_2 * y_1 + w_{02} * x_0 \quad (13)$$

$$y_2 = tgh(incl * u_2) \quad (14)$$

$$u_3 = w_3 * y_2 + w_{03} * x_0 \quad (15)$$

$$y_3 = tgh(incl * u_3) \quad (16)$$

Logo, a saída final da rede é dada por:

$$u_4 = w_4 * y_3 + w_{04} * x_0 \quad (17)$$

$$y_4 = tgh(incl * u_4) \quad (18)$$

Como já mencionado anteriormente, o ajuste dos pesos é feito por meio do método do gradiente descendente, estes ajustes estão descritos a seguir. O processo é realizado de trás para frente (fase *backward*), portanto, começa-se o ajuste nos pesos da última camada. Primeiramente define-se o erro quadrático médio (EQM) como:

$$EQM = \frac{1}{2} (d - y_4)^2 \quad (19)$$

Em que d é a saída desejada e y_4 é a saída da rede.

Então:

$$w_4 = w_4 + \alpha * (w_{t-1}^4 - w_{t-2}^4) - n * \frac{dEQM}{dw_4} \quad (20)$$

$$w_{04} = w_{04} + \alpha * (w_{t-1}^{04} - w_{t-2}^{04}) - n * \frac{dEQM}{dw_{04}} \quad (21)$$

Em que w_{t-1}^4 , w_{t-2}^4 , w_{t-1}^{04} , w_{t-2}^{04} e α são parâmetros relativos ao *momentum* e n é a taxa de aprendizado.

O mesmo pode ser feito para as demais camadas:

$$w_3 = w_3 + \alpha * (w_{t-1}^3 - w_{t-2}^3) - n * \frac{dEQM}{dw_3} \quad (22)$$

$$w_{03} = w_{03} + \alpha * (w_{t-1}^{03} - w_{t-2}^{03}) - n * \frac{dEQM}{dw_{03}} \quad (23)$$

$$w_2 = w_2 + \alpha * (w_{t-1}^2 - w_{t-2}^2) - n * \frac{dEQM}{dw_2} \quad (24)$$

$$w_{02} = w_{02} + \alpha * (w_{t-1}^{02} - w_{t-2}^{02}) - n * \frac{dEQM}{dw_{02}} \quad (25)$$

$$w_1 = w_1 + \alpha * (w_{t-1}^1 - w_{t-2}^1) - n * \frac{dEQM}{dw_1} \quad (26)$$

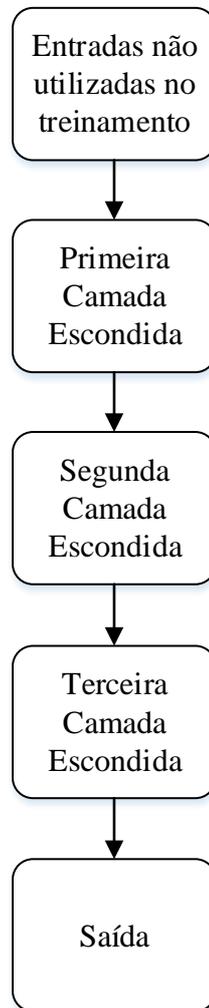
$$w_{01} = w_{01} + \alpha * (w_{t-1}^{01} - w_{t-2}^{01}) - n * \frac{dEQM}{dw_{01}} \quad (27)$$

3.2 Validação

O objetivo desta seção é detalhar a segunda etapa do funcionamento da RNA para o RAV. Terminada a primeira etapa de treinamento da rede, é necessária uma segunda etapa onde os pesos obtidos na primeira serão verificados. Este processo consiste basicamente em aplicar a fase *forward* às entradas que não foram utilizadas no treinamento e submetê-las aos pesos obtidos, o que é feito por meio das equações 11 à 18.

A validação é a etapa em que se verifica se a rede generaliza adequadamente dados que não pertenceram à etapa de treinamento. Para tanto, conforme citado anteriormente, 20% dos dados disponíveis são separados, por meio de sorteio aleatório, para esta etapa. Eles não são usados no treinamento, mas pertencem ao grupo de dados originais. A esquematização desta etapa está representada na figura 10.

Figura 10 - Fase de Validação da RNA



Fonte: Autora

4 RESULTADOS

Este capítulo tem o objetivo de relatar os resultados obtidos no treinamento e experimentos da RNA e, ainda, demonstrar a eficácia do método utilizado.

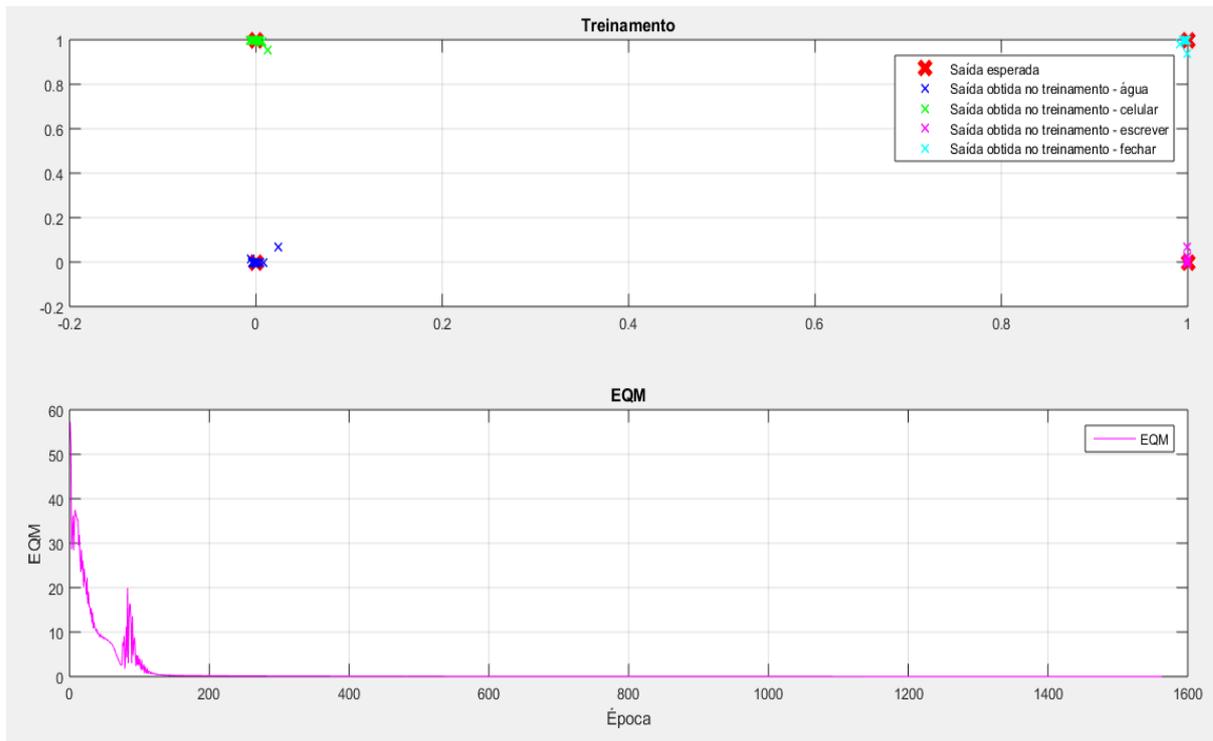
4.1 Resultados da fase de treinamento da RNA

Nesta seção estão os resultados obtidos durante a fase de treinamento. Primeiramente, o treinamento da rede é realizado para diferentes tipos de entrada. É importante ressaltar que, para o caso específico deste trabalho, o método LPC demonstrou alta robustez aos ruídos, como será demonstrado, e, além disso, a etapa de filtragem atenuou frequências importantes utilizadas para o cálculo dos coeficientes, o que não resultou em uma boa resposta. Portanto, as amostras utilizadas no cálculo dos parâmetros de entrada não são filtradas.

