UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

TAMIRIS DO VALE

FILTRAGEM INVERSA PARA IDENTIFICAÇÃO DO SINAL GLOTAL VIA SINTESE DE FALA

CURITIBA
2022

TAMIRIS DO VALE

FILTRAGEM INVERSA PARA IDENTIFICAÇÃO DO SINAL GLOTAL VIA SINTESE DE FALA

INVERSE FILTERING TO OBTAIN THE GLOTTAL SIGNAL VIA SPEECH OF SYNTHESIS

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia Elétrica do curso de Engenharia Elétrica da Universidade Tecnológica Federal do Paraná (UTFPR).
Orientador(a): Dr. Marcelo de Oliveira Rosa

CURITIBA
2022

TAMIRIS DO VALE

FILTRAGEM INVERSA PARA IDENTIFICAÇÃO DO SINAL GLOTAL VIA SINTESE DE FALA

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia Elétrica do curso de Engenharia Elétrica da Universidade Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 08/dezembro/2022

_____

Marcelo de Oliveira Rosa
Titulação: Doutorado.
Universidade Tecnológica Federal do Paraná

_____

Antônio Carlos Pinho
Titulação: Doutorado
Universidade Tecnológica Federal do Paraná

_____

Annemarlen Gehrke Castagna
Titulação: Mestrado
Universidade Tecnológica Federal do Paraná

CURITIBA

2022

Dedico este trabalho aos meus irmãos por serem o motivo de eu não desistir.

AGRADECIMENTOS

Chegar até aqui não foi fácil, foram meses intensos de pesquisa, escrita e desenvolvimento. Foi sofrido, mas graças a Deus tive muita gente para me ajudar, seja com incentivo, uma palavra amiga ou um colo quando eu pensava em desistir, e nesse momento que essa saga termina gostaria de agradecer a essas pessoas por tudo nessa caminhada.

Agradeço ao meu orientador Prof. Dr. Marcelo de Oliveira Rosa, pela sabedoria com que me guiou e pela paciência em me orientar, obrigada professor sem o senhor esse trabalho não seria possível.

Aos meus amigos que me aguentaram, torceram e me incentivaram nos momentos de desespero com uma palavra amiga ou só ficando por perto torcendo para que eu concluísse esse sonho, não citarei nominalmente para não correr o risco de esquecer de alguém, mas sintam-se agradecidos e homenageados.

Gostaria de deixar registrado também, o meu reconhecimento a minha mãe e em especial aos meus irmãos Henrique e Rodrigo pois acredito que sem o apoio deles seria muito difícil vencer esse desafio.

Enfim, o meu muito obrigada a todos os que por algum motivo contribuíram para a realização deste trabalho.

RESUMO

A fala é um sinal resultado da modulação realizada pela cavidade bucal e o trato supraglotal sobre o sinal glotal. O sinal glotal é um sinal importante porque traz informações sobre a laringe, podendo ser utilizado para a identificação de patologias da laringe. Para extrair o sinal glotal do sinal de fala e necessário a utilização de uma ferramenta de identificação de sistemas, a rede neural. Esse trabalho visa construir duas configurações a partir de arquiteturas de redes neurais para extração do sinal glotal a partir do sinal de fala e avaliar o desempenho de cada uma delas, e escolher qual rede mais se adequa ao problema proposto.

Palavras-chave: Redes neurais. Sinal de fala. Sinal glotal. Laringe.

ABSTRACT

The speech is a signal resulting from the modulation performed by the oral cavity and the supraglottal tract over the glottal signal. The glottal signal is an important signal because it brings information about the larynx, and can be used for the identification of laryngeal pathologies. To extract the glottal signal from the speech signal, it is necessary to use a system identification tool, the neural network. This paper aims to build two configurations from neural network architectures to extract the glottal signal from the speech signal and evaluate the performance of each one of them.

Keywords: Neural networks. Speech signal. Glottal signal. Larynx.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1. INTRODUÇÃO

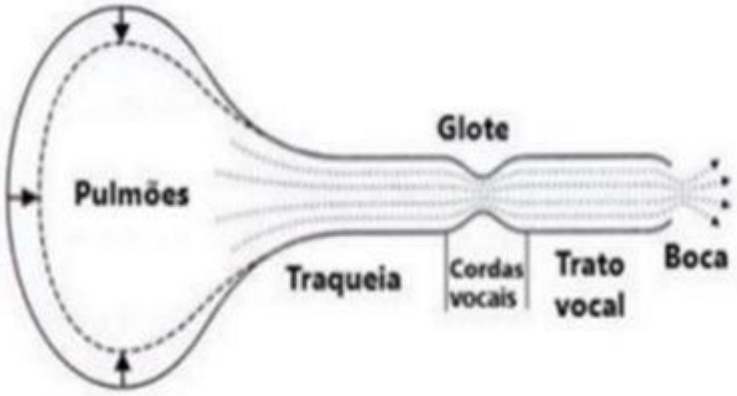
1.1 TEMA

A fala é um importante meio de comunicação que o ser humano possui sendo uma forma de comunicação que o difere dos demais animais. A palavras que compõe a fala são produzidas a partir de um conjunto complexo e finito de fonemas (ROSA,2022).

A produção da fala é um processo complexo, começando a partir de um escoamento de ar dos pulmões, devido à diferença de pressão entre os pulmões e a parte externa da boca. Esse fluxo de ar, passando pela laringe gera um trem de pulsos quase periódico chamado sinal glotal (CATALDO,2004).

Segundo Cataldo, Sampaio e Nicolato (2004), o sinal glotal é um sinal de baixa intensidade, que necessita de amplificação para que fonemas sejam caracterizados, sem que tal amplificação é realizada pelo trato vocal.

Figura 1: Sistema esquemático da produção de voz



Fonte: Cataldo, Sampaio, Nicolato (2004)

A produção do sinal glotal está atrelada às características anatômicas e fisiológicas da laringe, características que podem ser representadas por meio de analogias de sistemas físicos como o modelo de Ishizaka e Flanagan (ISHIZAKA E FLANAGAN,1972), que representa as cordas vocais como um sistema de duas massas. Apesar de ser um modelo simples, este carrega informações e

complexidade suficiente para sintetizar o sinal glotal e de voz (BROGIAN, PALMAS, VON KRUGER,2017).

Por carregar informações importantes do sistema que o gerou, a voz é utilizada como alternativa não-invasiva de análise das condições da laringe, pois determinadas patologias afetam o sinal glotal e por consequência, o sinal de voz. (ROSA,2002).

As equações que relacionam o sinal glotal e o sinal de voz, mesmo com o uso de modelos simples como o de duas massas (ISHIZAKA E FLANAGAN,1972) são complexas. Para se encontrar essas relações entre o sinal e o sistema que o gerou um dos métodos a ser utilizado é a identificação de sistemas. Dentre as ferramentas disponíveis para identificação de sistemas, foi utilizado o processo de filtragem inversa usando redes neurais artificiais para reconhecimento das relações entre o sinal de voz e o sinal glotal.

O foco deste trabalho foi encontrar a rede neural que dentro das duas arquiteturas de rede escolhidas que mais se adeque ao sistema produção de fala, e preparar uma estrutura de treinamento e estimação dos pesos da rede neural para que esta se comporte como um filtro que quando aplicado a sinais de fala, gere uma estimativa razoável do sinal glotal.

1.1.1 Delimitação do tema

O sinal glotal, resultado da passagem de ar pelas cordas vocais da laringe, tem suas componentes espectrais modificadas devido ao formato do trato supraglotal, no qual se inclui movimento da língua e do palato. Cada fonema gera formas distintas de trato supraglotal, funcionando como um filtro que amplifica componentes espectrais em frequências conhecidas como frequências formantes. Mesmo relacionadas com tipos de sons/fonemas distintos, tais frequências formantes acabam variando de pessoa para pessoa (ROSA,1998).

O trabalho proposto visa extrair a relação entre sinal de fala e o sinal glotal através de uma ferramenta de identificação de sistemas, a rede neural. A base de

sinais de entrada (sinal de fala) e de saída (sinais glotais) foi obtida a partir de um sintetizador parametrizado de sinais de fala e glotal, que empregam as equações matemáticas apresentadas no modelo de Ishizaka e Flanagan e representação do trato supraglotal usando tubos cilíndricos concatenados. Diferentes padrões de sinais de fala e glotal foram gerados a partir de diferentes valores para os parâmetros de trato supraglotal do sintetizador de maneira a representar fonemas etc.

Uma rede neural adequadamente treinada funciona como um filtro para sinais de vozes, gerando sinais glotais. A ferramenta computacional a ser utilizada será o MATLAB. Por meio desse software foram realizadas simulações do modelo do trato vocal para produzir os sinais glotal e de fala que treinaram as redes neurais estudadas.

1.2 PROBLEMAS E PREMISSAS

Segundo Scalassara (2009), o sinal glotal pode ser dividido em dois tipos: o que dá origem aos sons vocálicos e o que dá origem aos sons não vocálicos. Os sons vocálicos são gerados pela oscilação relaxada das pregas vocais, gerando um sinal quase periódico (um exemplo de som vocálico é o som ao se pronunciar uma vogal). Os sons não vocálicos, por sua vez, são os gerados sem a participação da laringe que abre sem se opor a passagem do ar, tornando o fluxo de ar turbulento. Um exemplo desse som é o que se produz ao pronunciar uma consoante tal como em /r/ em ‘Rato’).

O sinal glotal, ao passar pelo trato supraglotal tem algumas componentes amplificadas e outras atenuadas em frequência pelas componentes do trato vocal. Sendo assim o “caminho” direto da produção do sinal de voz e a indução de um fluxo de ar a partir dos pulmões, passando pela laringe que excita um sinal glotal, que por sua vez chega ao trato vocal e é modulado gerando o sinal de voz.

Com objetivo de reproduzir o processo inverso, ou seja, dado um determinado sinal de voz obter o sinal glotal que o gerou, utilizam-se ferramentas de

identificação de sistemas dentro das ferramentas possíveis optou-se pela rede neural.

As RNAs (Redes neurais artificiais) são modelos que tentam reproduzir o funcionamento do cérebro humano para resolução de problemas e realização de tarefas complexas que seriam difíceis de serem reproduzidas usando sistemas ditos convencionais (HAYKIN,2001). Os sistemas que usam redes neurais são considerados inteligentes, pois empregam algoritmos que permitem a adaptação do sistema com intuito de produzir a saída desejada. Esse processo de adaptação é realizado a partir de treinamento, supervisionado ou não, usando apenas os dados disponíveis com pouca informação relativa ao sistema que relaciona tais dados (ALVARENGA,2012).

O modelo mais utilizado é a *Multilayer Perceptron* (MLP), que possui diversas camadas de neurônios, dentre elas uma camada de entrada e a camada de saída (BARAVIERA, 2016), as camadas neurais ocultas possibilitam que a rede aprenda a partir da extração de parâmetros mais significativos a partir da entrada sendo propagado na rede camada a camada, cada neurônio tem como entrada a saída de todas as camadas anteriores (LEMOS,2014).

Outra modelo de rede utilizado é o *Nonlinear autoregressive exogenous* (NARX), comumente utilizada para sistemas dinâmicos não lineares, pois essa rede além de a camada de entrada receber realimentação com atraso de tempo ela recebe dados exteriores com atraso, fazendo com que o resultado seja dinâmico variando em função do tempo (SOUZA,2019).

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Construir um filtro inverso a partir de redes neurais que ao ser aplicado ao sinal de fala gere uma estimativa razoável do sinal glotal.

