

Redes Neurais Artificiais para classificação de patologias vocais

Manoela Rabello Kohler

Redes Neurais Artificiais para classificação de patologias vocais

Aluno: Manoela Rabello Kohler

Orientadora: Marley Vellasco

Trabalho apresentado com requisito parcial à conclusão do curso de Engenharia de Computação na Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil.

Resumo

Kohler, Manoela. Vellasco, Marley. Redes neurais artificiais para classificação de patologias vocais. Rio de Janeiro, 2011. 32p. Projeto Final II – Departamento de Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro.

O propósito desse trabalho é classificar doenças de voz através de uma Rede Neural Artificial. Para isso serão utilizados padrões de voz de uma base com vozes normais e com as seguintes patologias nas cordas vocais: nódulos ou paralisia unilateral. Serão extraídos os coeficientes glotais das vozes da base pelo método de filtragem inversa para fazer a classificação. A rede neural treinada anteriormente será implementada em um microcontrolador, para o qual serão introduzidas novas amostras de voz.

Palavras-chave

Redes neurais artificiais, classificação de padrões, microcontrolador, doenças da voz.

Abstract

Kohler, Manoela. Vellasco, Marley. Artificial neural networks for classification of vocal pathologies. Rio de Janeiro, 2010. 32p. Projeto Final II – Departamento de Engenharia Elétrica. Pontifícia Universidade Católica do Rio de Janeiro.

The purpose of this work is to use Artificial Neural Networks for pattern classification of voice obtained through acoustic measures done over digital voice signals. Voice pathologies (nodules on the vocal cords and unilateral vocal cord paralysis) and normal voice (no pathologies) will be classified by the Neural Network. The network will be programmed in a microcontroller to introduce new samples of voice that we want to classify either healthy, or with one of the pathologies above mentioned.

Keywords

Artificial neural networks, pattern classification, microcontroller, voice diseases.

Sumário

1	Introdução.....	7
1.1	Motivação e Domínio do Sistema	7
1.2	Objetivos	7
1.3	Base de Dados	8
1.4	Ambiente computacional.....	10
1.5	Organização.....	10
2	Redes Neurais Artificiais.....	11
2.1	Classificação por redes neurais	12
3	A voz.....	14
3.1	O sinal de voz	14
3.2	Doenças da voz.....	15
3.2.1	Nódulos	15
3.2.2	Paralisia das pregas vocais	16
3.2.3	Paralisia Unilateral	17
3.2.4	Paralisia Bilateral.....	18
4	Sinal Glotal	20
4.1	Parâmetros do sinal glotal	20
4.1.1	Parâmetros do domínio do tempo.....	21
4.1.2	Parâmetros do domínio da frequência	22
4.1.3	Parâmetros que representam variações e perturbações na frequência fundamental.....	23
5	Microcontroladores	25
5.1	Aplicações	25
5.2	PIC 18F452.....	25
6	Plano de Ação.....	27
6.1	Estudos iniciais	27
6.2	Preparação da base de dados	27
6.2.1	Audacity	27
6.2.2	Aparat e Praat	27
6.2.3	Matlab	28
6.3	WEKA.....	28
6.4	Matlab	28
6.5	Rede em linguagem C.....	29
6.6	MikroC.....	29

6.7	Proteus	29
6.8	Kit PicGenios.....	29
7	Análises e Resultados.....	31
7.1	Análise dos parâmetros para classificação	31
7.2	Análise da implementação do sistema no microcontrolador	33
7.3	Análise da distribuição dos dados.....	33
7.4	Resultado da classificação dos parâmetros do sinal glotal	34
8	Cronogramas	35
8.1	Cronograma Realizado	35
9	Conclusões e Trabalhos Futuros	38
9.1	Conclusões	38
9.2	Trabalhos Futuros	38
10	Referências.....	39

1 Introdução

1.1 Motivação e Domínio do Sistema

Este estudo trata do uso de recursos tecnológicos de Inteligência Artificial, em específico, Redes Neurais Artificiais, aplicadas ao diagnóstico de doenças da voz baseado em dados obtidos através de gravações de pacientes com nódulo e paralisia vocal.

A utilização de ferramentas automáticas para auxiliar o diagnóstico médico tem a vantagem de colocar os melhores conhecimentos disponíveis ao alcance de usuários com pouco treinamento específico na área para o rápido diagnóstico e tratamento da doença. Outra vantagem desejável é que estes instrumentos automáticos possam ser facilmente acessados pelos interessados, de forma que mais doutores não especialistas no assunto e pacientes possam diagnosticar doenças da voz, e utilizar este estudo como base para novas pesquisas na área.

Além disso, se tem como objetivo, deixar de lado o uso de técnicas invasivas para a detecção de patologias vocais como a videolaringoscopia - exames esses que são dispendiosos, de alto custo e podem causar mal estar ao paciente - substituindo tal técnica pela extração do sinal glotal através de algoritmos que fazem a filtragem inversa do sinal de voz.

1.2 Objetivos

- O objetivo principal desse trabalho é classificar doenças da voz com os parâmetros do sinal glotal, através de uma rede neural que será implementada em um microcontrolador.
- Utilizar, aprimorar e complementar conhecimentos adquiridos durante o curso de graduação.

Para isso, dividimos o objetivo principal em etapas a serem alcançadas ao longo de dois semestres:

- Obter os parâmetros através da filtragem inversa das amostras de voz, parâmetros esses, que servirão de entrada para a rede neural a ser implementada.
- A rede neural artificial será modelada para classificar os seguintes padrões de voz: normal, com nódulos e com paralisia unilateral das cordas vocais. Depois de treinada, a rede deverá conseguir interpolar com sucesso todas as novas entradas (nunca apresentadas no treinamento) entre as entradas dos seus respectivos grupos.
- Ao obter uma rede que classifique com precisão os novos padrões apresentados, a

mesma será reproduzida em linguagem de programação C para ser colocada em um microcontrolador, que possibilitará a introdução de novas amostras de voz através de um microfone acoplado ao microcontrolador.

- Objetiva-se criar uma forma alternativa, que não seja invasiva, nem dispendiosa para diagnosticar doenças vocais. Forma essa, que possibilitará a qualquer um, sem necessidade de conhecimentos específicos na área, realizar o teste.

1.3 Base de Dados

Locutor	Sexo	Idade	Descrição da Doença
Locutor 1	F	42	<i>nódulos bilaterais e pequena fenda em ampulheta</i>
Locutor 2	F	38	<i>nódulos bilaterais com fenda médio-posterior</i>
Locutor 3	F	24	<i>nódulos vocais moderados a severos com significativa fenda anterior e posterior (em ampulheta)</i>
Locutor 4	F	53	<i>nódulos vocais com fenda em ampulheta</i>
Locutor 5	F	53	<i>nódulos vocais com fenda em ampulheta</i>
Locutor 6	F	38	<i>nódulos vocais com fenda médio-posterior</i>
Locutor 7	F	34	<i>nódulos vocais com fenda médio-posterior</i>
Locutor 8	F	32	<i>nódulos fibrosos - fenda médio-posterior - grande esforço vocal</i>
Locutor 9	F	29	<i>nódulos vocais com fenda médio-posterior</i>
Locutor 10	F	33	<i>nódulos vocais com fenda em ampulheta</i>
Locutor 11	F	28	<i>nódulos vocais com discreta fenda em ampulheta</i>
Locutor 12	F	28	<i>nódulos e vocais com fenda médio-posterior</i>

Tabela 1 – Locutores com nódulo nas cordas vocais (F – Feminino, M – Masculino).