Para a etapa de treinamento, foram feitos testes com LPC de terceira, quinta, sétima e décima ordem. Acima dessas ordens os resultados já não demonstravam um ganho significativo que justificasse o aumento da complexidade computacional envolvida em adotá-los. As palavras utilizadas foram “água”, “celular”, “escrever” e “fechar”, todas em língua portuguesa no contexto da robótica assistiva aplicada para manipuladores e próteses.

Os resultados de classificação de cada palavra durante o treinamento para cada um destes testes, assim como a evolução do erro, podem ser vistos nas figuras 11 a 14. Nas tabelas 4.1 a 4.4 pode-se observar os valores de acerto de classificação de cada palavra e, ainda, o erro quadrático médio associado a cada uma delas.

Figura 11 - Treinamento LPC 3ª ordem



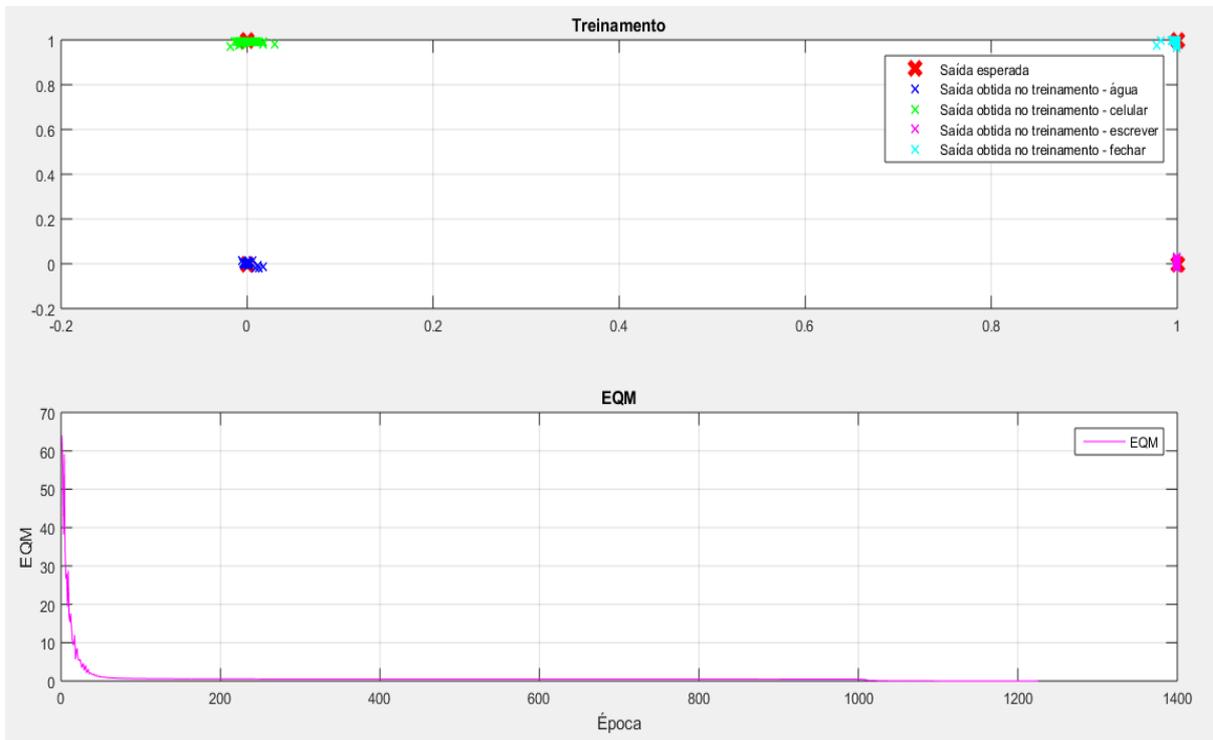
Fonte: Autora

Tabela 4.1 Acertos do treinamento de 3ª ordem

Treinamento LPC 3ª ordem		
Palavra	Acerto	EQM
Água	100%	0,0018
Celular	100%	0,0011
Escrever	100%	0,0044
Fechar	100%	0,0027
Acerto Total	100%	0,01

Fonte: Autora

Figura 12 - Treinamento LPC 5ª ordem



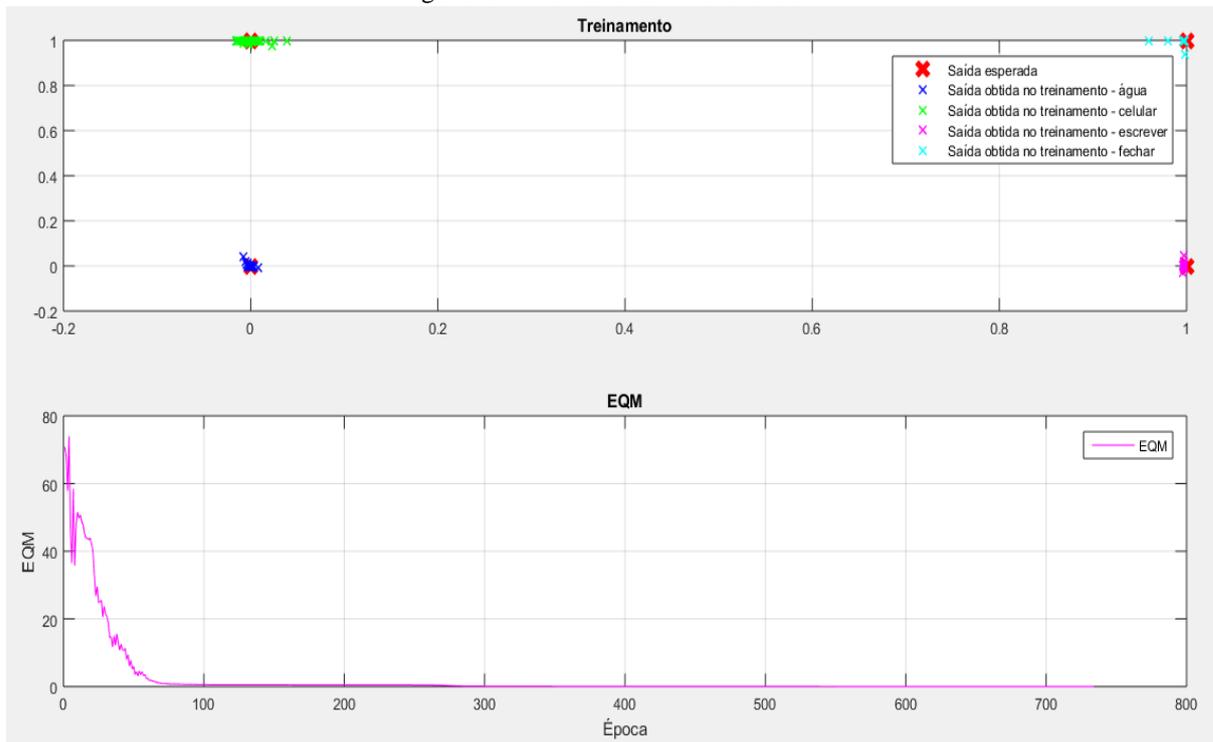
Fonte: Autora

Tabela 4.2 Acertos do treinamento de 5ª ordem

Treinamento LPC 5ª ordem		
Palavra	Acerto	EQM
<i>Água</i>	100%	0,0027
<i>Celular</i>	100%	0,0017
<i>Escrever</i>	100%	0,0036
<i>Fechar</i>	100%	0,0020
Acerto Total	100%	0,01