1.3.2 Objetivos Específicos

- ✓ Uma revisão bibliográfica acerca dos temas tais como produção da fala, sinal glotal, identificação de sistemas, redes neurais, envolvidos para desenvolvimento do trabalho proposto;
- ✓ Compreender o funcionamento do sintetizador de fala a ser utilizado;
- ✓ Desenvolver um banco de sinais glotais e de fala a partir da variação dos parâmetros no sintetizador;
- ✓ Encontrar dentre os modelos de redes neurais conhecidos o que se adequar ao problema proposto;
- ✓ Realizar a rotina de aprendizado do modelo de rede neural escolhido.
- ✓ Testar o resultado obtido a fim de verificar a eficácia do filtro obtido.

1.4 JUSTIFICATIVAS

Um dos principais motivos que levam ao estudo da produção de voz é a importância que a fala, como meio de comunicação, tem na vida do ser humano, sendo que alterações na voz podem ter consequências a vida social e profissional de um indivíduo (SILVA,2015).

Atualmente, o diagnóstico das patologias/que acometem a laringe é realizado através de exames, que são realizadas de forma invasiva, tais como a laringoscopia, a vídeo laringoscopia, naso fibroscopia exames esses que em sua execução trazem desconforto ao paciente (BROGIAN, PALMAS, VON KRUGER, 2017)

Além de desconforto, outra dificuldade encontrada no diagnóstico de patologia na laringe é a exigência de perícia e experiência do especialista que irá avaliar as queixas do paciente. Dessa forma o diagnóstico acaba sendo impreciso e subjetivo.

Com essas dificuldades de diagnóstico, faz se necessária busca de alternativas para a análise de patologias na laringe. Uma das alternativas que vêm sendo estudada é o uso do sinal de voz para auxílio no diagnóstico idealmente seria interessante analisar o sinal produzido pela laringe, denominado sinal glotal que traz efetivamente informações sobre o funcionamento da laringe, entretanto esse sinal não pode ser obtido diretamente.

Sendo assim o presente trabalho visa a obtenção do sinal glotal a partir de sinal de voz, com intuito de contribuir futuramente com a área médica para desenvolver formas alternativas de diagnósticos, pois o sinal glotal traz características importantes sobre a laringe.

1.5 PROCEDIMENTOS METODOLÓGICOS

Para o desenvolvimento do estudo “Filtragem Inversa para Identificação do Sinal Glotal via Síntese de Fala”, foi feita uma revisão bibliográfica acerca do tema, para compreensão do processo e equações que descrevem a produção da voz, do funcionamento do sintetizador, obtido através do estudo “Síntese de sinais de voz usando modelos biologicamente inspirados” realizado por Brogian, Palmas e Kruger (2017), também foi realizado um estudo de técnicas de processamento de sinais, em específico as redes neurais, com o objetivo de encontrar a que se adeque ao problema de identificação de sistemas proposto.

Em seguida foi construído com o auxílio do sintetizador disponível (Brogian, Palmas, Von Kruger,2017) para gerar um número significativos de pares de sinais (de voz e glotal relacionado) a fim de se reconhecer as relações existentes entre esses sinais. Este banco de pares de sinais foi utilizado para implementação de uma

rotina de treinamento das redes neurais MLP e NARX, redes que foram modificadas e avaliados seus respectivos desempenhos baseados no critério apresentado nos próximos capítulos.

A ferramenta computacional escolhida para desenvolver os processos descritos, foi *software* numérico MATLAB, sendo utilizada sua biblioteca de processamento de sinais e de redes neurais.

1.6 ESTRUTURA DO TRABALHO

Este trabalho de conclusão de curso terá como estrutura um total de 4 capítulos, e será dividido da seguinte forma:

Capítulo 1: Introdução - Visão geral de todo o trabalho em formato de uma proposta contendo o tema, delimitação do tema, problemas, objetivos, justificativas e procedimentos metodológicos.

Capítulo 2: Fundamentação teórica - Fundamentação teórica sobre o sistema de produção de fala, modelo matemático da laringe e do trato vocal, formação do sinal glotal, trato supraglotal, identificação de sistemas, redes neurais.

Capítulo 3: Processos metodológicos - Apresentação dos processos metodológicos utilizados, como a descrição do processo de criação de pares de sinal de fala e glotal que foram utilizados no processo de treinamento das redes neurais configuradas nos parâmetros escolhidos

Capítulo 4: Resultados e Discussões - Apresentação dos resultados obtidos e conclusões finais do estudo proposto, baseado no resultado obtido.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 SISTEMA DE PRODUÇÃO DE FALA

Considerada um dos principais meios de comunicação do ser humano, a fala tem como principais órgãos envolvidos em seu processo de produção, os órgãos que compõe o sistema respiratório e sistema de deglutição envolvidos são pulmões, traqueia, laringe e a cavidade supraglotal (boca, nariz e língua) que atuam em conjunto para a produção de fonemas e posteriormente a fala (MORI,2005) quando uma das estruturas envolvidas apresenta alguma variação (como patologias) ocorre a alteração no sinal gerado (ROSA ,2002).

Para facilitar a compreensão do sistema de produção da fala nas literaturas é comum que a sua estrutura de produção seja dívida em 3 grupos: respiração, vocalização, ressonância.

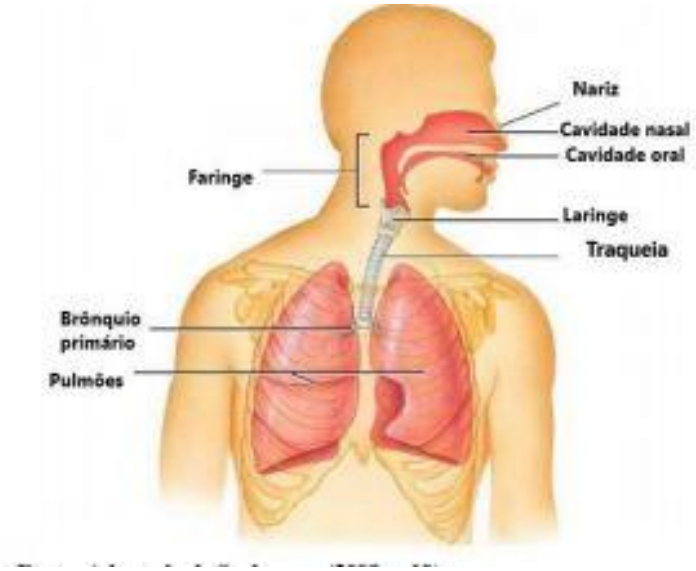
Na respiração, como próprio nome diz o sistema diretamente ligado nessa etapa é o respiratório, composto por um par de pulmões, sacos alveolares e bronquíolos.

Nesse sistema os pulmões funcionam como um reservatório, em que as trocas gasosas ocorrem, entregando massa de ar para os pares de brônquios (um para cada pulmão). Esses brônquios se unem a traqueia que permite que o ar circule para dentro e fora do pulmão. Além dessas estruturas tem-se o diafragma, que tem a finalidade de separar a caixa torácica da abdominal como também tem papel importante atividade respiratória e por consequência no processo de produção da fala (ROSA,1998).

Os processos respiratórios de expiração e inspiração são realizados por movimentos de expansão e contração dos pulmões que são realizados pelo movimento da caixa torácica, em conjunto com o diafragma e os músculos intercostais, sendo os movimentos expiratórios que causados pelo relaxamento natural dos pulmões (colapso passivo do tórax) os responsáveis pelo escoamento de

uma certa vazão de ar que ocorre no momento da fala. Essa vazão é definida pela diferença entre a pressão interna nos pulmões e a pressão externa (fora da boca) (ROSA,2002). O escoamento que é levado pela traqueia que funciona como um pequeno duto de ar (Rosa,1998).

Figura 2:Sistema Respiratório



Fonte: Scalassara (2009)

A figura 2 ilustra o sistema respiratório, apresentando a posição de cada órgão do sistema respiratório, órgãos que participam da produção de voz.

Passando pela traqueia, o fluxo de ar proveniente dos pulmões encontra o segundo grupo do sistema fonador, a vocalização, composto pela laringe, epiglote e cordas vocais.

No processo de vocalização o escoamento de ar produzido pelos pulmões, é conduzido pela traqueia até a laringe ou cavidade glotal que por sua vez gera um sinal pulsátil quasi-periódico, chamado de sinal glotal ou trem de pulsos.

Esse sinal glotal produzido segue para o trato supraglotal onde é realizada a amplificação do sinal para que seus componentes harmônicos sejam enfatizados gerando o que conhecemos por sinal de voz propriamente dito, e assim tem se o último processo da produção de fala, o sistema ressonador.

O sistema ressonador, conhecido como trato supraglotal é composto por boca, nariz e faringe esse sistema é o responsável por modular o sinal sendo a língua importante componente na variação do tamanho da área de ressonância, (ROSA,2011) e à medida que se modifica o formato do trato vocal, altera-se os sons básicos gerados pela laringe, o que faz com que existam uma variedade de timbres sonoros (RAZERA, 2004, p. 8)).

Para uma maior compreensão da geração do sinal glotal tem-se que compreender a laringe e seus músculos que são de suma importância para o sinal que virá a ser, após amplificação e atenuação de componente, o sinal de voz propriamente dito.

2.2 LARINGE E A PRODUÇÃO DO SINAL GLOTAL

A laringe além de ligada a proteção do sistema respiratório, impedindo que corpos estranhos atinjam o pulmão causando problemas de saúde, ela também está ligada a produção do sinal glotal juntamente com as cordas vocais, funcionando grosso modo como um controle de vazão de ar (ROSA,2002).

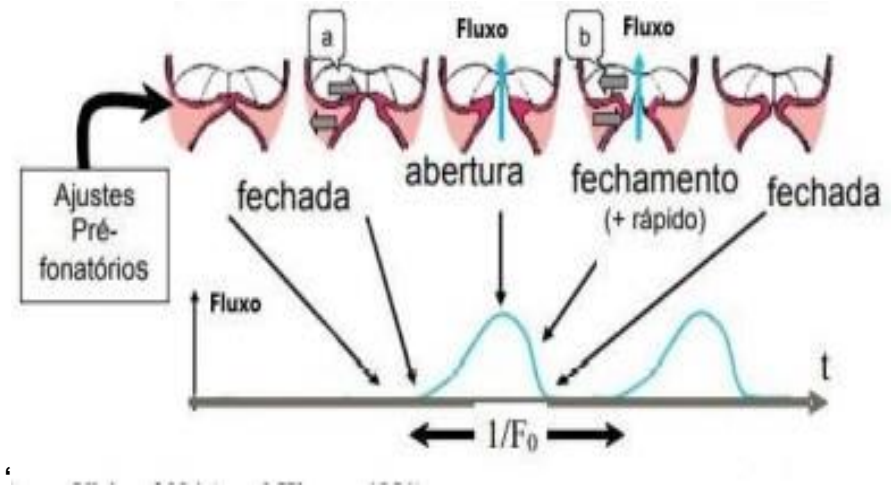
Sendo composta por cartilagens, osso hioide, músculos divididos com relação a origem e sua participação na fonação, sendo músculos intrínsecos (com origem dentro da laringe), e músculos extrínsecos ligados externamente, nervos ligamentos e lâminas de aponeurose(Razera,2004), a laringe pode ser dividida em quatro cavidade: vestíbulo ou pregas vestibulares, situado desde a epiglote até as falsas cordas vocais; os espaços entre as cordas vocais verdadeiras e falsas, são chamados ventrículos laríngeos; a glote e a cavidade infra glóticas que está localizada na extensão entre traqueia e as cordas vocais verdadeiras, sendo a última a primeira resistência aerodinâmica da laringe (Rosa,2002).