Locutor	Sexo	Idade	Descrição da Doença
Locutor 13	M	50	<i>paralisia da prega vocal direita com retração cicatricial no 1/3 médio - fenda fusiforme anterior sequela de trauma laríngeo</i>
Locutor 14	M	50	<i>paralisia idiopática da hemilaringe direita com leve arqueamento da borda livre</i>
Locutor 15	M	24	<i>paralisia da prega vocal direita com fenda fusiforme</i>

Locutor 16	<i>F</i>	<i>69</i>	<i>paralisia da prega vocal direita em posição para mediana com leve arqueamento e discreta fenda fusiforme - falsete paralítico</i>
Locutor 17	<i>F</i>	<i>45</i>	<i>paralisia da prega vocal e esquerda em posição mediana e para-mediana com coagitação e onda da mucosa adequada - sem fendas</i>
Locutor 18	<i>F</i>	<i>43</i>	<i>paralisia e idiopática da hemilaringe à direita em posição para-mediana</i>
Locutor 19	<i>M</i>	<i>66</i>	<i>paralisia pearrencial à esquerda com moderado arqueamento da borda livre (trauma de entubação)</i>
Locutor 20	<i>M</i>	<i>53</i>	<i>paralisia da prega vocal direita em posição para-mediana - rigidez da prega vocal esquerda para decorticação</i>

Tabela 2 – Locutores com paralisia nas cordas vocais (F – Feminino, M – Masculino).

Locutor	Sexo	Idade
Locutor 21	<i>F</i>	<i>56</i>
Locutor 22	<i>M</i>	<i>30</i>
Locutor 23	<i>F</i>	<i>41</i>
Locutor 24	<i>M</i>	<i>46</i>
Locutor 25	<i>F</i>	<i>61</i>
Locutor 26	<i>M</i>	<i>35</i>
Locutor 27	<i>M</i>	<i>63</i>
Locutor 28	<i>M</i>	<i>48</i>
Locutor 29	<i>M</i>	<i>26</i>
Locutor 30	<i>F</i>	<i>56</i>
Locutor 31	<i>F</i>	<i>56</i>

Tabela 3 – Locutores sem patologias nas cordas vocais (F – Feminino, M – Masculino)

1.4 Ambiente computacional

Para esse trabalho foram utilizados os seguintes softwares e hardwares:

- Aparat 0.3.1 (extração de parâmetros do sinal glotal)
- Praat (extração dos parâmetros jitter e shimmer)
- Matlab 2010a (rede neural)
- WEKA 3.6 (rede neural)
- Audacity 1.3 (amostras de vozes)
- Microsoft Visual Studio 2010 (rede neural em C)
- MikroC *compiler* for PIC 7.0.0.3 (rede neural + sistema do microcontrolador = geração do código hexadecimal)
- Proteus 7 (simulação do PIC 18F452)
- PICKit 2 v.2.61 (carrega .hex no microcontrolador)
- PIC 18F452 (kit contendo microcontrolador, teclado, LCD *display*, etc.)

1.5 Organização

Este trabalho está organizado como se segue.

No capítulo 2, há uma explicação sobre a teoria e o funcionamento de uma Rede Neural Artificial. Este mesmo capítulo também apresenta uma ideia geral do método de reconhecimento de padrões, forma pela qual o trabalho feito. O capítulo 3 apresenta uma breve pesquisa sobre a voz, o sinal de voz e as doenças da voz estudadas no presente trabalho e seus impactos sobre a voz. No capítulo 4 há uma explicação sobre o sinal glotal, e sobre os parâmetros do sinal glotal que foram usados nesse trabalho como entrada da rede neural para classificar as doenças da voz. No capítulo 5 há uma explicação sobre microcontroladores, suas aplicações e detalhes sobre o microcontrolador usado no presente trabalho. O capítulo 6 descreve o plano de ação que foi realizado para a conclusão do projeto final. O capítulo 0 apresenta análises e resultados obtidos com os dados e experimentos realizados. O capítulo 0 mostra um cronograma com as atividades realizadas durante os dois semestres de desenvolvimento deste projeto. E, finalmente, o capítulo 9 apresenta a conclusão do trabalho e trabalhos futuros.

2 Redes Neurais Artificiais

Redes neurais artificiais é uma área estudada pela inteligência computacional que vem sendo utilizada nas mais diversas áreas do conhecimento. Devido à sua capacidade de generalização e seus excelentes resultados, elas são amplamente utilizadas na resolução de problemas do mundo real envolvendo grandes massas de dados, sendo uma ferramenta poderosa na área da computação. [BRAGA, 2007]

A natureza desses problemas é, em geral, de:

- i) Classificação, isto é, atribuição a um padrão desconhecido uma entre várias classes de padrões conhecidos;
- ii) Previsão, isto é, prever situações futuras através de dados atuais;
- iii) Categorização, isto é, agrupamento em categorias bem definidas obtidas a partir da redundância existente entre os registros apresentados à rede;
- iv) Aproximação, em geral, de funções.

[BRAGA, 2007]

O uso das redes neurais para tais atividades está diretamente relacionado à sua capacidade de aprender por meio de exemplos e de generalizar a informação aprendida. Outra característica importante das RNAs é a capacidade de auto-organização, em que a rede atribui um padrão de entrada a uma classe entre um conjunto de classes conhecidas.

As redes neurais artificiais são sistemas capazes de resolver problema em situações onde é difícil criar modelos explícitos. Elas possuem a habilidade de inferir relações não lineares complexas e também a capacidade de resolver os problemas sem a necessidade de definição de listas de regras.

Essas técnicas computacionais utilizam como modelo o funcionamento do cérebro humano, onde os neurônios são treinados de forma a responder somente aos padrões de entrada, podendo ser utilizadas no reconhecimento de padrões sem a necessidade de um treinamento prévio.

As principais características que tornam desejáveis o estudo e o uso das redes neurais artificiais são: a capacidade das mesmas de aprender através da experiência, a capacidade de se adaptar a situações adversas e a sua tolerância a ruídos.

Uma rede neural artificial consiste em um modelo computacional em camadas de neurônios processadores interligados por conexões ponderadas (pesos). Cada neurônio aplica uma função, chamada função de ativação ao somatório de suas entradas mais um polarizador.

Nesse contexto, a proposta é mapear as características não lineares das medições acústicas através de redes neurais artificiais, buscando a classificação entre três tipos de

padrões de voz (normal, nódulo, paralisia).

A estrutura da rede poderá ser as seguintes três: rede *feed forward* com camada única, onde temos uma camada de entrada que se projeta sobre uma camada de saída de neurônios; rede *feed forward* com múltiplas camadas, que se distingue da anterior pela presença de uma ou mais camadas ocultas. A função dos neurônios ocultos é intervir entre a camada de entrada e saída de maneira útil, assim, a rede torna-se capaz de extrair estatísticas de ordem mais elevada; rede *backpropagation*, que se distingue da *feed forward* por ter um laço de realimentação, onde cada neurônio alimenta seu sinal de saída de volta para a entrada de todos os outros neurônios.

O aprendizado é um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de estímulos do ambiente onde está a rede. Os tipos de aprendizado são: aprendizado supervisionado, onde a rede é treinada através do fornecimento dos valores de entrada e de seus respectivos valores de saída desejados; e aprendizado não supervisionado, onde não existe saída desejada. A rede é treinada através das entradas e então, organiza os padrões em categorias.

O aprendizado pode ser dividido em dois: *batch*, que somente ajusta os pesos após a apresentação de todos os padrões; e incremental, que atualiza os pesos a cada apresentação de um novo padrão.

Também deverão ser observados os parâmetros utilizados no treinamento, como o momento, que diminui a probabilidade do processo de convergência parar em um mínimo local; a taxa de aprendizagem, que influencia a magnitude dos pesos.