Fonte: Autora

Figura 13 - Treinamento LPC 7ª ordem



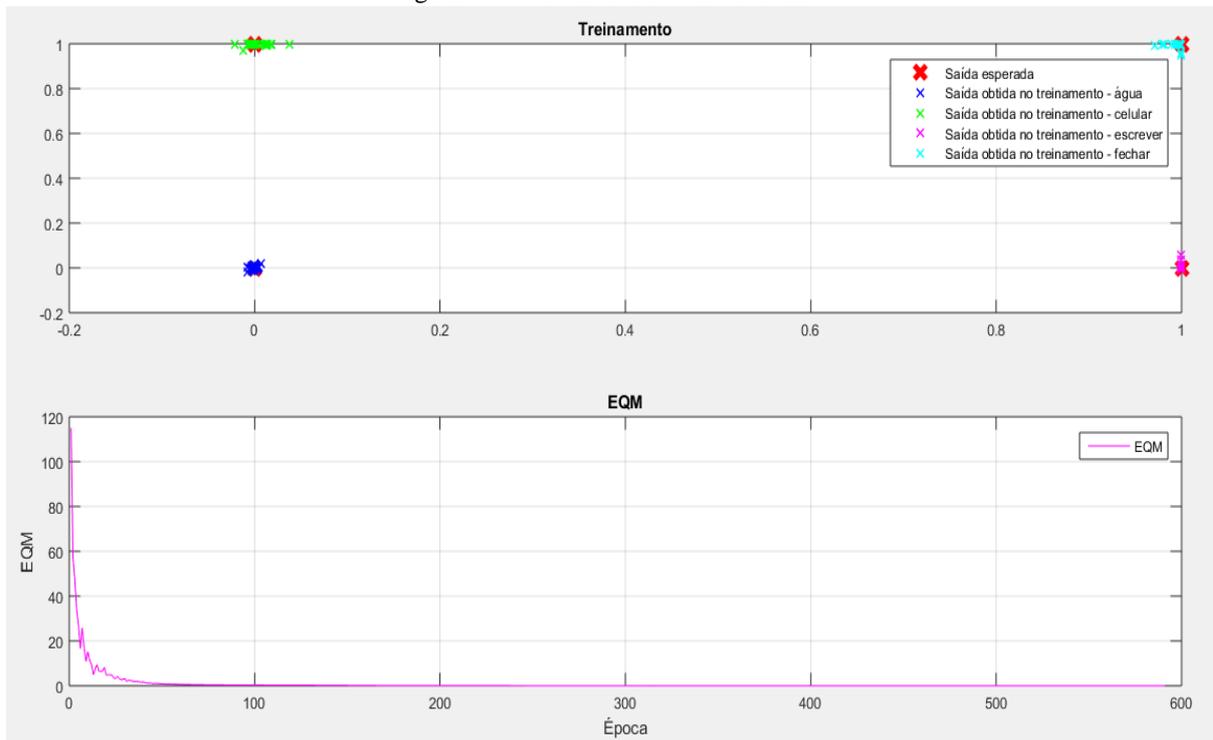
Fonte: Autora

Tabela 4.3 Acertos do treinamento de 7ª ordem

Treinamento LPC 7ª ordem		
Palavra	Acerto	EQM
<i>Água</i>	100%	0,0014
<i>Celular</i>	100%	0,0020
<i>Escrever</i>	100%	0,0033
<i>Fechar</i>	100%	0,0033
Acerto Total	100%	0,01

Fonte: Autora

Figura 14 - Treinamento LPC 10ª ordem



Fonte: Autora

Tabela 4.4 Acertos do treinamento de 10ª ordem

Treinamento LPC 10ª ordem		
Palavra	Acerto	EQM
<i>Água</i>	100%	0,0020
<i>Celular</i>	100%	0,0042
<i>Escrever</i>	100%	0,0021
<i>Fechar</i>	100%	0,0017
Acerto Total	100%	0,01

Fonte: Autora

Como pode ser observado nas figuras e tabelas anteriores, a fase de treinamento gerou excelentes resultados, tanto qualitativos quanto quantitativos, e a rede foi treinada em todas as quatro situações testadas. Vale ressaltar que o EQM total é o mesmo em todas as situações, pois o critério de parada da rede é exatamente que o erro seja igual ou inferior a 0.01. Obtendo o sucesso nesta fase, o próximo passo é verificar o quão bem essas redes podem generalizar os dados não contidos no grupo de treinamento, sendo esta a fase de validação.