A musculatura presente na laringe é dividida segundo Rosa (2011) de acordo com a sua participação na fonação sendo eles músculos extrínsecos e intrínsecos. Os extrínsecos, aqueles que estão ligados a estrutura externa da laringe responsáveis por além de manter ela fixa também permitem elevação e abaixamento

da laringe, sendo notas mais agudas produzidas por essa elevação e notas mais graves pelo abaixamento. E os intrínsecos, estão envolvidos diretamente na fonação, sendo responsáveis por abrir e fechar o espaço glotal. Sendo esse o movimento que controla a vibração das cordas vocais.

Os movimentos de abdução, abertura das cordas vocais que permitem a passagem de ar e depois de adução, fechando as cordas vocais interrompendo o fluxo de ar proveniente dos pulmões durante a fonação gerando uma série de pulsos de ar, denominados como trem de pulsos ou sinais glotais (Cataldo, Sampaio e Nicolato, 2004). A figura 3 apresenta um esquemático mostrando a relação do movimento das cordas vocais no momento da passagem do fluxo de ar durante a fonação com a produção do sinal glotal (Rosa,1998).

Figura 3: Esquemático da produção do sinal glotal



Fonte: Vieira, 2004 (apud Hirano, 1981).

Durante algum tempo, acreditou-se que movimento das cordas vocais, que são responsáveis pela geração do sinal glotal, eram realizados por impulsos neurológicos, mas como a vibração das cordas é em frequências mais altas que os pulsos neurais, essa teoria foi descartada (Rosa,2002) fazendo com que assim o papel da ação neural seria a ativação de músculos que indiretamente contribuem para que as cordas vocais se aproximem de tal forma que elas vibrem, e regulação o nível de tensão à qual estão submetidas são as mesmas (Cataldo,Brandão,2006).

Na busca pela compreensão mais adequada de como a laringe gera o sinal, a teoria mioelástica-aerodinâmica da laringe (TMA) foi proposta (VAN DER BERG,1958, TITZE, 1980) teoria descreve que as propriedades dos tecidos dos músculos da laringe são responsáveis pelo movimento das cordas vocais.

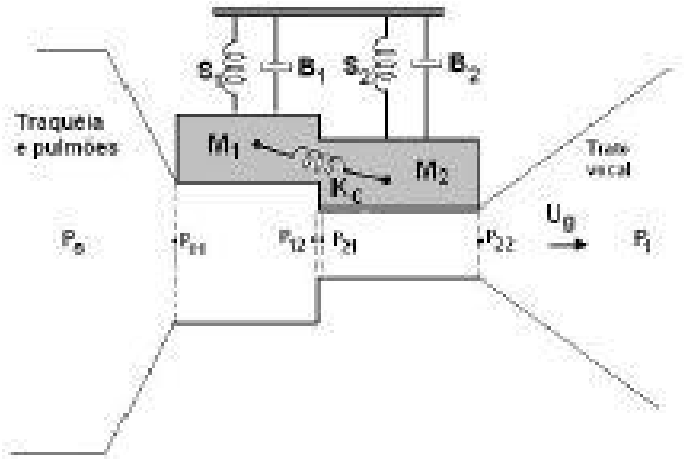
Essa teoria diz que interação da força aerodinâmica da respiração juntamente com a elasticidade dos músculos da laringe, produz o efeito Bernoulli de escoamento de ar. O ar em alta velocidade passa pela laringe, a pressão entre as cordas diminui, tendo assim a aproximação entre as cordas vocais em seguida ocorre um retrocesso elástico, fechando a glote e o ciclo de vibração recomeça. Também segundo essa teoria a pressão subglotal, as características dos tecidos das fibras das cordas, o tamanho da área glotal no momento da fonação e a tensão aplicada nas cordas verdadeiras e a massa das mesmas, são fatores que afetam a autossustentação da geração das vibrações (ROSA,2011).

Considerando como movimento das pregas descrito pela teoria mioelástica da laringe produz o sinal glotal, esse sinal se registrado através de uma modelagem matemática apropriada carrega consigo informações do sistema que o gerou, podendo ser utilizado para análises para detecção de patologias que afetam a laringe e como essas patologias afetam o sinal que posteriormente será amplificado gerando o sinal de fala (ROSA,2002).

Alguns estudos matemáticos foram desenvolvidos a fim de entender a laringe e seu processo de produção do sinal de glotal, diferenciando-se no grau de complexidade e no entendimento subjetividade do funcionamento das cordas vocais (ROSA,2002). Podemos destacar dentro desses modelos o modelo o de duas massas (ISHIZAKA E FLANAGAN,1972) que apesar de simples, consegue reproduzir satisfatoriamente tanto o movimento, quanto as características fisiológicas das cordas vocais. Nesse modelo cada uma das cordas é descrita como duas massas M_1 e M_2 em que cada uma das massas está associada a diferentes elasticidades, S_1 e S_2 e amortecimentos r_1 e r_2, que são as características viscoelásticas das cordas.

O modelo considera a dinâmica das cordas vocais como um sistema massa- mola e amortecedor com dois graus de liberdade, em que cada uma das cordas vocais é sistema de duas massas com movimentos simétricos e, portanto, equações análogas. Essas massas estão ligadas à laringe por duas molas não lineares S_1 e S_2 as massas são acopladas entre si por uma mola linear k_c.A figura 5 apresenta a ideia do de como é a configuração do modelo de duas massas.

Figura 4: Modelo mecânico de segunda ordem para as cordas vocais



Fonte: Ishizaka e Flanagan (1972)

As equações que descrevem o movimento e deslocamentos foram descritas por Ishikawa e Flanagan (1972) e foram utilizadas no sintetizador de voz (Brogian, Kruger, Palmas,2017)

As equações que descrevem o movimento e os deslocamentos são:

$$\boldsymbol{M_1\ddot{x}_1 + r_1\dot{x}_1 + S_1 + k_c(x_1 - x_2) = F_1}$$

$$M_2\ddot{x}_2 + r_2\dot{x}_2 + S_2 + k_c(x_2 - x_1) = F_2 \qquad \textbf{(1)}$$

em que:

$$s_i(x_i) = k_i(x_i + \eta_{ki}\, {x_i}^3), para\ x_i > -\frac{A_{goi}}{2l_g}$$

(2)

$$s_i(x_i) = k_i(x_i + \eta_{ki}\, {x_i}^3) + \; h_i \left\{ \left(x_i + \frac{A_{goi}}{2l_g} \right) + \eta_{ki} \left(x_i + \frac{A_{goi}}{2l_g} \right)^3 \right\} para \leq - \frac{A_{goi}}{2l_g}$$

sendo A_{g1} e A_{g2} as áreas da região existentes entre as cordas sendo a referência a linha de simetria entre elas. As molas S_1 e S_2 representam a tensão nas cordas vocais e possuem características não lineares, que a relação não linear entre a deflexão e força requerida para produzir essa deflexão é dada por:

$$\boldsymbol{f_{sj} = - k_j x_j \left(1 + \eta_{kj} {x^2}_j\right), \quad j = 1,2} \qquad \textbf{(3)}$$

sendo f_{sj} a força requerida para produzir deslocamento x_j, k_j a rigidez linear e η_{kj} descreve o comportamento não linear da mola S_j.

Considerando que no momento de colisão das massas, fechamento da glote, há duas forças no processo um que deforma das cordas vocais e uma restauradora durante o processo, que pode ser representada por uma mola S_{hj}.

A força para produzir essa deformação na massa tem a seguinte relação:

$$\boldsymbol{f_{hj} = h_j \left(x_j + \frac{A_{goj}}{2l_g} \right) \left\{ 1 + \eta_{kj} \left(x_j + \frac{A_{goj}}{2l_g} \right)^2 \right\},} \qquad \textbf{(4)}$$

Para $x_j + \frac{A_{goj}}{2l_g} \leq 0 \; j = 1{,}2.$

sendo h_j a rigidez linear, η_{kj} coeficiente positivo que representa a não-linearidade das cordas quando estão em contato, A_{goj}é a área da região existente entre as cordas vocais em repouso e l_gé o comprimento das cordas. E força resultante que age em M_j que age durante o fechamento será a soma de f_{sj} com f_{hj}.

2.3 MODELO DO TRATO VOCAL

O trato vocal é última etapa do processo de produção da fala, composto por elementos chamados articuladores essa parte do aparelho fonador se estende

desde a glote até os labios passando por língua e dentes. Funcionando como uma caixa de ressonância que atenua ou amplifica certas frequências do sinal glotal.

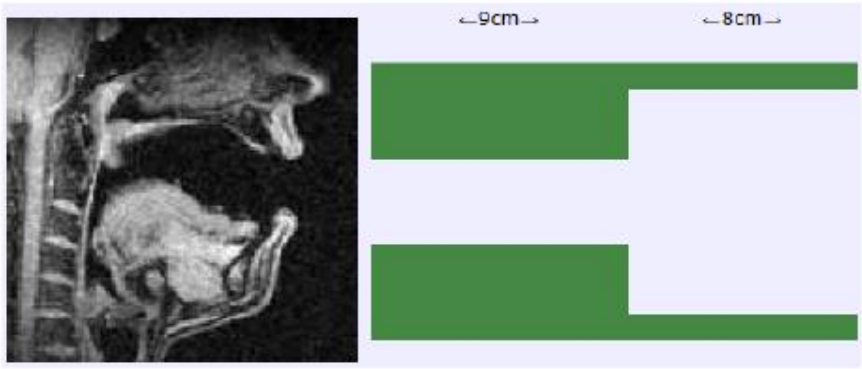
A maneira mais simples de se modelar o trato vocal segundo Granato (2005) é considerar como sendo um conjunto de n tubos cilindricos com seção transversal de área uniforme aberto de um dos lados representando os lábios e com uma fonte de excitação na outra. A figura 6 apresenta uma aproximação do trato por esses tubos.

Figura 5: Aproximação do trato vocal por tubos cilíndricos

Fonte :Adaptado de SAMPAIO; CATALDO; BRANDÃO (2006)

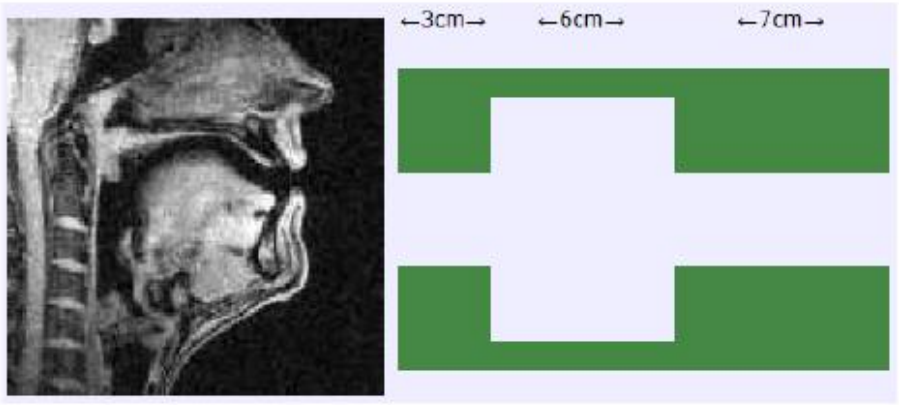
Considerando a modelagem do trato vocal por cilíndricos, as figuras 6,7,8 apresentam boas aproximações do trato vocal para os fonemas /a/, /i/ e /u/.