2.1 Classificação por redes neurais

As redes neurais artificiais podem ser empregadas a fim de classificar determinados dados de entrada segundo padrões, buscando assim associar uma resposta a cada um deles. Essa classificação de padrões pode ser feita de diferentes formas, de acordo com o tipo de rede neural adotado, bem como a regra de aprendizado que o mesmo emprega.

Classificação de padrões por redes neurais é muito utilizada para, por exemplo, detecção de fraudes, rating automático para crédito e validação automática de informações financeiras.

A precisão da rede na sua tarefa de classificação dependerá da variabilidade dos valores das amostras usadas como exemplo para seu treinamento, como também o número de amostras. O número de amostras deve ser não apenas suficiente, mas deve também ser composto por amostras cujos valores estejam suficientemente distribuídos entre os valores possíveis do sistema para permitir que a rede o modele com a precisão desejada.

Com os parâmetros do sinal glotal, as patologias são classificadas com uma rede neural *multilayer perceptron* (CATALDO, 2005). O modelo de classificação por rede neural é

desenvolvido para classificar os locutores em três grupos: locutores com nódulo nas cordas vocais, locutores com paralisia das cordas vocais e locutores com vozes normais.

Esses grupos serão as saídas da rede neural, que utilizou 70% da base de dados para treinamento, 20% para validação (para evitar um sobre treinamento da rede e para escolher o número de processadores e de camadas escondidas) e 10% para teste. Para o modelo dessa rede neural foi escolhido o método *early stopping* para evitar o sobre treinamento, em função disso foi separado um conjunto de validação.

3 A voz

3.1 O sinal de voz

A voz é uma das extensões mais fortes da personalidade humana, e nunca a voz teve tanto poder e tantas possibilidades de uso com as que hoje se dispõe, amplificada pela tecnologia. Através dos registros em discos e fitas de gravação, do rádio, do telefone e de todo o sistema eletroeletrônico de transmissão de ondas sonoras, quase nos acostumamos com a voz sem a presença do falante, ou mesmo do ouvinte.

Qualquer emissão humana falada, cantada ou até mesmo uma simples exclamação, apresenta três funções: [FOLMER, 1968]

- Função de representação: a voz comunica alguma coisa, ou seja, seu uso está relacionado ao conteúdo da mensagem verbal.
- Função de expressão: a voz revela alguma coisa do falante, como sua idade, seu nível sócio-econômico-cultural, seu estado emocional, etc.
- Função de Apelo: a voz deseja e provoca uma reação no ouvinte, o que significa que existe sempre uma intenção, frequentemente inconsciente, no tipo de voz que se utiliza no discurso.

Porém, em se tratando de som, apenas isto, outros tipos de análise são feitos sobre a pressão sonora realizada pelo aparelho fonador do homem. Por exemplo, a intensidade vocal é resultado da amplitude das vibrações das cordas vocais, também depende da frequência, porque os sons graves têm amplitude maior que os agudos. [BLOCH, 1958]

Ainda sobre a voz humana, qualidade vocal, termo antes chamado de timbre (atualmente mais usado para instrumentos musicais), é o termo atualmente empregado para designar o conjunto de características que a identificam. A qualidade vocal se relaciona à composição dos harmônicos da onda sonora. Isso acontece porque, o que de fato ouvimos, é o som resultante da superposição de vários sons de frequências diferentes. No entanto, a frequência do som ouvido é igual à do som de menor frequência, denominado de frequência fundamental. O que a diferencia é a presença dos sons harmônicos.

Segundo, as frequências fundamentais das vozes masculinas podem variar de 80 a 150 Hz, as femininas de 150 a 250 Hz, e as infantis encontram-se acima de 250 Hz. Apesar de os valores identificados das frequências fundamentais ficarem na faixa de 80 a 300 Hz, considerando homens, mulheres e crianças, a linguagem articulada compreende sons com frequências que variam de 250 Hz a 1500 Hz. [BEHLAU, 1995]

3.2 Doenças da voz

A voz é uma característica humana intimamente relacionada com a necessidade do homem de se agrupar e se comunicar. Ela é produto da nossa evolução, um trabalho em conjunto do sistema nervoso, respiratório e digestivo, e de músculos, ligamentos e ossos, harmoniosamente atuando para que se possa obter uma emissão eficiente. [KENT, 1992]

3.2.1 Nódulos

Nódulos são lesões de massa, benignas, bilaterais, de característica esbranquiçada ou levemente avermelhada, que se desenvolvem na região anterior das pregas vocais, na metade da área de maior vibração glótica, decorrentes essencialmente do abuso vocal.

O local sugere que os nódulos são resultados de fonotrauma, o estresse físico das cordas vocais que ocorre pelo uso pesado da voz (em grande quantidade, por longo tempo, com frequência fundamental muito grave, com Velocidade de fala aumentada, ataques vocais bruscos, etc.) ou pelo uso da voz sob condições acústicas adversas.

É um fato que os nódulos são encontrados quase que exclusivamente em mulheres jovens e meninos pré-adolescentes, o que indica que o tamanho da laringe também pode ser um fator que predispõem a formação de nódulos. Os nódulos são um problema bem conhecido dos cantores profissionais e amadores. Alguns têm a hipótese de que é devido ao estilo particular de cantar e outros têm proposto que está simplesmente relacionado ao aumento da quantidade de uso da voz ou a intensidade do cronograma de execução.



Figura 1: Os nódulos são simétricos meados de lesões nas pregas vocais

Os nódulos não são lesões pré-cancerosas ou cancerosas. Em geral, a natureza simétrica bilateral de nódulos, e o fato de que ocorrem em pessoas jovens é suficiente para eliminar qualquer dúvida sobre a malignidade e fazer uma biópsia desnecessária.

Nódulos geralmente causam rouquidão indolor, que é resultada pelas irregularidades no fechamento das pregas vocais, assim como irregularidades na vibração, ambas os resultados do volume do nódulo. A rouquidão geralmente piora como uso da voz, ou durante um resfriado ou

garganta inflamada, e melhora com o repouso vocal.



Figura 2: Nódulos prejudicam a voz por bloquear o total fechamento das pregas vocais e por perturbar as vibrações das cordas vocais durante a vocalização.

Os nódulos são simétricos ou quase simétricos. Eles podem variar de tamanho de paciente para paciente. Eles são amplos e se assemelham a um monte de tecido. Nódulos são da mesma cor que o resto da prega vocal, em contraste com a maioria dos pólipos, que são vermelhos. Os nódulos não crescem dramaticamente no curto prazo, embora possam crescer com trauma repetido do uso da voz.



Figura 3: Nódulos sempre ocorrem em par.

3.2.2 Paralisia das pregas vocais

Paralisia de pregas vocais é a imobilidade de uma das pregas vocais por causa de dano ou disfunção do seu nervo principal. Este nervo viaja a partir do cérebro passa pelo pescoço e chega ao peito antes de girar para cima de volta para a laringe. Porque ele passa pelo pescoço duas vezes, ele é chamado de nervo laríngeo recorrente. O nervo do lado esquerdo é maior que o da direita e por isso é mais propenso a lesões. Como resultado, a prega vocal esquerda é duas vezes mais propensa a ter paralisia do que a direita.

Paralisia de pregas vocais pode ser unilateral ou bilateral. Elas têm causas ligeiramente diferentes e os sintomas e tratamentos são muito diferentes.