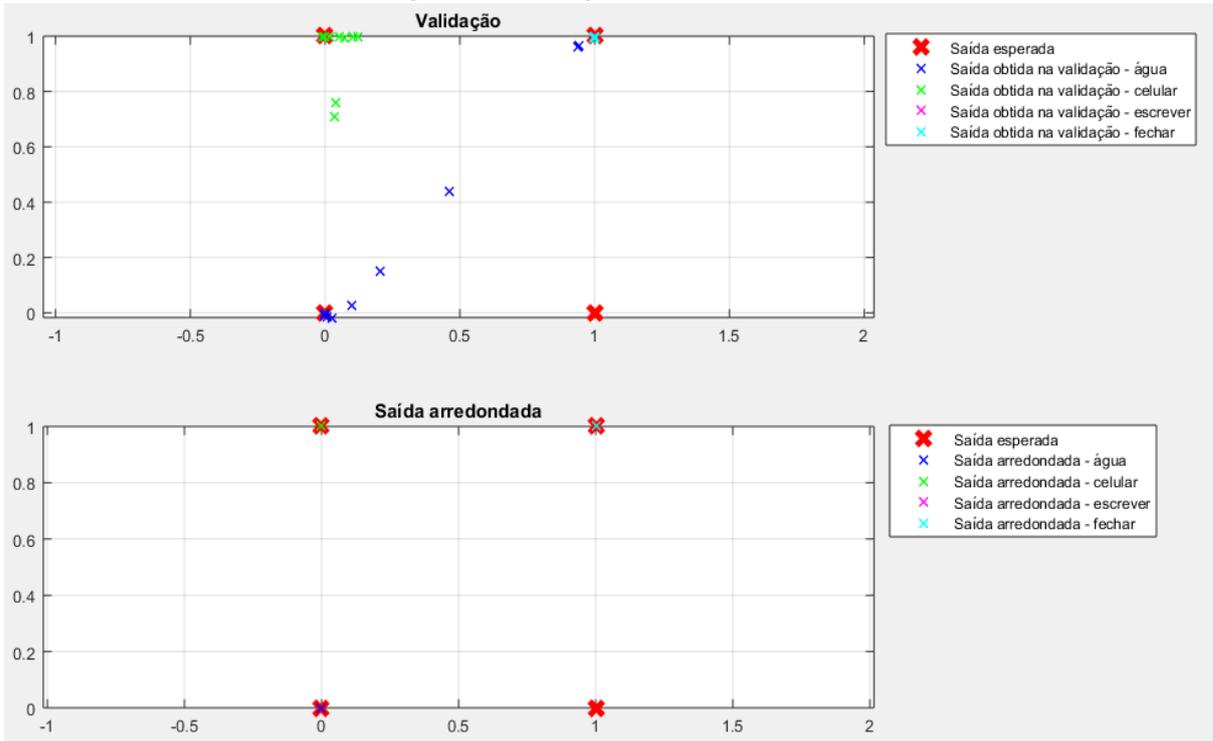
4.2 Resultados da fase de validação

Após a fase de treinamento, é realizada a etapa de validação e, neste trabalho, foram realizados dois testes de validação. No primeiro teste, foi utilizada uma parcela das entradas que não foram utilizadas no treinamento (20%), ou seja, amostras pertencentes ao mesmo grupo. Já para o segundo teste, foi utilizado um outro grupo de amostras, gravados em outro momento, utilizados inteiramente nesta etapa. Este procedimento se deve ao objetivo prático do trabalho, pois, uma vez que a rede é treinada pelo usuário, ela será utilizada em outros locais, com sinais gravados em momentos e contextos diferentes.

4.2.1 Teste 1 – Amostras do mesmo grupo

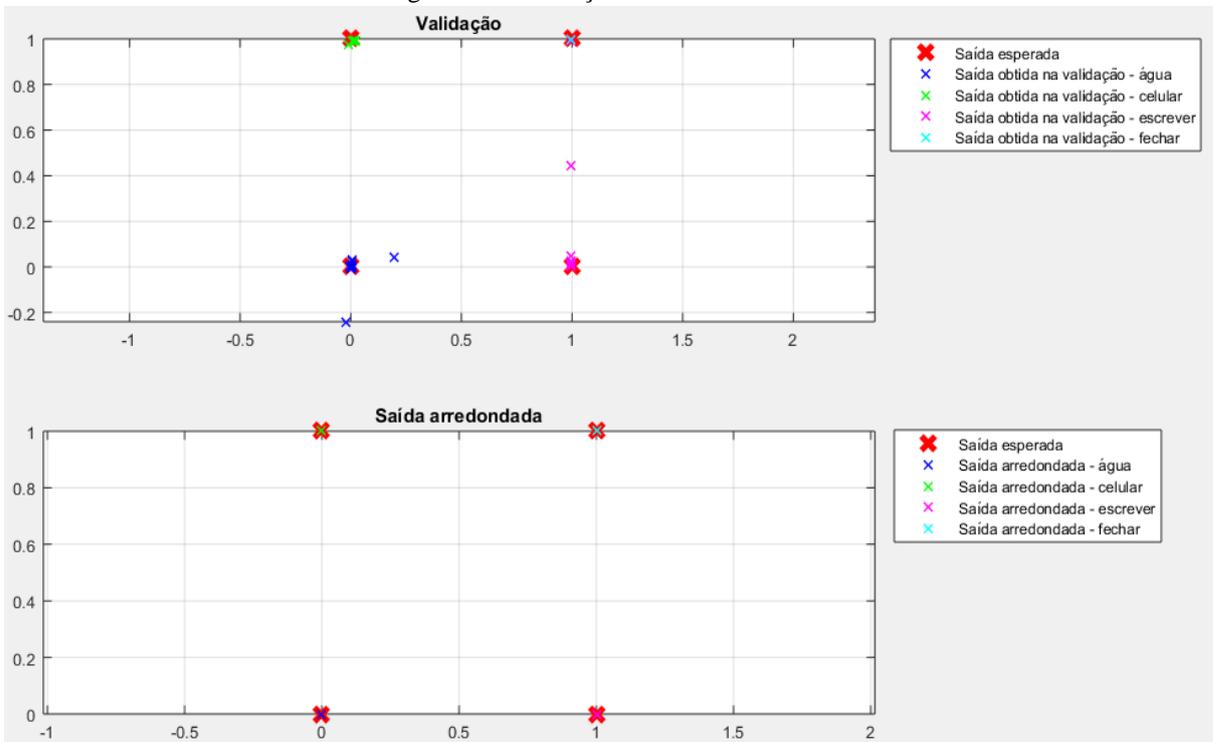
Nesta seção, serão demonstrados os resultados obtidos ao se validar os dados com uma parcela dos dados que não foram utilizados na fase de treinamento, porém, pertencentes ao mesmo grupo. Ao todo, foram validadas 40 amostras, sendo 10 de cada grupo. Os resultados estão nas figuras 15 a 19. Ao final da seção encontra-se uma tabela contendo os índices de acerto de cada palavra e o índice geral para cada ordem de LPC testada.