Figura 6: Aproximação do trato para a vogal /a/



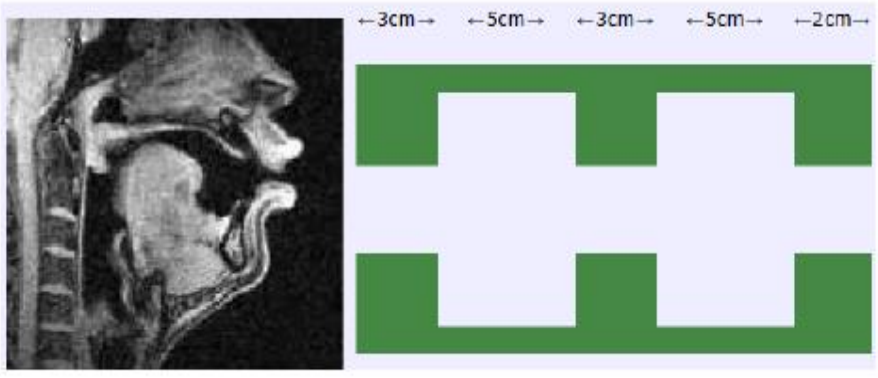
Fonte: Silva Jr (2015)

Figura 7: Aproximação do trato vocal para vogal /i/



Fonte :Silva Jr. (2015)

Figura 8: Aproximação do trato vocal para a vogal /u/

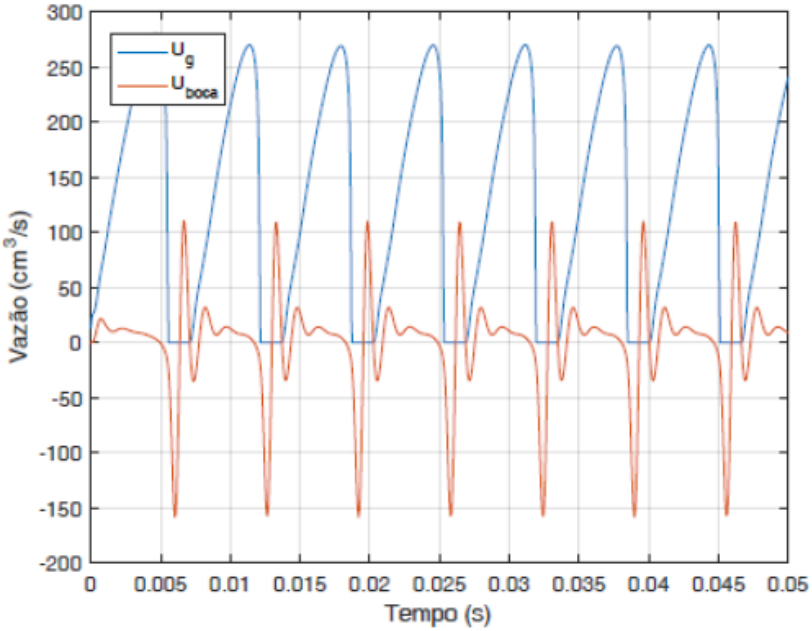


Fonte: Silva Jr (2015)

Com base nas considerações de modelagem do trato vocal apresentada, e do no modelo apresentado por Ishizaka e Flanagan (1972), é possível a síntese de sinais de voz. Em seu estudo, Brogian *et al* (2017) considerando as equações

apresentadas e realizando considerações acerca das soluções das equações que descrevem o escoamento de ar, apresentaram um modelo para síntese de voz da vogal /a/ como apresentado na figura 9.

Figura 9: Vazão sinal glotal e vazão irradiada pela boca para síntese do fonema /a/

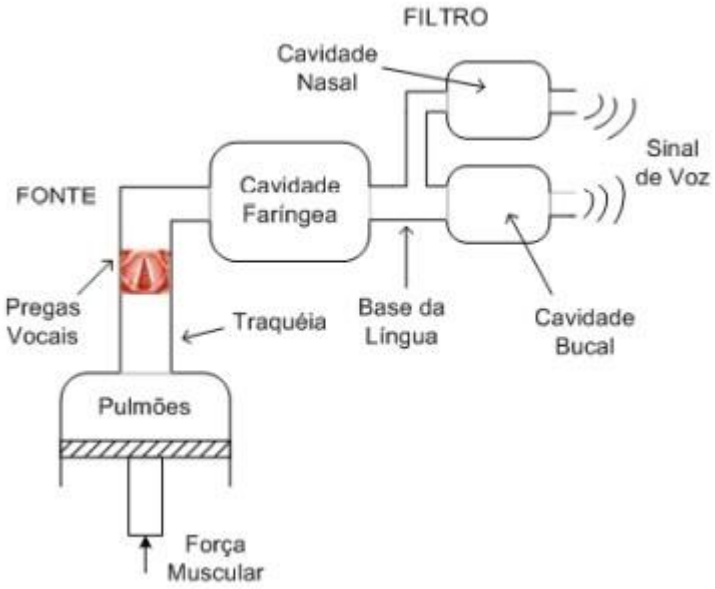


Fonte: Brogian, Palmas e Von Kruger (2017)

Nas últimas décadas, diversos métodos e técnicas têm sido desenvolvidos para a estimativa do sinal glotal a partir do sinal de voz. Muitas dessas abordagens são baseadas na teoria do filtro-fonte de Fant: o sinal glotal e a função de transferência do trato vocal são independentes e, portanto, linearmente separáveis do sinal de fala (MATTI,2008).

A teoria do filtro-fonte da produção da fala afirma que a fala pode ser descrita como um som fonte sendo modulada por um filtro que muda dinamicamente. Esta é uma simplificação da relação entre a glote e o trato vocal e implica que os sinais de fala são produzidos por um sinal fonte, representando o sinal glotal, que é modulado por uma função de transferência (filtro) determinada pela forma do trato supraglotal. Como apresentado na figura 10, que mostra um esquemático do sistema, em que o sistema subglotal (pulmões, brônquios e traqueia) são responsáveis pela geração de energia da fonação, grosso modo pode se entender como uma fonte e o trato vocal superior é considerado como filtro pois o sinal gerado pelo sistema subglotal tem sua frequência modificada de acordo a seletividade do trato superior (DAJER,2006).

Figura 10: Esquemático modelo fonte-filtro



Fonte: Dajer (2006)

A maioria dos métodos baseia-se num processo designado filtragem inversa. A filtragem inversa representa, portanto um desconvolução, ou seja, procura obter o sinal glotal aplicando o inverso da função de transferência do trato vocal ao sinal de voz. Apesar da simplicidade do conceito, o processo de filtragem inversa não é simples uma vez que o sinal de saída pode incluir ruído e não é simples modelar com precisão as características do filtro do trato vocal.

Com objetivo de se encontrar a partir dessa técnica de filtragem inversa visando encontrar um filtro que descreva o trato vocal e considerando como entradas (sinal de voz sintetizado) e saída (sinal glotal), recursos de identificação de sistemas foram utilizados sendo eles modelagem de caixa preta que emprega como ferramentas redes neurais como rede MLP e rede NARX.

2.4 IDENTIFICAÇÃO DE SISTEMAS E REDES NEURAIS

Para construir um filtro que se aplicado ao sinal de voz apresenta como resposta o sinal glotal, ou seja, um filtro que modele o trato vocal, tem-se que compreender a identificação de sistemas, área de conhecimento que realiza a modelagem matemática do sistema.

A identificação de sistemas está dividida em 3 grandes grupos, de acordo com a natureza dos elementos: caixa branca, caixa preta e caixa cinza. Na modelagem do tipo caixa branca, para modelar o sistema as leis físicas são previamente conhecidas e utilizadas durante o processo de identificação de sistema. No modelo tipo caixa preta a modelagem do sistema é realizada utilizando somente os dados dinâmicos do sistema, como sinais de entrada e saída é construído através de dados dinâmicos do sistema, tais como entrada e saída. Já no modelo caixa cinza utiliza-se tanto do modelo de caixa branca como caixa preta (Aguirre,2007).

Para esse trabalho que busca extrair o sinal glotal do sinal de fala, apesar de conhecer algumas equações físicas que descrevem partes do sistema, elas não foram utilizadas, sendo assim o modelo de caixa preta da identificação de sistemas foi utilizado sendo apenas considerado os dados de entrada (sinal de fala) e o de saída (glotal).

Numa configuração caixa preta, a ideia é parametrizar uma função dinâmica genérica, de modo que o sistema tenha sua dinâmica aproximada por tal função, desde que os parâmetros adequados sejam escolhidos (Oroski,2017). Redes neurais são um dos métodos de modelagem caixa preta.

2.5 REDES NEURAIS

Redes neurais são sistema de processamento com características em comum com as redes neurais biológicas, que formam o sistema nervoso, como a capacidade de estruturar uma base de conhecimentos a partir de processos de aprendizagem do ambiente que está inserida e modelar o conceito de pesos sinápticos para armazenar o conhecimento adquirido (HAYKIN,2001).

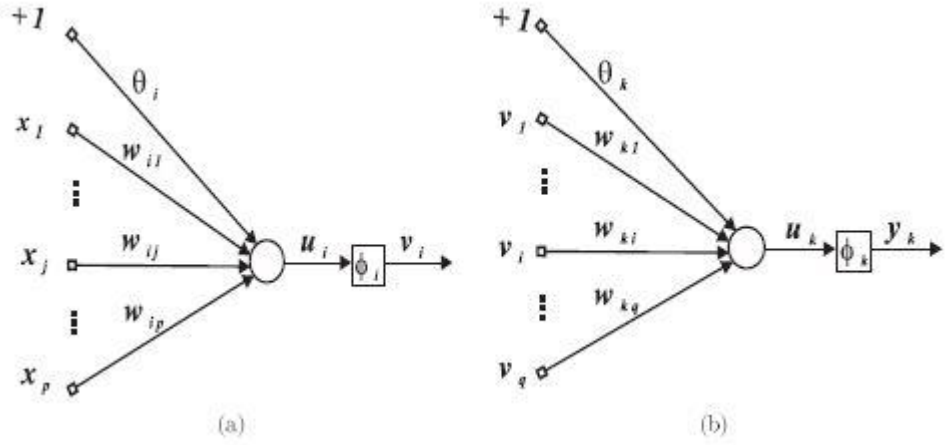
São muitas as arquiteturas de redes neurais possíveis para identificação. Redes neurais de modo geral podem ser divididas quanto ao tipo de aprendizado: aprendizado supervisionado e redes com aprendizado não supervisionado. No caso supervisionado cada entrada na rede vem acompanhada de uma saída desejada, a fim de permitir modificação dos parâmetros em função do erro entre a resposta da

rede e a saída desejada. Já os não supervisionados são os que rede neural detecta padrões e características das estatísticas do espaço de entrada.

No caso do trabalho apresentado, em que cada sinal de fala vem com um sinal glotal desejado em que a performance e o “sucesso” da rede está ligado ao erro gerado, redes de aprendizado supervisionado foram escolhidas, dentre as quais foram utilizadas as arquiteturas de MLP e NARX.

A rede MLP (*Multilayer Perceptron*) é uma rede na qual uma camada de entrada recebe os sinais, podendo existir uma ou mais camadas intermediárias compostas por neurônios somadores com função de ativação não linear e uma camada de saída também com neurônios somadores. Como na figura 12 (a) neurônios da camada escondida e (b) neurônios de saída.

Figura 11: Camada de uma rede MLP



Fonte: Menezes Junior (2016)

Seu alto grau de conectividade determinado pelas sinapses da rede, interligações entre neurônios de diferentes camadas, em que cada uma delas está associada um valor chamado peso sináptico. A figura 11 apresenta uma MLP de apenas uma camada. Assim que definido o número de camadas e neurônios da rede MLP, o processo de aprendizado é iniciado a partir de ajustes dos pesos sinápticos e limiares de ativação por meio de algoritmo de retropropagação do erro. Para cada valor de entrada apresentado a rede no instante n,o ajuste dos parâmetros da MLP acontece em duas fases: uma direta e uma reversa.