3.2.3 Paralisia Unilateral

A maioria paralisia unilateral das pregas vocais acontece por um dos três motivos: lesão do nervo durante inúmeras cirurgias comuns, pressão sobre o nervo por um tumor ao lado dela, ou inflamação do nervo que para de trabalhar, geralmente atribuídas à infecção viral. Juntos, estes três cenários representam mais de 85% dos casos de paralisia das pregas vocais. Há dezenas de outras causas menos comuns, como acidente vascular cerebral, e outras doenças neurológicas, além de efeitos colaterais de certas drogas e toxinas.

Paralisia de pregas vocais pode ser um resultado acidental de diversas cirurgias comuns, como cirurgias de coração e pulmão, já que o nervo principal mergulha no peito antes de retornar para a laringe. A paralisia da prega vocal não é necessariamente um sinal de que o nervo foi cortado. O nervo também pode parar de funcionar se estirado ou comprimido e às vezes aos muito pouco manuseio. Por esta razão, uma das pregas vocais pode ser paralisada após mesmo a mais suave das operações. Finalmente, ramos do nervo também podem ser danificados pelo tubo de respiração colocado em anestesia geral.

Na paralisia unilateral, as pregas vocais são incapazes de fechar, o que provoca problemas de voz e deglutição. A voz é rouca, soprosa e macia, e falar em locais com muito ruído é um desafio. Os pacientes ficam sem fôlego quando falam porque é necessário muito ar para fazer as pregas vocais vibrarem. Isto é comumente confundido com a falta de ar causada por um problema pulmonar. Às vezes, músculos normalmente não envolvidos na produção da voz agem para tentar juntar as pregas vocais, o que pode causar um pescoço dolorido depois de falar por longos períodos. Ocasionalmente, as alterações na voz serão acompanhadas de tosse ao engolir. Isto é notado especialmente quando se bebe líquido.



Figura 4: A prega vocal do lado direito desta foto foi paralisada após uma operação de tireoide.



Figura 5: Um esforço para produzir a voz move a outra prega à linha média, mas uma abertura substancial permanece entre as duas. Isto produz uma voz suave e sussurrada.

3.2.4 Paralisia Bilateral

A paralisia bilateral das pregas vocais normalmente acontece pelas mesmas causas descritas na paralisia unilateral.

Na paralisia bilateral, as pregas vocais não conseguem abrir, o que causa o estreitamento e bloqueio das vias respiratórias. O tamanho do espaço deixado entre as pregas vocais imóveis determina o grau de obstrução. Há quase sempre respiração ruidosa e falta de ar. Às vezes, isso é confundido com asma, pelos médicos e pacientes, o que é um engano perigoso, pois a paralisia bilateral de pregas vocais tem uma chance muito real de causar uma obstrução fatal das vias aéreas. Dois cenários em que isso pode acontecer são: inesperadamente após uma cirurgia, ou com inchaço das pregas vocais, com em um resfriado comum, bloqueando a pequena passagem de ar restante de uma pessoa com paralisia bilateral.

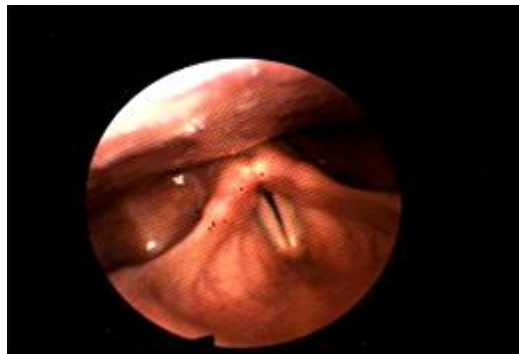


Figura 6: Esta pessoa tem paralisia bilateral de pregas vocais por causa de um acidente vascular cerebral. Esta é a sua maior abertura das pregas vocais que ele tem para respirar.



Figura 7: Essa pessoa ainda tem um pequeno movimento da prega vocal, à direita da foto, mas este é a maior abertura das pregas vocais para respirar. Ela fez uma traqueostomia.

4 Sinal Glotal

A produção da voz se inicia com uma contração-expansão dos pulmões. Cria-se assim uma diferença entre a pressão de ar nos pulmões e a pressão do ar na frente da boca, causando um escoamento de ar. O escoamento passa pela laringe e, antes homogêneo, vai se transformando em uma série de pulsos (conhecidos como sinal glotal) de ar que chegam à boca e na cavidade nasal. Os pulsos de ar são modulados pela língua, pelos dentes e lábios, isto é, pela geometria desses órgãos, de forma a produzir o que conhecemos por voz. O sinal glotal, porém, possui propriedades importantes de difícil reprodução que estão intimamente ligadas às características anatômicas e fisiológicas da laringe. [CATALDO, 2004]

Atualmente, a teoria mais aceita para a descrição do sinal glotal é a teoria chamada de aerodinâmica mioelástica, que postulou que os movimentos de abrir e fechar as cordas vocais são regidos pelas propriedades mecânicas dos tecidos musculares que constituem, principalmente, as cordas vocais e pelas forças aerodinâmicas que se distribuem ao longo da laringe durante a fala. [VAN DEN BERG, 1858]

Uma das principais formas de se caracterizar o sinal de voz é através de medições acústicas, pelo fato de que tais medidas podem revelar importantes características fisiológicas através de variações de seus valores. Quando são obtidos conjuntos de medições acústicas, é possível criar padrões de voz. A partir daí, os parâmetros obtidos através do sinal glotal podem ser usados pela rede neural artificial para classificar patologias.

4.1 Parâmetros do sinal glotal

Para discriminar o sinal glotal, extraem-se parâmetros desse sinal, aproveitando sua periodicidade, em vogais. Os parâmetros que descrevem o fluxo glotal podem ser usados em múltiplas aplicações, tais como: pesquisas sobre a produção da voz, codificação, síntese, reconhecimento automático de voz, uso clínico, verificação e identificação automática de locutor e para quantificar a contribuição do pulso glotal na transmissão de sentimentos.

Para obtermos o sinal glotal, era necessário fazer uso de técnicas invasivas, onde se introduziam micro câmeras para gravar sons logo após a passagem do ar pelas pregas vocais. Hoje em dia é possível obter o sinal glotal usando métodos não invasivos, pela utilização de um filtro inverso no sinal de voz, que consiste em eliminar a influência do trato vocal e da radiação da voz causada pela boca, preservando as características do sinal glotal. (ALKU, 1992)

Para cada registro de voz, o correspondente sinal glotal foi obtido por filtro inverso PDIAIF e os parâmetros foram extraídos usando os softwares Aparat (Software Aparat) e Praat (Software Praat). Os seguintes parâmetros foram obtidos: *frequência fundamental (f_0)*, *jitter*,

shimmer, K_o , K_a , NAQ , AQ , CIQ , $OQ1$, $OQ2$, OQa , QoQ , $SQ1$, $SQ2$, $DH12$, e HRF . Os parâmetros foram separados de acordo com o grupo ao qual pertencem. OQ foi dividido em $OQ1$ e $OQ2$, os coeficientes de abertura são calculados pelas primeira e segunda aberturas do fluxo glotal. A diferença entre $OQ1$ e $OQ2$ é que $OQ1$ é calculado pelo fechamento do fluxo glotal até o fechamento do próximo fluxo glotal, e $OQ2$ é calculado pela abertura até o fechamento do fluxo glotal. O coeficiente SQ , também foi dividido em dois pelas primeira e segunda aberturas do fluxo glotal.