Figura 15 - Validação 1 LPC de 3ª ordem



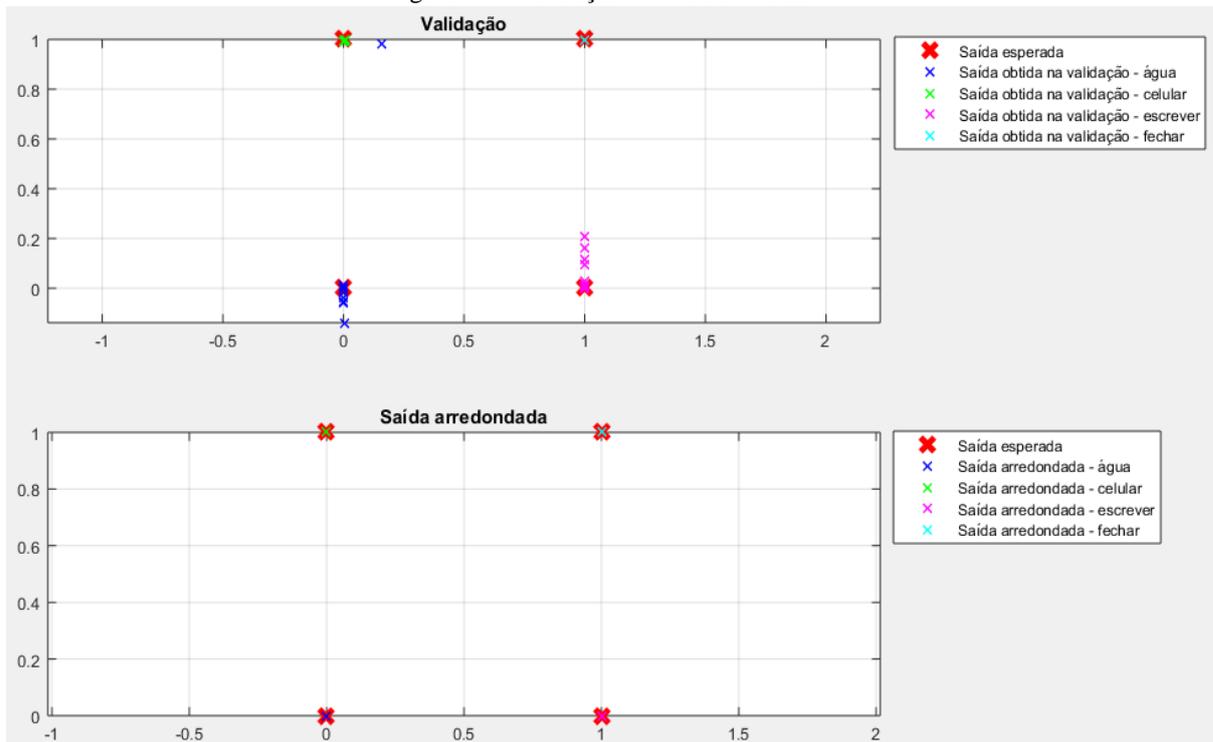
Fonte: Autora

Figura 16 - Validação 1 LPC de 5ª ordem



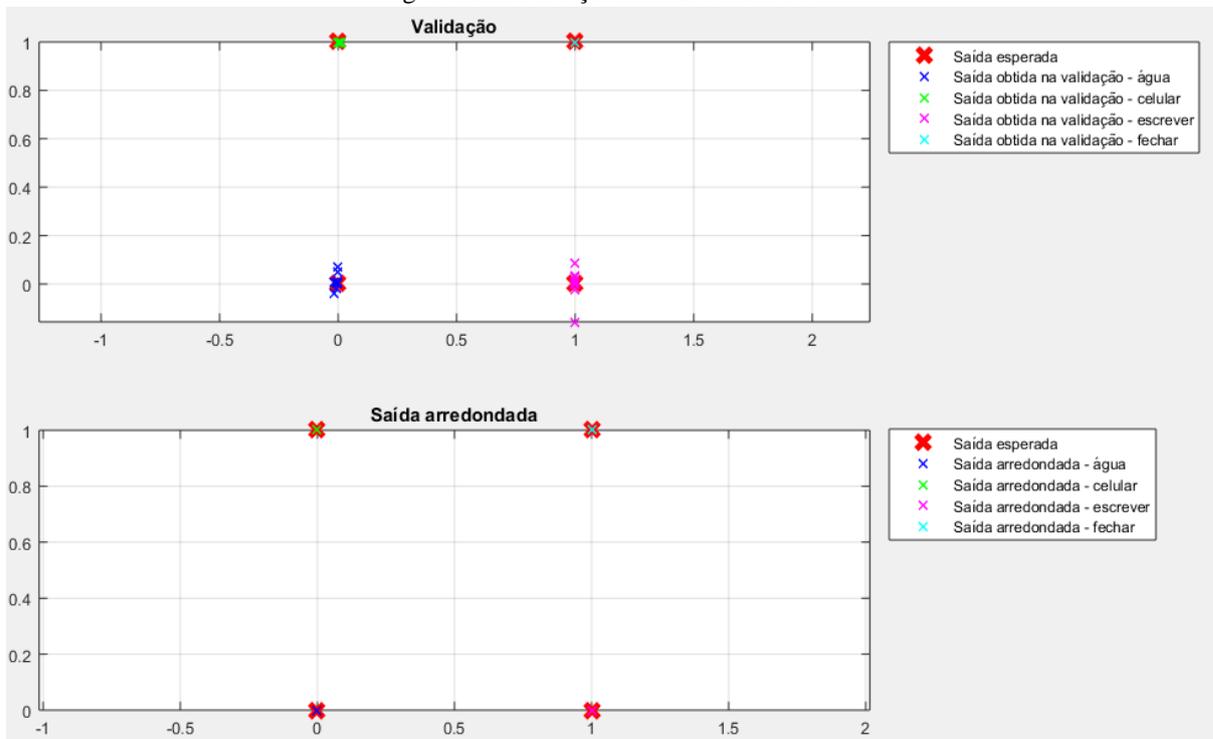
Fonte: Autora

Figura 17 - Validação 1 LPC de 7ª ordem



Fonte: Autora

Figura 18 - Validação 1 LPC de 10ª ordem



Fonte: Autora

Tabela 4.5 Resultados quantitativos para a validação

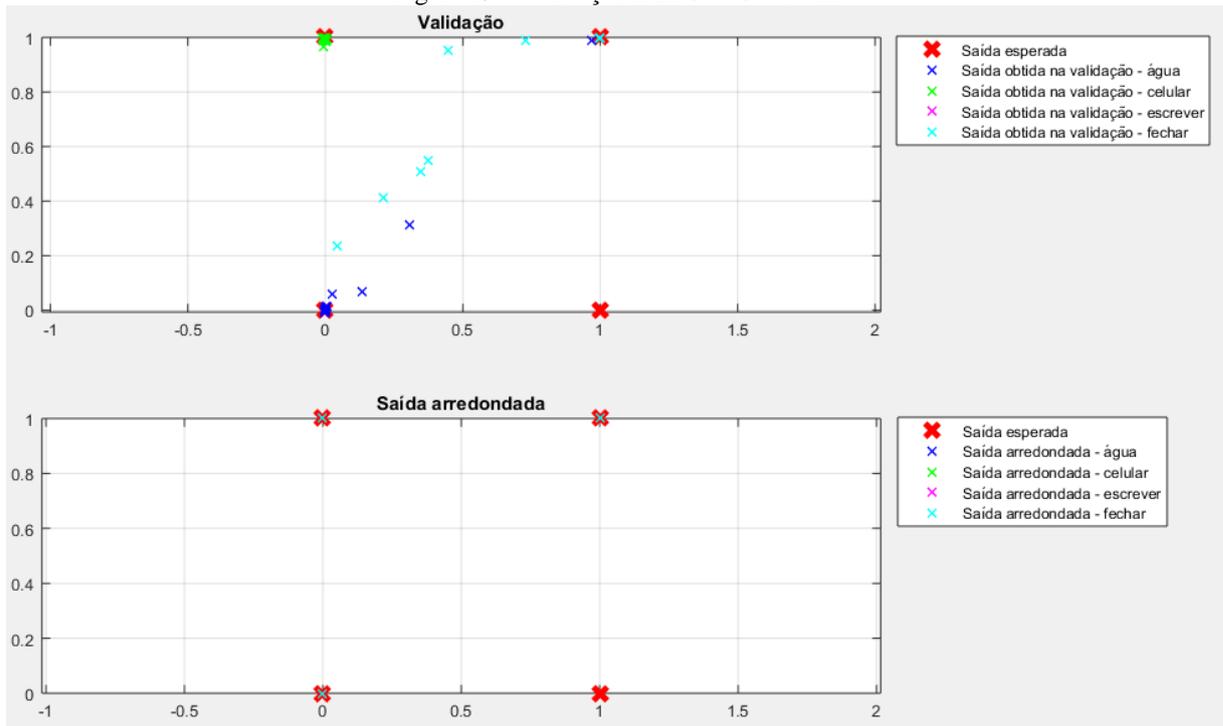
Validação 1 LPC 3ª ordem		Validação 1 LPC 5ª ordem		Validação 1 LPC 7ª ordem		Validação 1 LPC 10ª ordem	
Palavra	Acerto	Palavra	Acerto	Palavra	Acerto	Palavra	Acerto
<i>Água</i>	80%	<i>Água</i>	100%	<i>Água</i>	90%	<i>Água</i>	100%
<i>Celular</i>	100%	<i>Celular</i>	100%	<i>Celular</i>	100%	<i>Celular</i>	100%
<i>Escrever</i>	0%	<i>Escrever</i>	90%	<i>Escrever</i>	100%	<i>Escrever</i>	100%
<i>Fechar</i>	100%	<i>Fechar</i>	100%	<i>Fechar</i>	100%	<i>Fechar</i>	100%
Acerto Total	70%	Acerto Total	97.5%	Acerto Total	97.5%	Acerto Total	100%