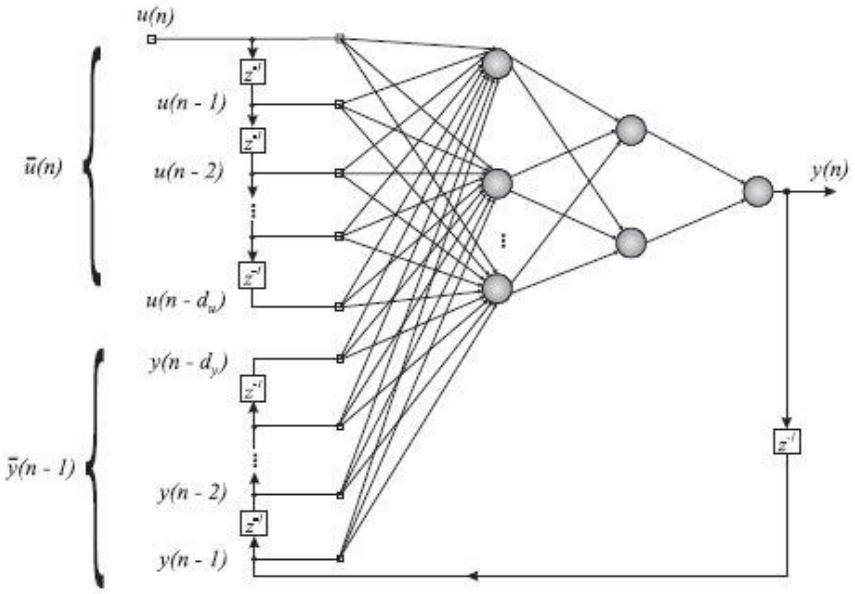
Na fase direta são calculados as ativações e saída de todos os neurônios da camada escondida e todos os neurônios da camada de saída, e na etapa reversa acontece o cálculo de gradiente e ajuste de peso de todos os neurônios da camada escondida e da camada de saída, e depois é feito o cálculo do erro entre a saída desejada para o neurônio da camada de saída e a resposta gerada é utilizada para atualizar os pesos e limiares.

Para avaliar o desempenho da rede treinada é importante avaliar a sua resposta a dados de entrada diferentes daqueles utilizados durante o treinamento, calculando-se o valor de erro médio para estes vetores. Esses dados utilizados para avaliar o desempenho são divididos em 2, um para teste e outro para validação.

Assim, tem-se três conjuntos de dados, um para treinamento, de tamanho $N_1 <$ N e outros dois de tamanho $\frac{N_2}{2}$, sendo $N_2 = N - N_1$. Em geral, escolhe-se N_1 tal que a razão $\frac{N1}{N}$ esteja na faixa de 0,75 a 0,90, ou seja, se $\frac{N1}{N}$ for aproximadamente 0,70 tem-se que 70% dos vetores de dados para serem selecionados aleatoriamente, sem reposição e utilizados durante o treinamento. Os 30% por cento restantes são usados para testar e validar a rede.

Outro modelo de rede neural útil é o modelo NARX *(Nonlinear AutoRegressive model with eXgoneous*) dividida em dois modos de identificação: paralelo e série-paralelo. O modelo paralelo, ou recorrente, a saída estimada é realimentada e incluída na saída do regressor e o modelo série paralelo a saída do regressor é formada apenas por valores atuais de saída do sistema.

Figura 12: Rede NARX com d_u entradas de d_y atrasos de saídas



Fonte: Menezes (2006)

A rede NARX pode ser descrita pela equação matemática:

$$\boldsymbol{y(n) = f(y(n-1), y(n-2), \dots, y(n-d_y), u(n-1), u(n-2), \dots, u(n-d_u))} \qquad \textbf{(5)}$$

em que $u(n)\ \epsilon\ \mathbb{R}$ e $y(n)\ \epsilon\ \mathbb{R}$ são a entrada e saída do modelo no instante n e $d_u > 0$ e $d_y > 0$ $d_u \leq d_y$ sendo d_u e d_y como atrasos na entrada e na saída. E $f(\cdot)$ é uma função não-linear desconhecida.

NARX é definida de modo que o seu regressor de entrada $U(n)$ contenha d_u amostras da variável observada $x(n)$, espaçadas de $\tau > 0$ unidades de tempo, enquanto o regressor de saída $y(n-1)$ contém valores reais ou estimativas da mesma variável, porém amostradas em instantes consecutivos. À medida que o treinamento da rede NARX avança, as estimativas $y(n) = b\,x(n+1)$ tornam se cada vez mais próximas dos valores desejados $d(n) = x(n+1)$, indicando convergência do processo de treinamento.

Assim, as saídas da rede que estão sendo realimentadas para o regressor de saída $y(n-1)$ tendem a replicar o comportamento de curto-prazo da série temporal real, pois o atraso entre as estimativas é unitário. Já o regressor de entrada

fornece informação de médio/longo prazo sobre a dinâmica da série temporal, visto que o atraso é sempre muito maior que a unidade.

3 MÉTODOS E PROCEDIMENTOS

3.1 CRIAÇÃO DOS DADOS

Para a criação de um longo par de sinais glotal e de fala, que foi utilizado para alimentar as redes neurais, utilizou-se o sintetizador de voz e sinal glotal criado por Brogian, Palmas e Von Kruger (2018), que como já explicado anteriormente, foi obtido através do modelo de duas massas de Ishizaka e Flanagan (1972).

O sintetizador para gerar os sinais de sons vocálicos possui tanto parâmetros fixos como variáveis. Os fixos são os de simulação da laringe, responsáveis para a produção do sinal glotal, apresentados na tabela 1.

Tabela 1: Parâmetros de simulação da laringe

Paramêtros	dimensão	unidade
M_1	0.125	g
M_2	0.025	g
k_1	8000	dyn/cm
k_2	8000	dyn/cm
k_c	25000	dyn/cm
Geometria da laringe		
l_g	1.4	cm
A_{g01}	0.05	cm²
A_{g02}	0.05	cm²
d_1	0.25	cm
d_2	0.05	cm

Fonte: Autoria própria (2022)

sendo l_g comprimento das cordas, A_{g01} e A_{g02} áreas glotais, d_1e d_2, M_1e M_2 massas das cordas vocais, k_1 e k_2 coeficientes de elasticidade das massas, k_c coeficiente entre as massas.

Já como parâmetros variáveis tem-se os do trato supraglotal (vocal) representado por tubos concatenados e os de pressão subglotal P_S. Os tubos concatenados que representam o trato supraglotal deste trabalho são modelados em 4 tubos (sendo um deles considerado o da boca), divididos em dimensões de 0,25 cm até 1,5 cm de raio e 4 cm de comprimento cada tubo. Juntamente com a

variação dos raios dos tubos, a pressão do pulmão também foi variada de 5000 a 8000 cm/H_2O. A tabela 2, apresenta os valores utilizados para a combinação dos parâmetros de modo a gerar 1024 combinações diferentes de tubos e pressão.

Tabela 2: Parâmetros dos tubos do trato supraglotal e pressão

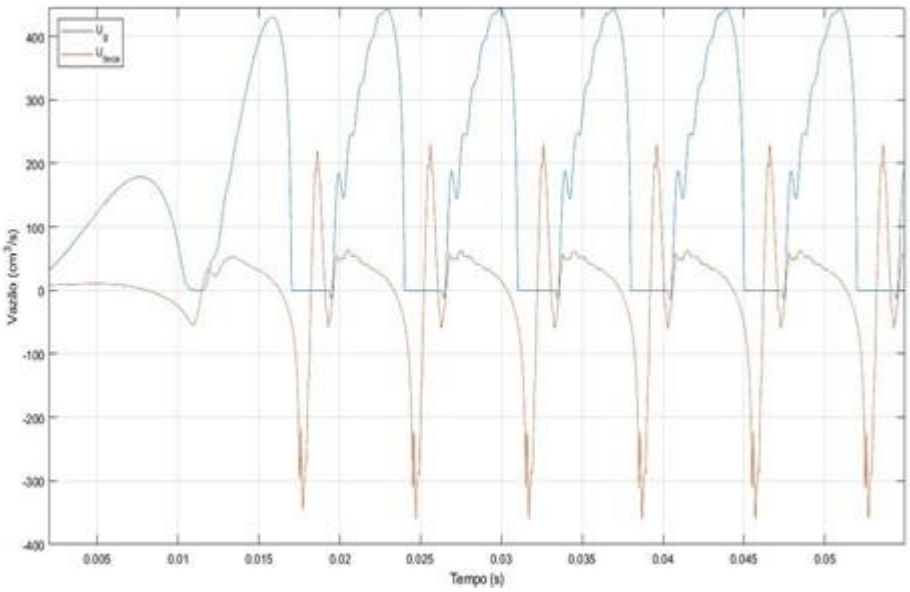
Parâmetros	Valores				Unidade
Pressão	5000	6000	7000	8000	cm/H2O
Tubo 1	0,25	0,667	1,083	1,50	cm
Tubo 2	0,25	0,667	1,083	1,50	cm
Tubo 3	0,25	0,667	1,083	1,50	cm
Tubo 4	0,25	0,667	1,083	1,50	cm

Fonte: Autoria própria (2022)

Dois exemplos de combinações geradas com os valores da tabela 2 é um trato supraglotal com 4 tubos sendo o tubo 1 de 0,25cm, o tubo 2 de 0,667 cm, o tubo 3 de 1,5 cm e o tubo 4 de 1,083 cm e pressão de 6000 cm/H_2O e outro trato supraglotal com 4 tubos sendo o tubo 1 de 0,667 cm, o tubo 2 com 1,083 cm, tubo 3 com 0,25cm e o tubo 4 com 0,667 cm e pressão de 8000 cm/H_2O.

Para cada combinação dos parâmetros descritos anteriormente, gera-se um par de sinais glotal e de fala que foram utilizados para teste do ajuste de parâmetros das redes neurais MLP e NARX. Esse par foi inserido na rede como sinal de fala sendo a entrada x(t) e o sinal glotal como o de saída y(t). A figura abaixo apresenta o resultado da síntese de algum dos parâmetros resultando na vazão glotal.

Figura 13:Sinal glotal e de voz sintetizados



Fonte: Autoria própria (2022)

Em seguida da criação desse par de sinais, inicia-se a configuração no MATLAB da rede neural para treinamento, teste e validação.

Para a configuração da rede MLP, utiliza-se uma rede similar a *feedforward,* em que a rede só possui uma direção e sem realimentação, em que os parâmetros de entrada possuem uma linha de atraso associada permitindo com que a rede tenha uma resposta dinâmica aos dados de entrada da série temporal. Modificados esses atrasos da entrada avalia-se o desempenho da rede de acordo com os resultados apresentados.

Uma característica da rede MLP é existência de camadas intermediárias que não se ligam diretamente com a camada de saída. Tais camadas de neurônios são chamadas ocultas que possibilitam a rede aprender tarefas complexas extraindo progressivamente as características mais significativas dos padrões da entrada. No estudo realizado com as redes utilizou-se apenas uma camada oculta, variando apenas o número de neurônios dessa camada oculta.

Configurada a rede com relação aos tamanhos das entradas e o número de neurônios da camada oculta, foi escolhida a rotina de treinamento da rede, que é

responsável pelo processo de aprendizagem. O processo de aprendizagem de uma RNA se dá através do ajuste dos pesos da rede (SANTANA,2012).