Algoritmos que obtêm o sinal glotal podem ser classificados em duas categorias: semiautomáticos e manuais. Nesse trabalho o algoritmo de filtragem inversa utilizado foi o semiautomático, chamado PSIAIF (*Pitch Synchronous Iterative Adaptive Inverse Filtering*) (PULAKKA, 2005) (JULIANO, 2008). O algoritmo PSIAIF foi escolhido pelo seu alto desempenho e seu fácil desenvolvimento. Existe um toolbox em Matlab, chamado Aparat (Software Aparat), que foi construído baseado especialmente no método PSIAIF para obter o sinal glotal e extrair suas principais características e parâmetros. Os parâmetros podem ser divididos em três grupos: parâmetros do domínio do tempo, do domínio da frequência, e os que representam variações da frequência fundamental. (PULAKKA, 2005)

4.1.1 Parâmetros do domínio do tempo

Os parâmetros que podem ser extraídos do sinal glotal são descritos abaixo (PULAKKA, 2005), (JULIANO, 2008).

- i. Frequência fundamental (f_0): é a frequência de abertura e fechamento da glote. É uma das principais medidas utilizadas pelos estudiosos para caracterizar a voz humana em diversas etapas da vida. A frequência fundamental tem uma relação direta com massa, elasticidade e comprimento das pregas vocais. Por meio da análise da frequência fundamental, podemos determinar a idade e sexo do indivíduo.
- ii. Fase de fechamento (K_0): É a duração da fase de fechamento glotal. Tal duração é indicada na Figura 8 **Erro! Fonte de referência não encontrada..** [PULAKKA, 2005]
- iii. Fase de abertura (K_a): é a duração da fase de abertura glotal. Tal duração é indicada na Figura 8. [PULAKKA, 2005]
- iv. Quociente de abertura (OQ): é a razão entre o tempo total de abertura da prega vocal e o tempo total de um ciclo do sinal glotal. É inversamente proporcional à intensidade da voz. O quanto menor é seu valor, maior é a intensidade da voz. (ROSA, 2005) (PULAKKA, 2005)
- v. Quociente de fechamento (CIQ): é a razão entre a fase de fechamento (K_0) e o tempo total de um ciclo do sinal glotal. (PULAKKA, 2005) É inversamente proporcional à intensidade da voz. O quanto menor é seu valor, maior é a intensidade da voz.

(ROSA, 2005) (PULAKKA, 2005)

- vi. Quociente de amplitude (AQ): é a razão entre a amplitude do sinal glotal (máxima abertura) e o valor mínimo do sinal glotal. (ALKU, 1996) Está relacionada com a fonação do locutor. (JULIANO, 2008)
- vii. Quociente de amplitude normalizada (NAQ): é calculado pela razão entre o quociente de amplitude (AQ) e o tempo total da duração de um pulso glotal. (ALKU, 2002)
- viii. Quociente de abertura definido pelo modelo Liljencrants-Fant (OQa): é outro quociente de abertura, que é calculado pelo modelo de Liljencrants-Fant para filtro inverso. Detalhes sobre esse modelo Poe ser encontrado em (GOBL, 2003).
- ix. Quociente de abertura (QoQ): é a relação entre o tempo da abertura do sinal glotal no exato momento da oscilação e o tempo de fechamento. (JULIANO, 2008) Ele vem sido usado em alguns trabalhos para classificar emoções. (LAUKKANEN, 1996)
- x. Quociente de velocidade (SQ): definido como a razão entre a duração da fase de abertura e a fase de fechamento do sinal glotal. (PULAKKA, 2005)

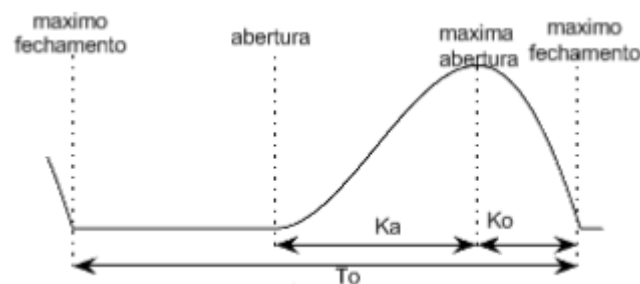


Figura 8 - Sinal glotal (K_0 e K_a)

4.1.2 Parâmetros do domínio da frequência

- i. Diferença entre harmônicos (DH12): também conhecido como H_1-H_2 e é a diferença entre os valores do primeiro e segundo harmônicos do sinal glotal (TITZE, 1992) (AIRAS, 2008). Esse parâmetro tem sido usado para medir a qualidade vocal.
- ii. Fator de relação harmônica (HRF): relaciona o primeiro harmônico (H_1) com a soma de energia dos outros harmônicos (H_k) [17]. Também vem sido usado para medir a qualidade vocal.

4.1.3 Parâmetros que representam variações e perturbações na frequência fundamental

- i. É a perturbação ou variabilidade da frequência fundamental ciclo a ciclo, ou seja, a perturbação da frequência do sinal vocal, como mostrado na Figura 9. [CHILDERS, 1990], [CATALDO, 2005]. Este parâmetro está relacionado ao aspecto perceptual de aspereza na voz. [MONTAGNOLI, 1999]. Mudanças no jitter podem ser causadas por dificuldades psicológicas ou neurológicas. (ROSA, 2005)

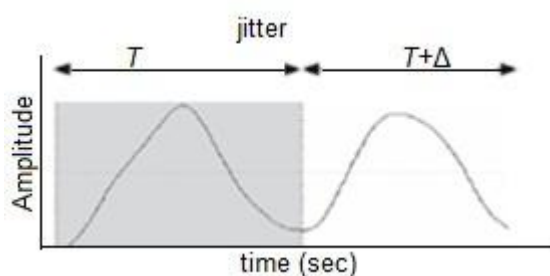


Figura 9: Jitter

- ii. É a variação da amplitude do sinal de voz, obtida através da diferença entre a amplitude do sinal de voz obtida em um período e a do período seguinte, como mostrado na Figura 10. [CHILDERS, 1990] [CATALDO, 2005] Esse parâmetro está intimamente relacionado ao aspecto perceptual de rouquidão. [MONTAGNOLI, 1999]. Mudanças no shimmer acontecem, principalmente quando há lesões nas cordas vocais, como pólipos, edemas e carcinomas. (ROSA, 2005)

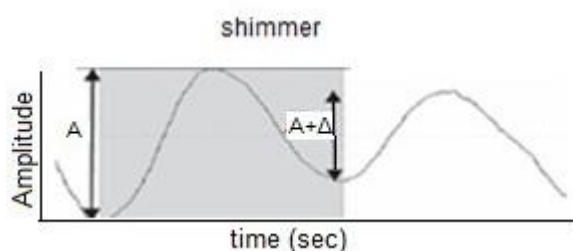


Figura 10: Shimmer

Os parâmetros Jitter e Shimmer são especialmente importantes em uso clínico, pois mudanças em seus valores indicam, geralmente, doenças vocais, como lesões, edemas, carcinomas, e até mesmo doenças neurológicas ou psicológicas. A análise de frequência fundamental e medidas de perturbação, como jitter e shimmer, permitem descrever quase completamente a voz humana. Isso colabora significativamente no processo de avaliação e

tratamento dos distúrbios vocais. [ARAÚJO, 2002]

Os parâmetros K_0 e K_a são especialmente importantes para o estudo do envelhecimento da voz, pois com o passar dos anos, há uma degeneração de gordura dos tecidos acarretando

uma diminuição da elasticidade dos ligamentos, assim, as mudanças anatômicas na laringe decorrentes da idade podem limitar a extensão do fechamento das pregas vocais. [LINVILLE, 1996]

5 Microcontroladores

Os microcontroladores são chips inteligentes, que têm um processador, pinos de entrada/saída e memória. Através da programação dos microcontroladores podemos controlar suas saídas, tendo como referência as entradas ou um programa inteiro. [Microchip Technology Incorporated, 2010]

Um microcontrolador difere de um microprocessador em vários aspectos. Primeiro e o mais importante, é a sua funcionalidade. Para que um microprocessador possa ser usado, outros componentes devem-lhe ser adicionados, tais como memória e componentes para receber e enviar dados. Em resumo, isso significa que o microprocessador é o verdadeiro coração do computador. Por outro lado, o microcontrolador foi projetado para ter tudo em um só. Nenhum outros componentes externos são necessários nas aplicações, uma vez que todos os periféricos necessários já estão contidos nele. Assim, nós poupamos tempo e espaço na construção dos dispositivos. [Microchip Technology Incorporated, 2010].