De acordo com as figuras e a tabela 4.5, pode-se observar que, apesar de ter sido treinado pela rede, o LPC de 3ª ordem não fornece bons parâmetros para classificação dos grupos. Enquanto o de 10ª ordem, apresentou o melhor resultado, com um índice de acerto de 100%, o que demonstra a eficácia do método utilizado.

4.2.2 Teste 2 – Amostras de grupo distinto

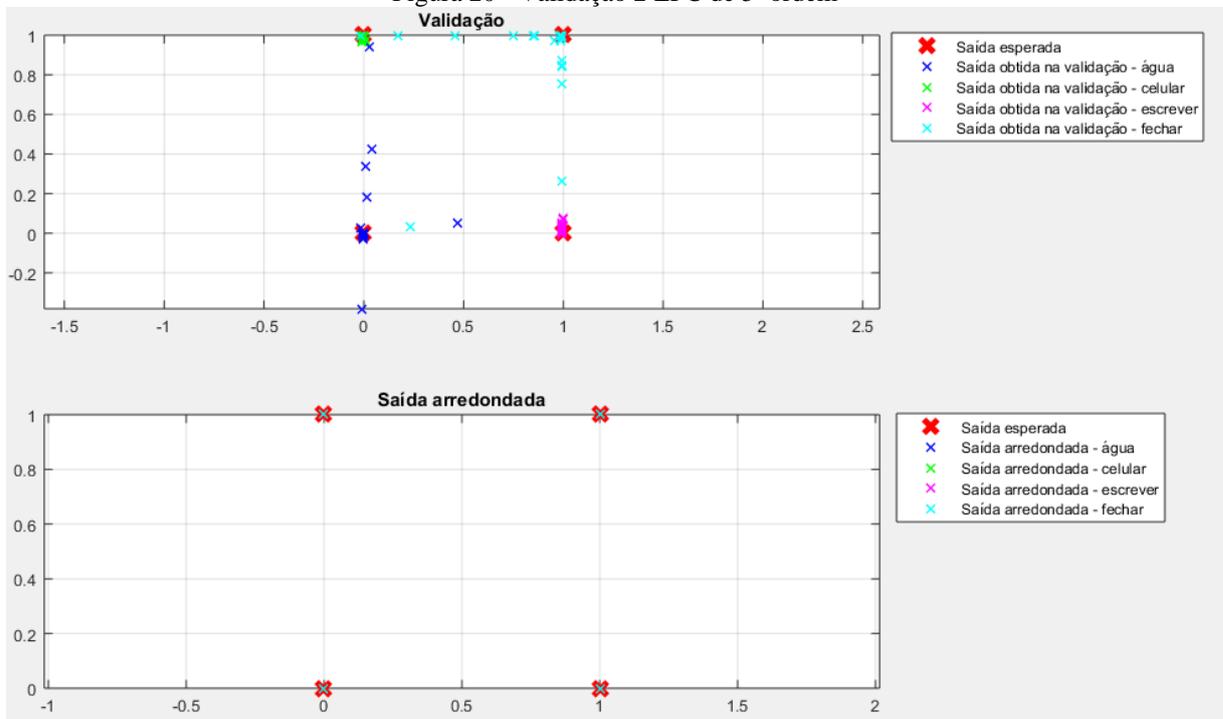
Nesta seção, serão demonstrados os resultados obtidos ao se validar os dados com um grupo de amostras gravados separadamente do grupo utilizado no treinamento. Ao todo foram gravadas 100 amostras, sendo 25 de cada palavra. Os resultados estão mostrados nas figuras 19 a 22. Ao final da seção encontra-se uma tabela contendo os índices de acerto de cada palavra e o índice geral para cada ordem de LPC testada.