O algoritmo utilizado para o treinamento da rede neural deste estudo é um algoritmo de retropropagação chamado *Scaled conjugate Gradient Backpropagation* (SCG) esse algoritmo faz parte do conjunto de algoritmos de treinamento supervisionado. Esse algoritmo possui maior convergência, pois segue na direção de descida mais íngreme, ou seja, segue na direção do valor mais negativo do gradiente. Essa direção é chamada conjugada e o tamanho do passo é ajustado a cada iteração (BARROS,2018).

A rede utiliza dados separados previamente para realizar o teste da rede e depois validação no caso da rede treinada nesse trabalho a divisão de treinamento foi de 70% dos dados para treinamento 15% para teste e 15% para validação. A definição dos dados foi feita de maneira aleatória pelo *software.*

Os critérios para avaliar a rede foram o desempenho, a magnitude do gradiente de desempenho e o número de verificações de validação. A magnitude do gradiente e o número de verificações de validação são usados para encerrar o treinamento. O gradiente se tornará muito pequeno à medida que o treinamento atingir um mínimo do desempenho. O desempenho da rede é calculado pelo MSE (erro médio quadrático), descrito pela equação:

$$\boldsymbol{MSE = \frac{1}{n}\sum_{1}^{n}(saída\ esperada_n - sinal\ gerado\ pela\ rede_n)^2} \qquad \textbf{(6)}$$

O critério de parada da rede foi o número de iterações sucessivas em que o desempenho da rede não diminui, atingindo esse número a rede era finalizada. Esse número de iterações escolhido foi o valor padrão do MATLAB de 6 iterações. Outro critério para finalização do treinamento foi a magnitude do gradiente, se esse valor for menor que $1e^{-6}$ a rede é finalizada. O número de épocas (iterações) da rede foi ajustado para que se atingir o valor de 1000 a rede é finalizada, tornando-se mais um dos critérios de parada utilizados.

Para rede Neural NARX configuração da rede tanto no algoritmo de treinamento, análise de desempenho e critérios de parada foram os mesmos que se utilizou para configuração da rede MLP. O ponto diferencial dessa rede para a MLP é que a rede NARX possui realimentação, isso é além de atrasos na entrada a rede possui atraso na saída que realimenta a rede. Os valores de atraso na entrada e saída foram variados de maneira gradativa com objetivo de avaliar o desempenho da rede a essas variações dos atrasos. Para rede NARX foi utilizado também apenas uma camada oculta e variando-se apenas o número de neurônios dessa camada.

Realizados os treinamentos das configurações, as redes e seus respectivos resultados de performance foram salvos no software e tabelados de forma a facilitar a visualização dos desempenhos das configurações testadas afim de evidenciar os melhores e piores resultados, e fazer análise da relação do resultado com o aumento do valor de atraso e aumento do número de neurônios na camada oculta. Esses resultados são discutidos no capítulo a seguir.

4 RESULTADOS E DISCUSSÕES

Nesse capítulo são apresentados os resultados das configurações dos parâmetros das redes neurais explicitados no capítulo anterior. De modo a facilitar a compreensão, e visualização dos resultados eles são apresentados em tabelas, evidenciando-se os melhores e piores resultados. Sendo o pior resultado, marcado na tabela com vermelho e o melhor com a cor verde.

Além da análise da rede de acordo com os valores de performance atrelado ao erro médio quadrático foi realizada uma avaliação qualitativa do resultado, observando a resposta do sistema em forma curva, para analisar se o sinal gerado na saída da rede se aproxima do sinal glotal esperado.

4.1 REDE MLP

Nas redes MLP, como não há métodos para determinar a configuração da rede, partiu-se de redes com pequenos atrasos na entrada e baixa quantidade de neurônios na camada oculta e aumentado gradativamente esses valores até uma rede com configuração de atraso na entrada e neurônios na camada oculta com valor de 20. Essa variação dos valores de atraso na entrada e neurônios na camada oculta e as performance de cada rede pode ser visualizada na tabela 3.

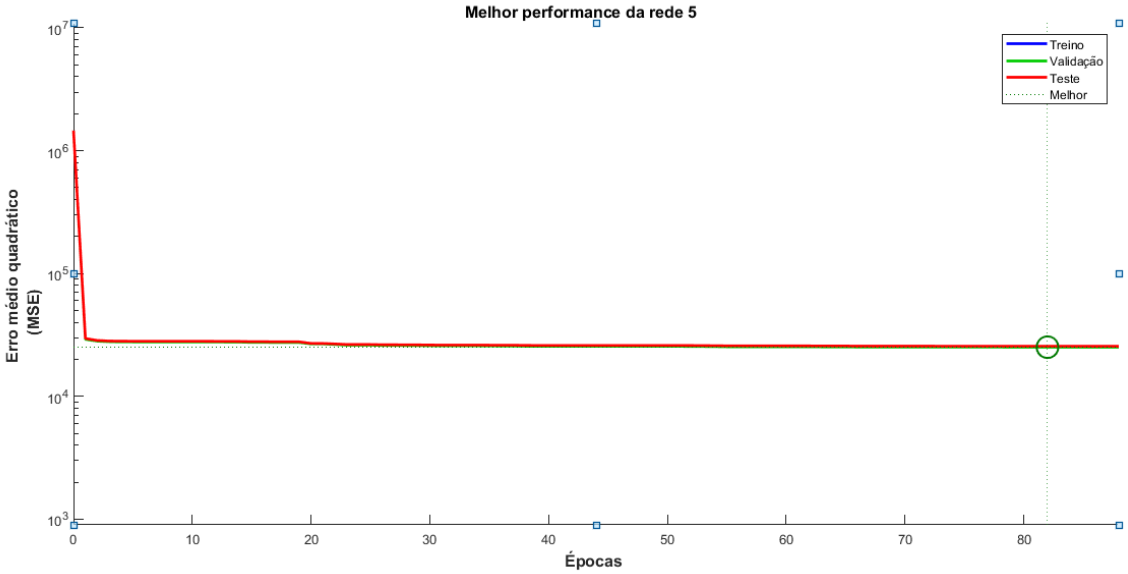
Tabela 3: Valores de Rede MLP e suas performances

REDES MLP	Atrasos da entrada	Neurônios na camada oculta	Performance (MSE)
Rede 1	2	2	2,407E+04
Rede 2	2	5	2,444E+04
Rede 3	2	10	2,509E+04
Rede 4	2	20	2,437E+04
Rede 5	5	2	2,516E+04
Rede 6	5	5	2,484E+06
Rede 7	5	10	2,423E+04
Rede 8	5	15	2,392E+04
Rede 9	5	20	2,308E+04
Rede 10	10	2	2,114E+04
Rede 11	10	5	2,203E+04
Rede 12	10	10	2,178E+04
Rede 13	10	15	2,127E+04
Rede 14	10	20	2,331E+04
Rede 15	15	10	1,800E+05
Rede 16	15	15	1,776E+04
Rede 17	15	20	1,790E+05
Rede 18	20	20	1,460E+04

Fonte: Autoria Própria (2022).

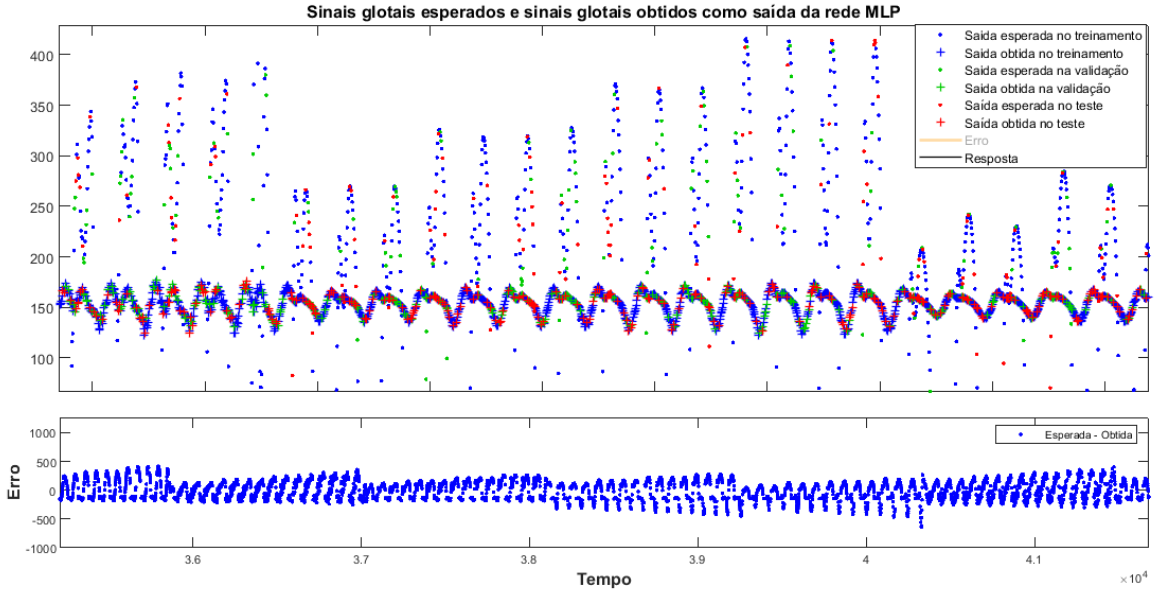
O *software* MATLAB apresenta recursos visuais para acompanhamento do treinamento da rede, como a curva de performance apresentada na figura 14 que mostra variação dos erros médios quadráticos e o menor valor do erro para a rede 5 que apresentou pior desempenho dentre as redes treinadas.

Figura 14: a) Performance da rede MLP 5



Fonte: Autoria própria (2022)

Figura 15: Curva de resposta apresentada pela rede MLP 5

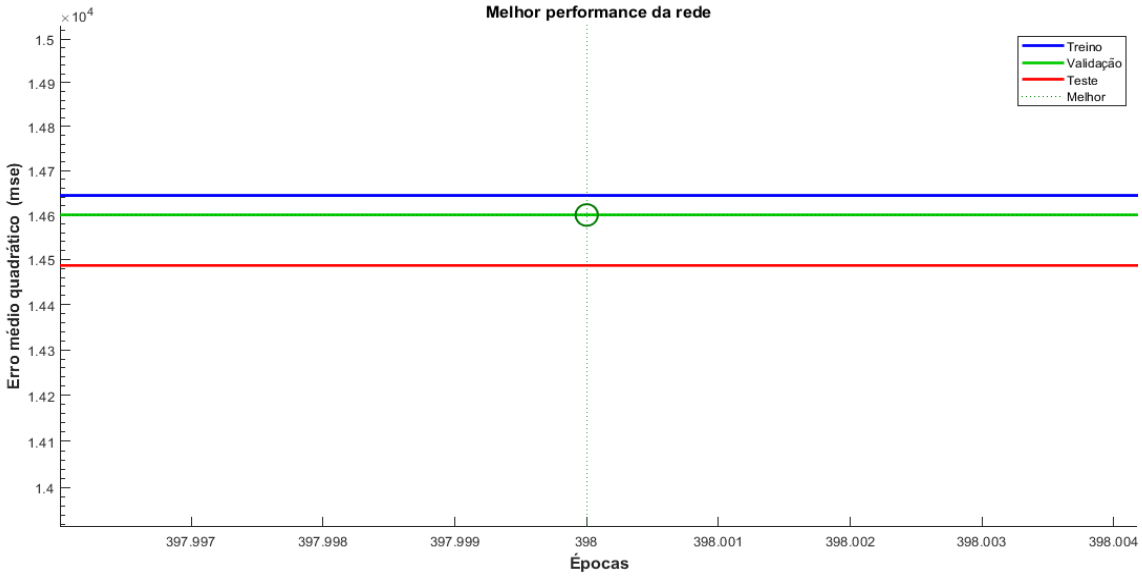


Fonte: Autoria própria (2022).