5.1 Aplicações

Microcontroladores são geralmente utilizados em automação e controle de produtos e periféricos, como sistemas de controle de motores automotivos, controles remotos, máquinas de escritório e residenciais, brinquedos, sistemas de supervisão, etc. Por reduzir o tamanho, custo e consumo de energia, e se comparados à forma de utilização de microprocessadores convencionais, aliados a facilidade de desenho de aplicações, juntamente com o seu baixo custo, os microcontroladores são uma alternativa eficiente para controlar muitos processos e aplicações.

5.2 PIC 18F452

O PIC 18F452 é um modelo de microcontrolador PIC da família de 16 bits de núcleo de processamento, apesar de ainda lidar com dados de 8 bits. A vantagem dessa família de PICs é que eles possuem mais instruções em código de máquina (75 contra 35 da série 16F) que é otimizada para ser usada com compiladores de linguagem C e memória linear. A sua velocidade de processamento também é maior, na maioria até 10 MIPS (40MHz de *clock*) com alimentação entre 2 e 5,5V. O seu encapsulamento DIP consiste em 40 pinos. [Microchip Technology Incorporated, 2010]

Esse modelo possui 32k bytes (16k instruções) de memória de programa flash, 1536 bytes de memória RAM e 256 bytes de memória EEPROM. Os principais periféricos internos são:

- 8 canais para conversão A/D a 10 bits
- AUSART (*addressable universal synchronous assynchronous receiver transmitter*) com suporte para RS485
- MSSP (*master synchronous serial port*) para SPI e I2C
- PSP (*parallel slave port*)
- 4 *timers* (3 de 16 bits e 1 de 8 bits)
- 2 módulos CCP (*capture compare PWM*)

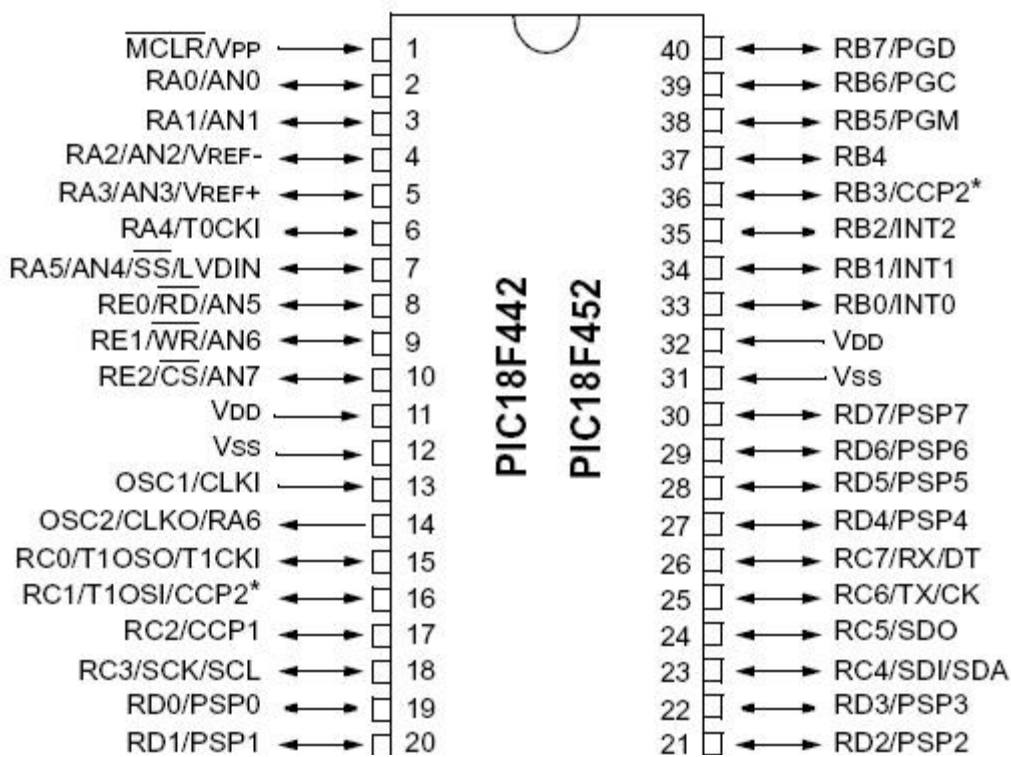


Figura 11 – PIC 18f452

6 Plano de Ação

6.1 Estudos iniciais

Para atacar o problema de modo a chegar a um resultado, procurando que ao final do trabalho, a estrutura seguisse a proposta feita, foi feito inicialmente um estudo aprofundado de:

- ✓ redes neurais artificiais: topologias de rede, aprendizado, treinamento, etc.
- ✓ implementação de redes neurais no software WEKA
- ✓ implementação de redes neurais em MATLAB

Como o projeto diz respeito à voz, patologias que a afetam, também foi feito um estudo dos fundamentos da produção da voz humana e de como se faz o processamento da voz, assim como a extração de características do sinal de voz, sinal glotal e seus parâmetros, e filtragem inversa.

Sendo o objetivo final, produzir um equipamento que possa receber novas amostras de voz através de um microfone, também foi necessário o estudo de microcontroladores e sua linguagem específica em C.

6.2 Preparação da base de dados

6.2.1 Audacity

A partir da base de dados, especificada no primeiro capítulo, com amostras de voz de homens e mulheres em todas as faixas de idade, com nódulos nas cordas vocais, com paralisia unilateral das cordas vocais e saudáveis, que contém poucos locutores para cada categoria, no caso das patologias: 12 casos de nódulo, 8 casos de paralisia e 11 casos com vozes saudáveis, serão retirados de cada arquivo *wave* 8 amostras através do software *Audacity*. Cada amostra tem duração de 0,1 segundos, resultando em um banco com 248 amostras, o que já foi suficiente para treinar nossa rede neural.

6.2.2 Aparat e Praat

A partir da base de dados foi necessário fazer o pré-processamento da voz, onde o sinal da voz (analógico) foi convertido em um sinal digital, para posterior extração de características de sinal da glotal que possui propriedades importantes que estão intimamente ligadas às características anatômicas e fisiológicas da laringe. Os parâmetros que descrevem o fluxo glotal foram usados para a detecção de patologias.

Tais parâmetros foram extraídos com a ajuda do software *Aparat*, que nos retorna, para cada amostra de voz, um arquivo *.mat* (Matlab Data) com várias características da voz. Nesse trabalho, nos interessará, inicialmente, os seguintes parâmetros: *frequência fundamental*, K_0 , K_a , NAQ , AQ , CIQ , $OQ1$, $OQ2$, OQa , QoQ , $SQ1$, $SQ2$, $DH12$, HRF , *jitter (local)* e *shimmer (local)*.

Outros quatro parâmetros muito importantes para a classificação de doenças vocais são K0, Ka, jitter e shimmer. K0 e Ka podem ser obtidos indiretamente através dos coeficientes fornecidos pelo Aparat. O jitter e o shimmer foram obtidos através de outro software: Praat.

6.2.3 Matlab

Foi feito um script em Matlab para extração automática de todos esses parâmetros de todas as amostras de voz.