Figura 19 - Validação 2 LPC de 3ª ordem



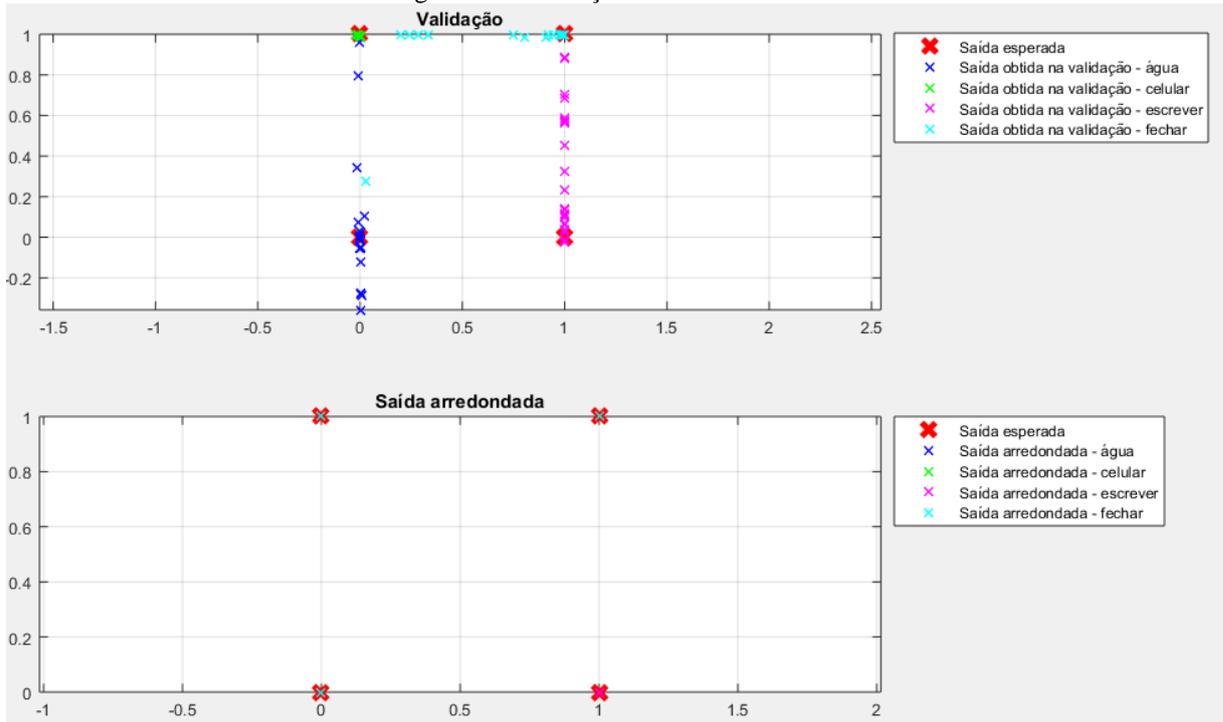
Fonte: Autora

Figura 20 - Validação 2 LPC de 5ª ordem



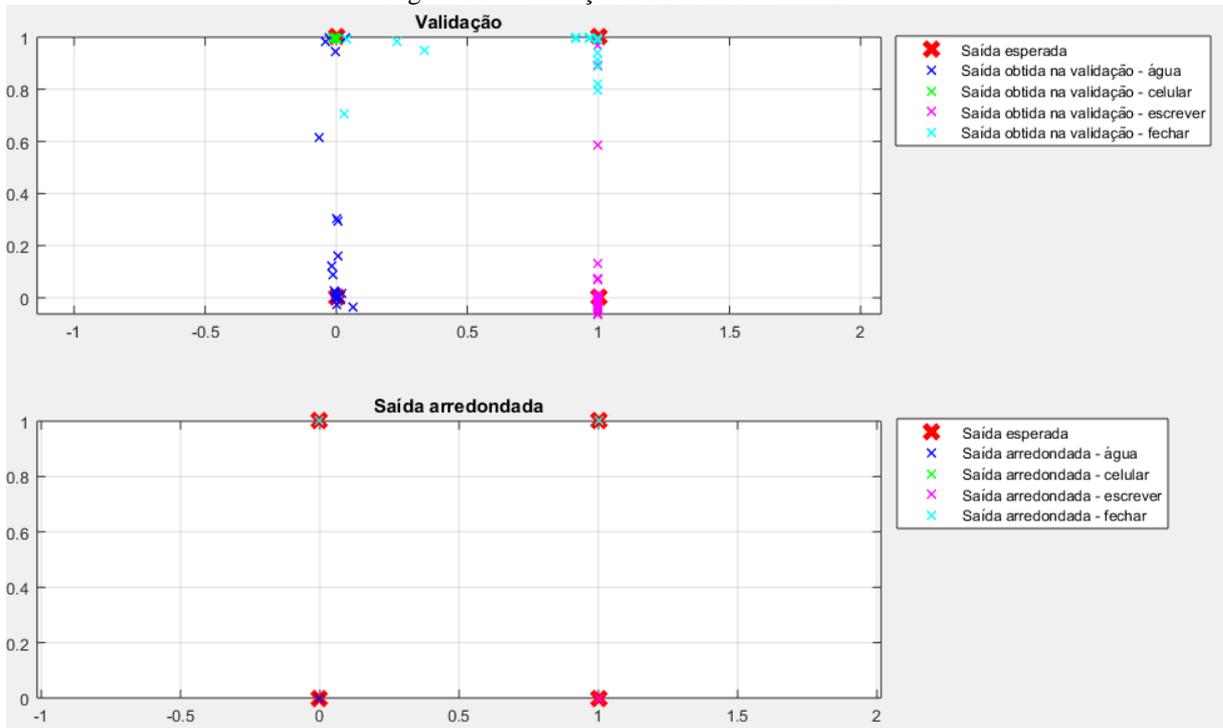
Fonte: Autora

Figura 21 - Validação 2 LPC de 7ª ordem



Fonte: Autora

Figura 22 - Validação 2 LPC de 10ª ordem



Fonte: Autora

Tabela 4.6 Resultados quantitativos para a validação

Validação 2 LPC 3ª ordem		Validação 2 LPC 5ª ordem		Validação 2 LPC 7ª ordem		Validação 2 LPC 10ª ordem	
Palavra	Acerto	Palavra	Acerto	Palavra	Acerto	Palavra	Acerto
<i>Água</i>	92%	<i>Água</i>	92%	<i>Água</i>	88%	<i>Água</i>	68%
<i>Celular</i>	100%	<i>Celular</i>	100%	<i>Celular</i>	100%	<i>Celular</i>	100%
<i>Escrever</i>	0%	<i>Escrever</i>	100%	<i>Escrever</i>	68%	<i>Escrever</i>	88%
<i>Fechar</i>	84%	<i>Fechar</i>	80%	<i>Fechar</i>	84%	<i>Fechar</i>	84%
Acerto Total	69%	Acerto Total	93%	Acerto Total	85%	Acerto Total	85%

Como já havia ocorrido na primeira validação, os parâmetros do LPC de 3ª não se mostraram eficazes. Porém, ao contrário do que ocorreu no teste anterior, neste teste, os parâmetros de 10ª ordem não geraram bons resultados. Os parâmetros de 5ª ordem foram os que apresentaram a melhor resposta, com 93% de acerto, o que demonstra mais robustez.

5 CONCLUSÕES E TRABALHOS FUTUROS

5.1 Conclusões

Inicialmente o trabalho foi dividido em duas partes. A primeira era a elaboração de um algoritmo que fosse capaz de realizar o reconhecimento automático de voz utilizando Redes Neurais Artificiais, dividida em treinamento da rede e validação dos dados. A segunda era a aplicação deste algoritmo em uma prótese robótica, para que o controle pudesse ser realizado por meio de comando de voz.

Para o treinamento da rede realizado na primeira etapa, o objetivo foi atingido e a rede foi treinada para cada uma das situações testadas (LPC de 3^a, 5^a, 7^a e 10^a ordem). Quanto a validação, os resultados podem ser divididos em dois tipos: a primeira situação é a situação na qual, de um mesmo conjunto, foram retiradas as amostras de treinamento e as amostras de validação, já para a segunda situação foram usadas amostras de conjuntos diferentes para o treinamento e para a validação.

Na primeira situação, foram obtidos resultados excelentes, com 100% de acerto para o LPC de 10^a ordem e 97,5% de acerto para o LPC de 5^a e 7^a ordem. Estes resultados confirmam a eficácia da RNA implementada e, ainda, a robustez do LPC aos ruídos.

Já para a segunda situação, a taxa mais alta de acerto foi do LPC de 5^a ordem, 93%, o que representa um bom resultado. Esta diferença na porcentagem de acertos se deve a grande discrepância de ruídos, a qual não foi possível a atenuação somente com a utilização do LPC ou dos filtros mais simples propostos. Era esperado ainda que a taxa de acerto fosse mais alta para as ordens de LPC superiores (7^a e 10^a), o que não ocorreu. Este resultado pode ser explicado pelo fato de que quanto mais alta a ordem, mais específica a rede se torna e, conseqüentemente, mais sensível. Sendo assim, ao se utilizar os parâmetros de 7^a e 10^a, a rede se tornou altamente específica em relação aos ruídos presentes nas amostras do treinamento, e, ao se modificar esses ruídos na validação, os resultados foram inferiores aos parâmetros de 5^a ordem, que não são tão específicos e sensíveis.

Ainda assim, conclui-se que o objetivo da implementação da RNA para reconhecimento de voz foi realizado com êxito.