Pode-se observar na figura 15 que a saida resposta da rede treinada se comparada a saída esperada, ou seja o sinal glotal, apresenta grandes discrepâncias.

Já a rede 18 apresentou no treinamento da rede MLP o menor valor para erro médio quadrático, e por esse motivo foi considerada dentre as redes testadas a que oteve melhor desempenho que pode ser observado na figura 16 que apresenta a curva dos erros médios quadráticos da fase de treino, teste e validação.

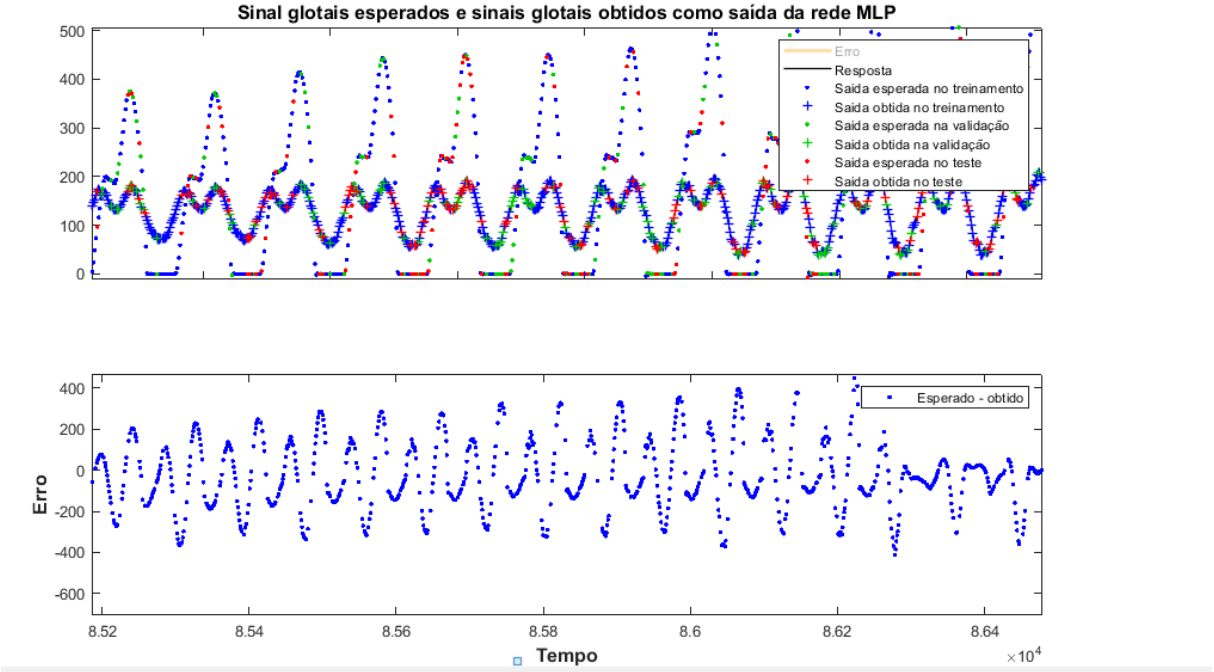
Figura 16: Curva de performance da rede MLP 18



Fonte: Autoria própria (2022)

Observando qualitativamente a curva de resposta da rede 18 apresentada na figura 17, pode-se verificar que a resposta da rede se assemelhou mais ao sinal glotal esperado do que curva gerada pela rede 5.

Figura 17: Resposta de saída da rede MLP 18



Fonte: Autoria Própria (2022)

4.2 REDES NARX

Após a simulação e teste das arquiteturas de MLP, procedeu-se com a criação de redes NARX, de acordo com os critérios apresentados no capítulo anterior. A tabela 4 apresenta as redes criadas e seus desempenhos baseados no valor do erro médio quadrático, quanto menor o erro apresentado melhor é o desempenho da rede.

Já de antemão pode-se observar que redes com maior valor de neurônios na camada oculta tendem a produzir melhores resultados do que as com menos neurônios, isso se deve a maior sinapse da rede fazendo com que o erro propagado diminua a cada iteração dos neurônios da rede. Outra questão que é que os atrasos na camada de entrada também influenciam a performance da rede conforme se observa na rede 2, que quanto menor o atraso na entrada e no número de neurônios na camada maior é o valor do erro, sendo assim pior o desempenho da rede treinada.

Tabela 4: Variação de parâmetros redes NARX

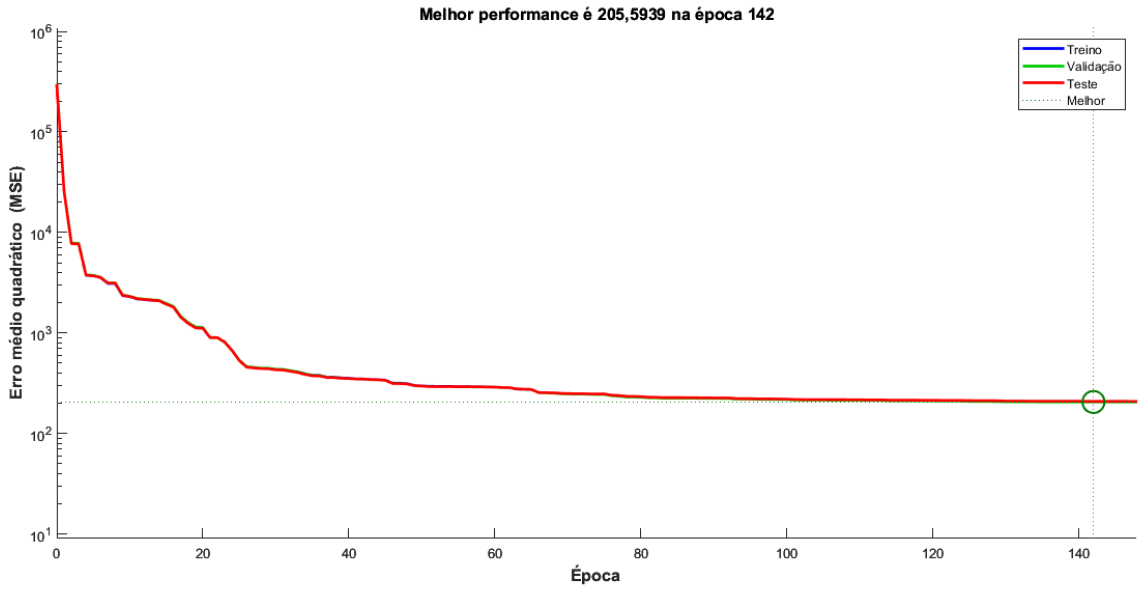
REDES NARX	Atraso na entrada	Atraso na saída	Neurônios camada oculta	Performance (MSE)
Rede 1	2	2	2	202,521
Rede 2	2	5	2	205,594
Rede 3	2	5	5	196,495
Rede 4	2	5	10	200,209
Rede 5	3	5	10	176,597
Rede 6	5	5	10	170,987
Rede 7	5	10	10	179,589
Rede 8	5	10	15	141,352
Rede 9	5	15	15	136,792
Rede 10	10	10	5	169,667
Rede 11	10	5	10	195,219
Rede 12	10	5	5	158,096
Rede 13	10	10	10	161,428
Rede 14	10	10	15	197,623
Rede 15	10	15	15	126,803
Rede 16	15	15	15	144,227

Fonte: Autoria própria (2022)

De acordo com a tabela 4 apresentada a rede 2 grifada de vermelho pois ela apresenta o maior erro sendo assim é considerada a pior rede dentro das

configurações apresentadas. Já grifada em verde a rede 15 que apresentou o melhor desempenho, sendo assim o menor erro médio quadrático (MSE). Essa rede apresenta valores de atraso de saída e número neurônios grandes se comparados as redes testadas, com base nas observações realizadas na tabela 4 verifica-se que na rede neural NARX quanto maior o valor do atraso e o número de neurônios menor é o erro. A figura 20 apresenta o desempenho da rede configurada com pior desempenho, rede essa que valida a relação também de valores pequenos de atrasos e neurônios aumentam o valor do erro.

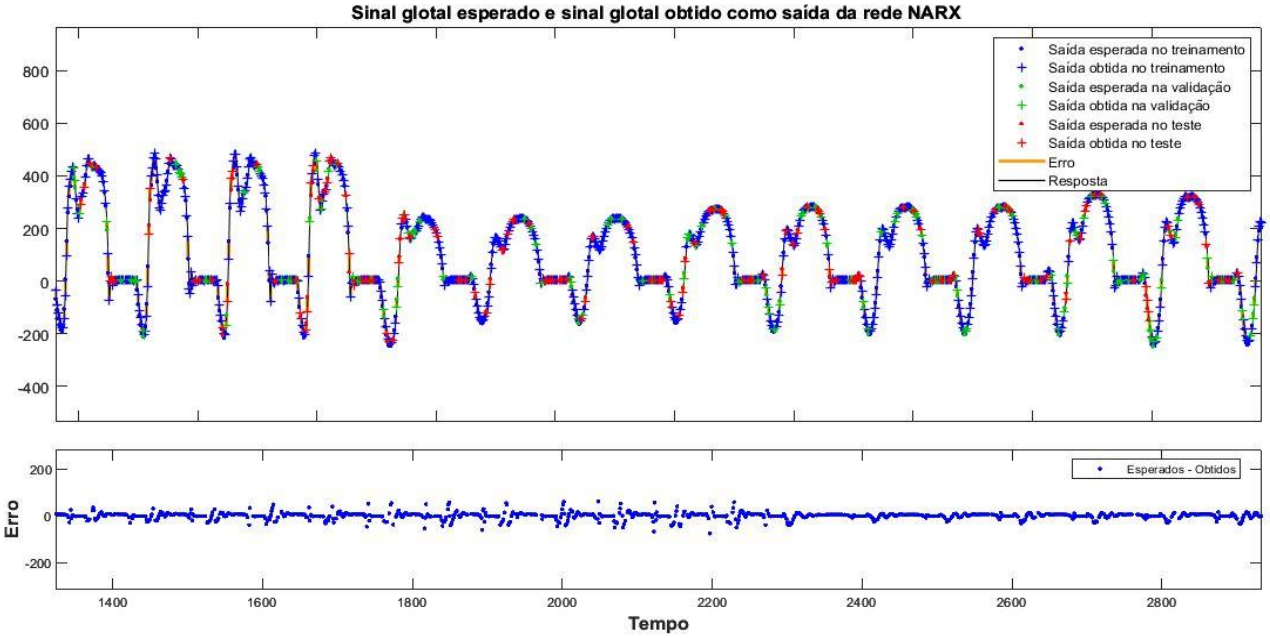
Figura 18: Curva de desempenho Rede NARX 2



Fonte: Autoria Própria (2022)

Apesar de apresentar um erro alto na tabela 4, pode se observar a partir da figura 19 que a rede NARX 2 apresenta visualmente uma curva que já se aproxima do sinal glotal esperado e que os pontos de saída esperada e obtida não são mais tão dispersos quanto os observados no modelo de rede neural MLP. Também vale destacar na figura que a mudança de amplitude do sinal se dá devido a variação de tamanhos tubos do trato supraglotal, explicado no capítulo 3.