6.3 WEKA

Inicialmente foram usados todos os parâmetros para criar uma rede no software WEKA, onde a visualização da distribuição de dados é melhor e a implementação da rede é mais simples. Além disso, é possível criar um filtro facilmente, que nos retorna os parâmetros mais relevantes para a classificação, descartando os parâmetros que só atrapalham ou que não fazem diferença para a rede, apenas a tornam mais complexa.

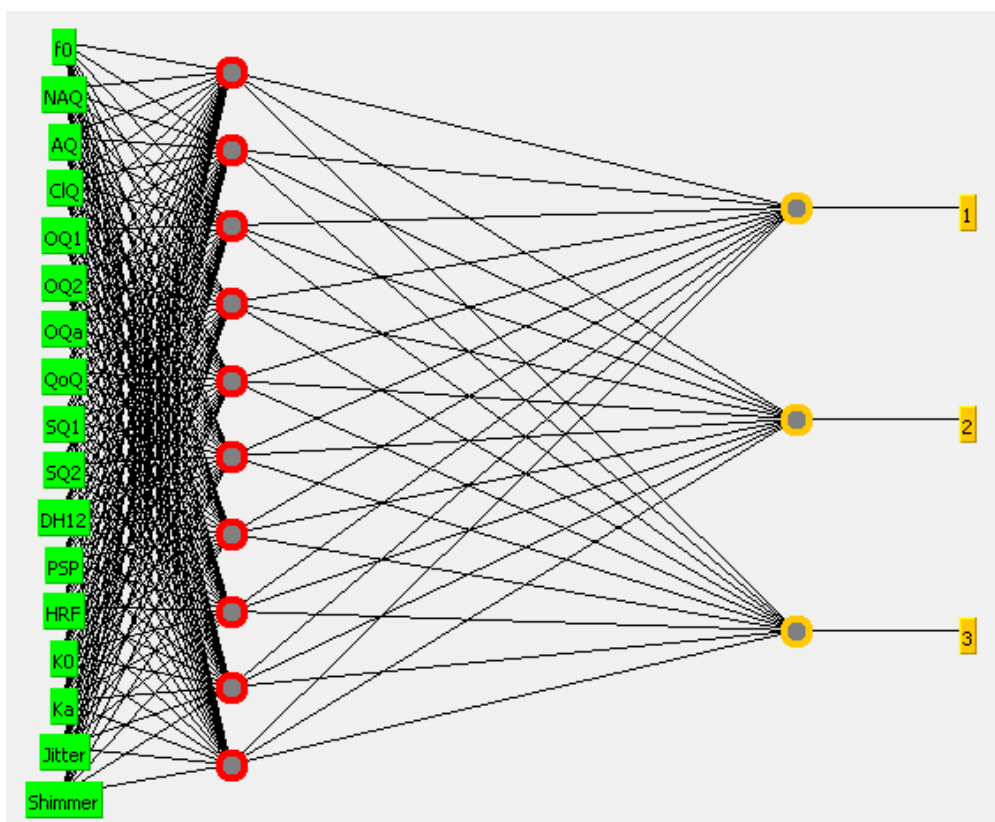


Figura 12 – Modelo da rede neural gerada pelo WEKA

6.4 Matlab

A partir daí foi usada uma rede neural em MATLAB para classificar a voz, levando em conta os seguintes aspectos: melhor arquitetura de rede, parâmetros da rede, normalização de pesos, critério de parada do treinamento, taxa de aprendizagem e momento.

6.5 Rede em linguagem C

Quando a rede obteve um resultado de classificação satisfatório, próximo de 100% de acerto dos novos padrões apresentados, a mesma foi reproduzida em linguagem de programação C para ser colocada em um PIC, usando os bias e pesos fornecidos pelo Matlab ao se criar e treinar uma rede neural.

6.6 MikroC

Com a Rede implementada e testada, o código hexadecimal é gerado pelo software MikroC.

6.7 Proteus

Foi usado um software de teste, chamado Proteus, para testar o funcionamento do sistema a ser implantado no microcontrolador. O Proteus permite fazer a simulação do protótipo, montando todo o esquemático de hardware do kit PicGenios, como mostrado na **Figura 13**.

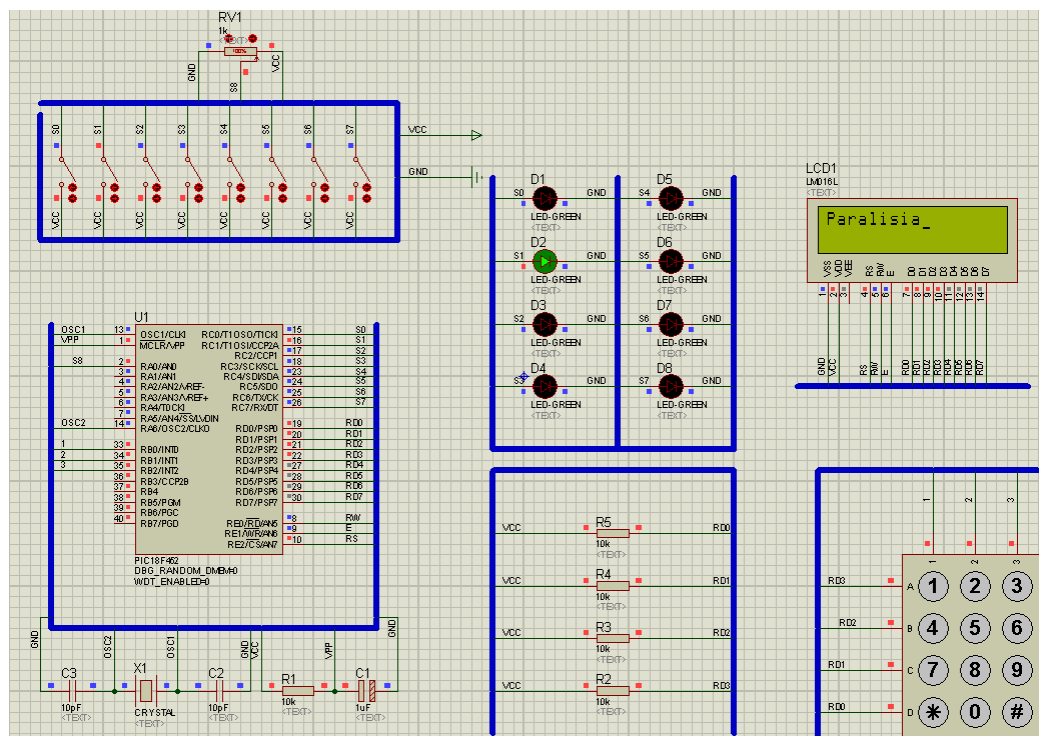


Figura 13 – Screenshot da montagem do hardware do KIT PIC18f452 no software Proteus

6.8 Kit PicGenios

O código hexadecimal gerado pelo software MikroC é então embarcado no microcontrolador e qualquer novo padrão pode ser apresentado através do teclado do kit, que enviará através a resposta da classificação através do LCD do microcontrolador.