5.2 Trabalhos Futuros

Para trabalhos futuros, podem ser realizadas melhorias na primeira etapa para obtenção de taxas mais altas de acerto quando amostras de diferentes grupos forem utilizadas, para que a rede possa ser mais precisa ao ser utilizada na prática. Como já mencionado, a diferença entre as taxas de acerto das duas situações se deve, principalmente, à diferença de ruídos das amostras. Para amenizar estes ruídos, o LPC e os filtros mais simples testados não foram suficientes, o que demonstra a necessidade de uma filtragem mais acurada. É importante ressaltar, que o filtro utilizado não deve atenuar frequências relevantes para o cálculo dos parâmetros do LPC, pois, caso isso ocorra, os resultados obtidos podem apresentar uma taxa de acerto inferior à taxa obtida sem o uso do filtro.

Já a segunda etapa, a execução dos comandos em uma prótese real, poderá ser executada em uma prótese impressa em impressora 3D, onde um determinado movimento dos motores da prótese é associado a saída da rede, que foi obtida de acordo com o comando de voz executado.

REFERÊNCIAS

- [1] MOHAMAD, Siti Nur Ateeqa; JAMALUDIN, Ahmad Ammar; ISA, Khalid. "*Speech semantic recognition system for an assistive robotic application*". IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Selangor, Malaysia, 2016.
- [2] KELLER, James; THAKUR, Dinesh; LIKHACHEV, Maxim; GALLIER, Jean; KUMAR, Vijay. "*Coordinated Path Planning for Fixed-Wing UAS Conducting Persistent Surveillance Missions*". IEEE Transactions on Automation Science and Engineering, Volume 14, Issue: 1, pages 17-24, 2017.
- [3] FLUSHING, Eduardo Feo; GAMBARDELLA, Luca; DI CARO, Gianni. "*On Using Mobile Robotic Relays for Adaptive Communication in Search and Rescue Missions*". 2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Lausanne, Switzerland, 2016.
- [4] Notas de aula do Professor Leonardo Olivi. Disciplina Robótica Móvel, Faculdade de Engenharia, Universidade Federal de Juiz de Fora, UFJF, 2016.
- [5] *Sistema IBGE de Recuperação Automática – SIDRA*. Disponível em <<https://sidra.ibge.gov.br/Tabela/3425>>. Acesso em 27/04/2017.
- [6] DA SILVA, Luiz Augusto Zillmann. "*Controle de Prótese Mecânica Aplicado à Robótica Assistiva através de Eletromiografia*". Trabalho de Conclusão de Curso, Faculdade de Engenharia, Universidade Federal de Juiz de Fora, UFJF 2016.
- [7] MADAN, Akansha; GUPTA, Divya. "*Speech Feature Extraction and Classification: A Comparative Review*". International Journal of Computer Applications (0975 – 8887). Março de 2014.
- [8] POTAMIANOS, Gerasimos; JELINEK, Frederick. "*A study of N-gram and decision tree letter language modeling methods*". Speech Commun. 24 (3), 171–192. Junho de 1998.
- [9] ROSENFELD, Ronald. "*Adaptive Statistical Language Modeling: A Maximum Entropy Approach*". Tese de Doutorado, School of Computer Science, Carnegie Mellon University. Abril de 1994.
- [10] BENGIO, Y; DUCHARME, R; VINCENT, P. "*A neural probabilistic language model*". Journal of Machine Learning Research 3, 1137–1155. Agosto de 2003.
- [11] DA SILVA, Vinicius R. V. G. "*Algoritmos para Redução de Ruídos em Sinais de Áudio*". Trabalho de Conclusão de Curso, Departamento de Engenharia Eletrônica e de Computação, Escola de Engenharia, Universidade Federal do Rio de Janeiro, UFRJ. Dezembro de 2007.
- [12] BRANDÃO, Alexandre Santos. "*Redes Neurais Artificiais Aplicadas ao Reconhecimento de Comandos de Voz*". Departamento de Engenharia Elétrica e de Produção, Centro de Ciências Exatas e Tecnológicas, Universidade Federal de Viçosa, UFV. Junho de 2005.

[13] KURZEKARA, Pratok K.; DESHMUKH Ratnadeep R.; WAGHMARE, Vischal B.; SHRISHRIMAL, Pukhraj P. “A Comparative Study of Feature Extraction Techniques for Speech Recognition System”. *International Journal of Innovative Research in Science, Engineering and Technology*. Vol. 3, Issue 12. Dezembro de 2014.

[14] *FAKSTAFF*. Disponível em <<http://www.facstaff.bucknell.edu/mastascu/elessonshtml/signal/signal2.htm>>. Acesso em 09/05/2017.

[15] *KEYSIGHT TECHNOLOGIES*.

Disponível em <http://rfmw.em.keysight.com/rfcomms/refdocs/wcdma/wcdma_meas_occup_bw_desc.html> Acesso em 09/05/2017.

[16] *UOL EDUCAÇÃO*.

Disponível em <<https://educacao.uol.com.br/disciplinas/matematica/desvio-padrao-entenda-este-conceito-de-estatistica.html>>. Acesso em 09/05/2017

[17] UNICAMP. *Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Departamento de Sistema e Controle de Energia*. Disponível em <<http://www.dsce.fee.unicamp.br/~antenor/pdf/qualidade/b1.pdf>>. Acesso em 09/05/2013.

[18] USP. *Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação*. Disponível em <<http://conteudo.icmc.usp.br/pessoas/andre/research/neural/#hist>>. Acesso em 11/05/2013.

[19] DA SILVA, Ivan Nunes; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade. “*Redes Neurais Artificiais: para engenharia e ciências aplicadas*”. Editora Artliber, 2010.

[20] *GSIGMA Grupo de Sistemas Inteligentes de Manufatura*.

Disponível em <http://www.gsigma.ufsc.br/~popov/aulas/rna/neuronio_artificial/index.html>. Acesso em 11/05/2017.

[21] *RESEARCHGATE*.

Disponível em <https://www.researchgate.net/figure/228731551_fig4_Architectural-graph-of-a-multilayer-perceptron-with-two-hidden-layers>. Acesso em 12/05/2017.

[22] TEBELSKIS, Joe. “*Speech Recognition using Neural Networks*”. Tese de doutorado, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania. Maio de 1995. “An eficiente noise robust speech recognition system using artificial neural networks” *Communication and Signal Processing (ICCSP)*. Abril de 2016.

[23] SANTOSH, Gupta; BHURCHANDI, Kishor M.; KESKAR Avinash G. “*An eficiente noise robust speech recognition system using artificial neural networks*”. *Communication and Signal Processing (ICCSP)*, Melmaruvathur, India. Abril de 2016.

[24] UFBA. *Universidade Federal da Bahia Escola Politécnica, Departamento de Engenharia Elétrica*. Disponível em <http://www.dee.eng.ufba.br/home/simas/PDS_aula03.pdf>. Acesso em 13/05/2016.

[25] SANDRU, Daniel Florin; NANU, Sorin; SILEA Ioan; MICLEA Razvan Catalin. “*Kalman and Butterworth filtering for GNSS/INS data*”. Electronics and Telecommunications (ISETC), 2016 12th IEEE International Symposium. Timisoara, Romania. Dezembro de 2016.

[26] SHRAWANKAR, Urmila; THAKARE, Vilas. “*TECHNIQUES FOR Feature in Speech Recognition System: A Comparative Study*”.

[27] KESARKAR, Manish P. “*Feature Extraction for Speech Recognition*”. M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay. Novembro de 2003.