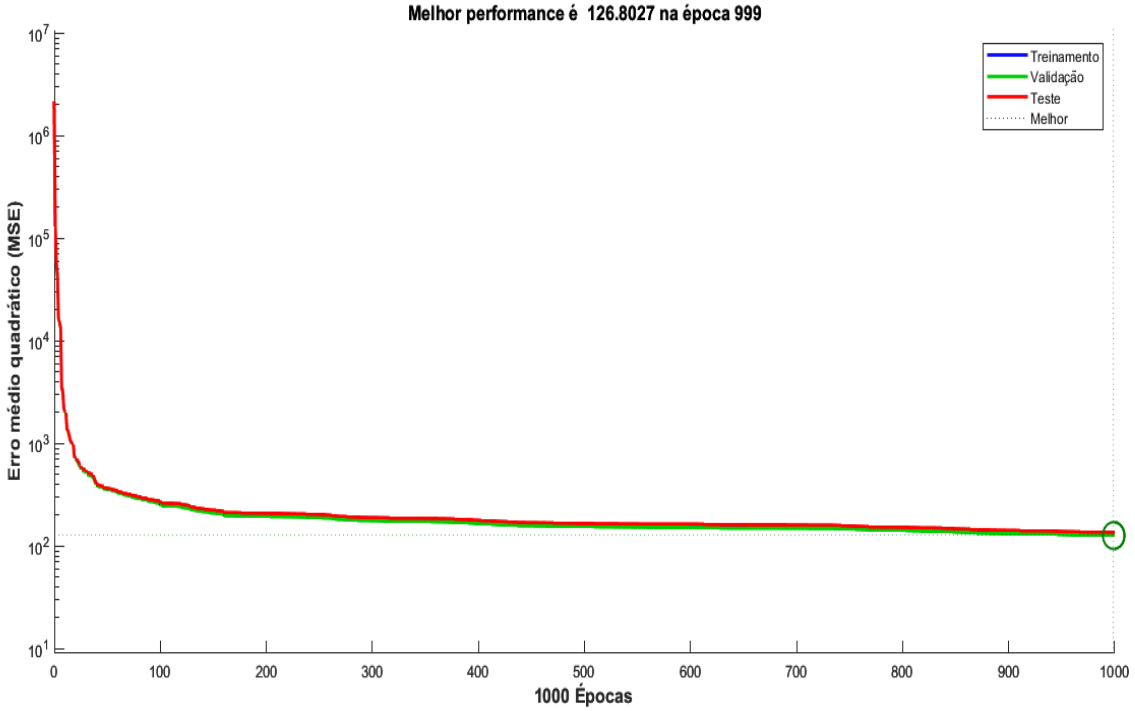
Figura 19: Resposta rede NARX 2



Fonte: Autoria Própria (2022)

E por fim tem-se a rede NARX de número 15 que apresentou melhor desempenho dentre as redes treinadas.A figura 20 apresenta a curva de desempenho dessa rede, vale destacar que a rede foi a única dentro das apresentadas nesse estudo que teve sua rotina de treinamento finalizada por atingir o número máximo de épocas estabelecido como critério de parada, 1000 épocas.

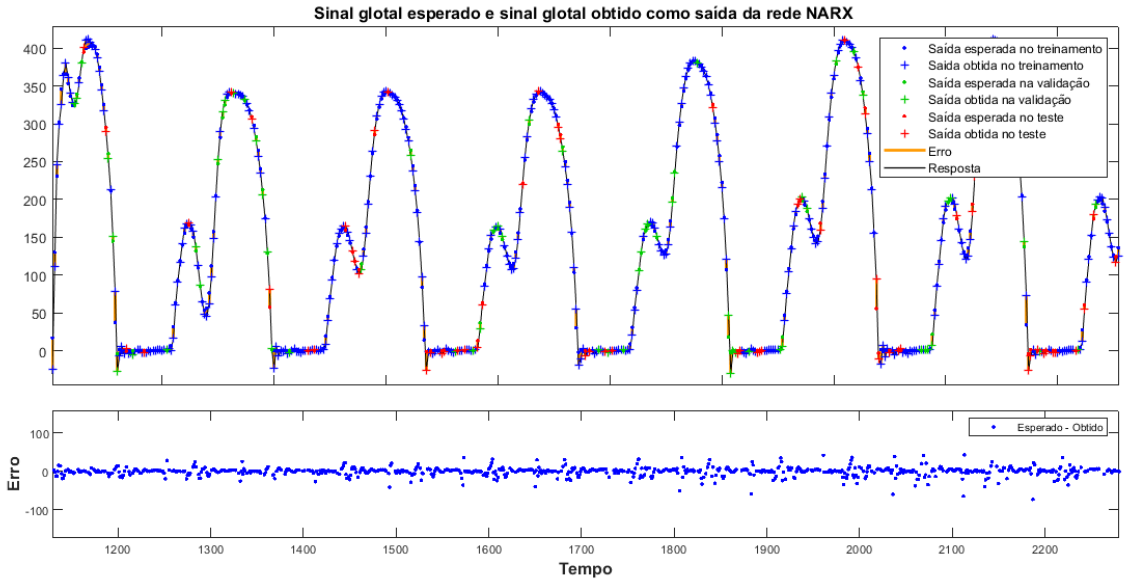
Figura 20: Performance da rede NARX 15



Fonte: Autoria Própria (2022)

Por fim tem-se na figura 21, a resposta da rede Neural 15 em que se verifica que essa foi a rede que teve o resultado mais aproximado do sinal glotal esperado.

Figura 21: Resposta da rede NARX 15



Fonte: Autoria Própria (2022)

5 COMPARAÇÕES DAS REDES

Assim que finalizado testes das configurações definidas para cada uma das arquiteturas, pode-se comparar os resultados. Realizado a razão entre os resultados da rede 18 da arquitetura MLP, com melhor de desempenho tendo MSE=1,460E+04 e da rede 15 da arquitetura NARX, considerada a rede com melhor desempenho tendo MSE= 126,803 tem se que a rede NARX possui um resultado 115 vezes melhor. Como observado na equação:

$$razão\ entre\ melhor\ resultado = \frac{Melhor\ desempenho\ rede\ MLP}{Melhor\ desempenho\ rede\ NARX} \qquad (7)$$

6 CONCLUSÃO

Neste trabalho buscou-se avaliar a capacidade das arquiteturas de rede escolhidas de realizar o aprendizado que utilizando os dados de sinal de fala e glotal sintetizados e apresenta-se como resposta um sinal glotal, que com base nos parâmetros utilizados apresentou um resultado razoável.

Com base nesse resultado apresentados pelas redes testadas e a comparação entre as redes, conclui-se que a rede neural de NARX é a rede mais indicada para estudos futuros que visem identificar a relação do sinal de fala com o sinal glotal isso porque ao utilizar realimentação com os sinais da saída atrasados para atualizar os pesos, a rede aumenta o aprendizado fazendo com que o erro diminui a cada iteração fazendo com que os resultados sejam melhores e mais rápidos que o da rede MLP.

Para trabalhos futuros podem ser observados duas considerações importantes para a melhoria dos resultados. Sendo a primeira consideração o refinamento do tubo glotal realizando aumentando número de seções no tubo afim de suavizar o formato, representando assim um modelo mais realístico do trato supraglotal.

Outra consideração seria a utilização de mais camadas ocultas juntamente com o aumento dos neurônios da camada oculta, pois como foi observado as redes com mais neurônios, em ambas as arquiteruras apresentou erros menores, sendo assim melhor desempenho.

REFERÊNCIAS

AGUIRRE, L. A. **Introdução à Identificação de Sistemas - Técnicas Lineares e Não Lineares Aplicadas à Sistemas Reais**. Belo Horizonte MG: Editora UFMG, 2007.

BARROS, Victor Pedroso Ambiel**. Avaliação do desempenho de algoritmos de retropropagação com redes neurais artificiais para a resolução de problemas não-lineares.** 2018. 136 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2018.

BROGIAN, Georgya; PALMAS, Jacqueline; KRUGER, Vanessa Christine von. **Síntese de sinais de voz usando modelos biologicamente inspirados**. 2018. 78 f. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica) - Universidade Tecnológica Federal do Paraná, Curitiba, 2018.

CATALDO, E; SAMPAIO, R; NICOLATO, L. **Uma discussão sobre modelos mecânicos de laringe para síntese de vogais**. ENGEVISTA, v. 6, n. 1, p. 47-57, abr. 2004.

CATALDO, E; LUCERO, J.C; NICOLATO, L; SAMPAIO, R**. Comparison of Some Mechanical Models of Larynx in the Synthesis of Voiced Sounds**. 2006. J. Braz. Soc. Mech. Sci. & Eng. vol.28 no.4 Rio de Janeiro Oct./Dec. 2006

DAJER, Maria Eugenia. **Análise de sinais de voz por padrões visuais de dinâmica vocal.** Tese (Doutorado em Engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2010.

HAYKIN, Simon. **Redes neurais: princípios e prática**. trad. Paulo Martins Engel. - 2.ed. -Porto Alegre: Bookman, 2001.

HAGAN, M. T. et al. **Neural network design**. Boston: Pws Pub., 1996.

ISHIZAKA, K; FLANAGAN, J. L. **Synthesis of Voiced Sounds From a Two- Mass Model of the Vocal Cords**. The Bell System Technical Journal. Vol. 51. N. 6. July-August, USA, 1972.

MENEZES JÚNIOR, J. M. P. **Redes neurais dinâmicas para predição e modelagem não-linear de séries temporais.** 2006. 116 f. Dissertação (Mestrado em Engenharia de Teleinformática) – Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2006.

OROSKI, E. (2015). **Identificação de Sistemas Não Lineares Utilizando Modelos NARX, Funções Ortonormais e Otimização Heurística**, Tese de Doutorado em Engenharia de Sistemas Eletrônicos e Automação, Publicação PGEA.DM-104/15 TD, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 124p.

RABINER, L. R.; SCHAFER, R. W. **Digital processing of speech signals**. Prentice Hall, Englewood Cliffs, NJ, EUA, 1978.

ROSA, Marcelo de Oliveira. **Análise acústica da voz para pré-diagnóstico de patologias da laringe**. 1998. Dissertação (Mestrado em Engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 1998. doi:10.11606/D.18.1998.tde-11122015-144509.

ROSA, Marcelo de Oliveira. **Laringe digital**. 2002. Tese (Doutorado em Engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2002. doi: 10.11606 /T.18.2002.TDE-19112015-110520.

ROSA, Marcelo de Oliveira. **Modelagem da Laringe: da biologia ao computador.** 2011. Revista de Letras. V.30, 1/4, 70-81, jan.2010/dez.2011.

SAMPAIO, R.; CATALDO, E.; BRANDÃO, A. **Análise e processamento de sinais.** Sociedade Brasileira de Matemática Aplicada e Computacional. São Paulo, 2006.

SANTOS, Juliana F. **Reconhecimento de Padrões em medidas acústicas para identificação de patologias da Laringe**. Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica), Universidade Tecnológica Federal do Paraná, Curitiba, 2016.

SCALASSARA, Paulo Roberto. **Utilização de Medidas de Previsibilidade em Sinais de Voz para Discriminação de Patologias de Laringe.** Tese (Doutorado em Engenharia Elétrica) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2009.2

SILVA Jr., Celso Donizete. **Modelagem Acústica do Trato Vocal Humano Pelo Método dos Elementos de Contorno.** 2015. Projeto de Graduação em Engenharia Mecânica, Faculdade de 78 Tecnologia da Universidade de Brasília, Brasília, 2015.

SOUZA, Felipe Maraschin Pereira de. **Aplicação de uma rede neural artificial NARX para obtenção do comportamento dinâmico de um atenuador de impacto de alumínio do tipo honeycomb.** 2019. Programa de pós-graduação em Engenharia Mecânica, Universidade Federal da Paraíba, Paraíba, 2019.

TITZE, Ingo. R. **Principles of voice production.** 1994. Prentice-Hall, NJ: Englewood Cliffs, NJ.