Figura 14 – Kit PicGenios

7 Análises e Resultados

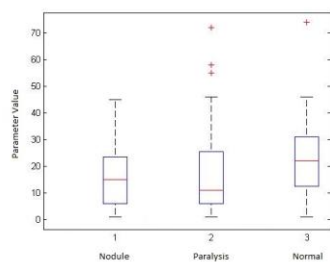
7.1 Análise dos parâmetros para classificação

Para que se possa avaliar a influência de cada parâmetro de entrada na classificação de doenças da voz, o *boxplot* do Matlab foi utilizado. O *boxplot* é mostrado para cada parâmetro extraído do sinal glotal das gravações da base de dados, para que se possa ver seus comportamentos em cada tipo de patologia, em vozes normais e comparar seus comportamentos. *Boxplot* é um modo conveniente de representar graficamente os grupos de dados numéricos, e foi utilizado neste trabalho para analisar a influência de cada parâmetro em classificar corretamente cada doença. (ALAGAMBIGAI, 2009)

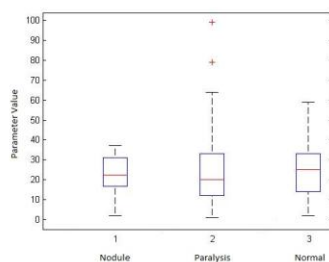
A linha central da caixa marca a mediana do conjunto de dados. A parte inferior da caixa delimita 25% dos pontos abaixo da mesma e a parte superior delimita 75% dos pontos abaixo. As hastes inferiores e superiores se estendem, respectivamente, pelo restante dos pontos marcados no gráfico. Os valores fora desses limitadores, que destoam muito do restante dos valores, são considerados *outliers*.

O *boxplot* permitiu avaliar a simetria dos dados, sua dispersão e a existência ou não de *outliers* nos mesmos, sendo de muita ajuda na comparação dos três conjuntos de dados correspondentes às vozes com nódulo, paralisia das cordas vocais e vozes normais. O *boxplot* foi construído com a função *boxplot()* do Matlab.

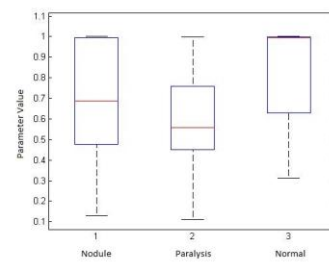
A frequência fundamental não sofre grandes alterações entre os três grupos comparados. O parâmetro K_0 , que mostra a fase de fechamento das cordas vocais é maior em vozes normais, como mostrado na Figura 15 (a). Os parâmetros AQ, NAQ, OQa e QoQ mostram que as vozes normais têm mais intensidade em comparação com patologias, como mostrado na Figura 15 (f), (g), (h) e (h). Os parâmetros SQ1 e SQ2 têm valores mais baixos em vozes normais, que indica uma redução na estrutura das cordas vocais quando se tem doenças, especialmente paralisia, como mostrado na Figura 15 (l) e (m). Em vozes com patologias, a estrutura das cordas vocais é muito comprometida, o que é indicado pelo *jitter* e *shimmer* na Figura 15 (o) e (p). Esses parâmetros variam mais nas vozes com paralisia e atingem valores muito alto também, provando ter afetado muito as cordas vocais.



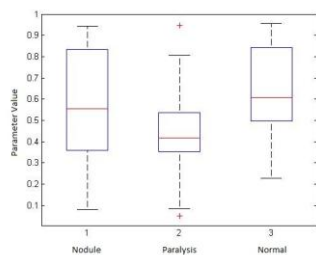
a



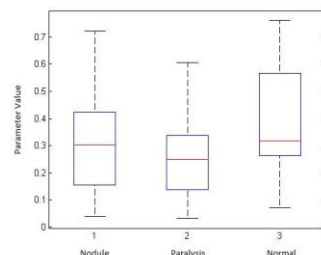
b



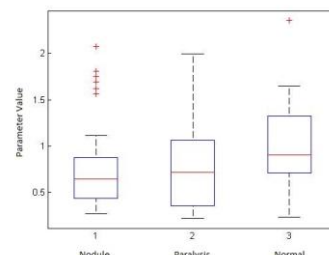
c



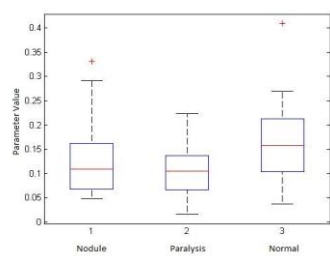
d



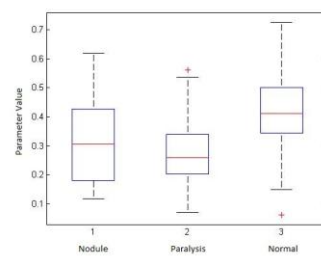
e



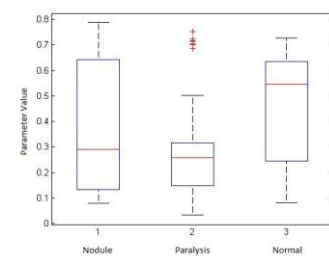
f



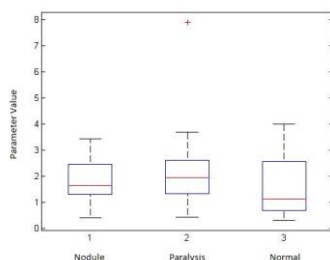
g



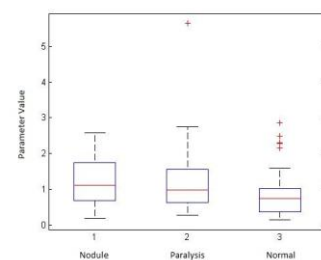
h



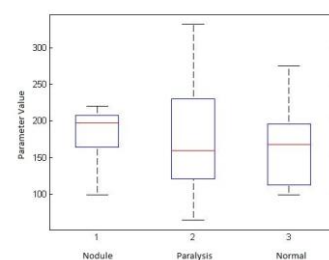
i



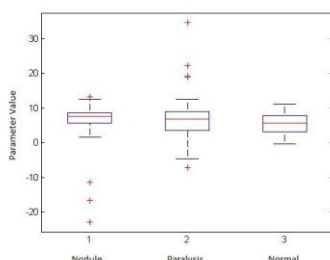
j



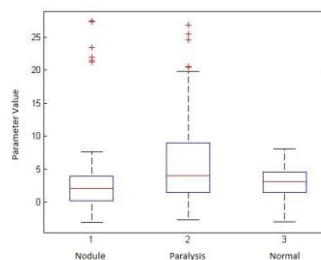
k



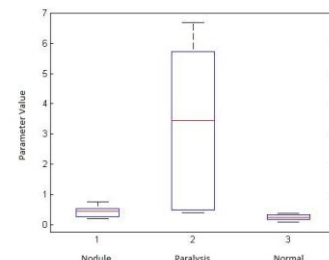
l



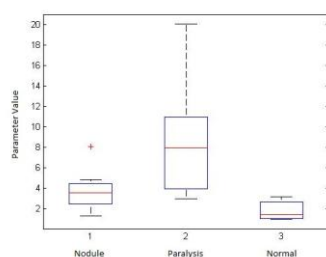
m



n



o



p

Figura 15 - *Boxplot* para cada parâmetro extraído do sinal glotal. K_0 (a), K_a (b), OQ1 (c), OQ2 (d), CIQ (e), AQ (f), NAQ (g), OQa (h), QoQ (i), SQ1 (j), SQ2 (k), F0 (l), DH12 (m), HRF (n), Jitter (o), Shimmer (p).

7.2 Análise da implementação do sistema no microcontrolador

O objetivo inicial era criar um equipamento médico que pudesse ser operado por leigos para se fazer a classificação de vozes entradas no microcontrolador por um microfone acoplado ao mesmo. Por motivo de tempo disponível para a realização do sistema, o microfone não pode ser acoplado, e a entrada dos coeficientes foi feita pelo teclado do KIT PIC18f452. Todos os testes realizados tiveram resultados positivos, classificando cada amostra de voz corretamente.

7.3 Análise da distribuição dos dados

Ao final de testes iniciais com resultados insatisfatórios, foi feita uma análise da distribuição dos dados pelo software Weka, resultando no gráfico da Figura 16.

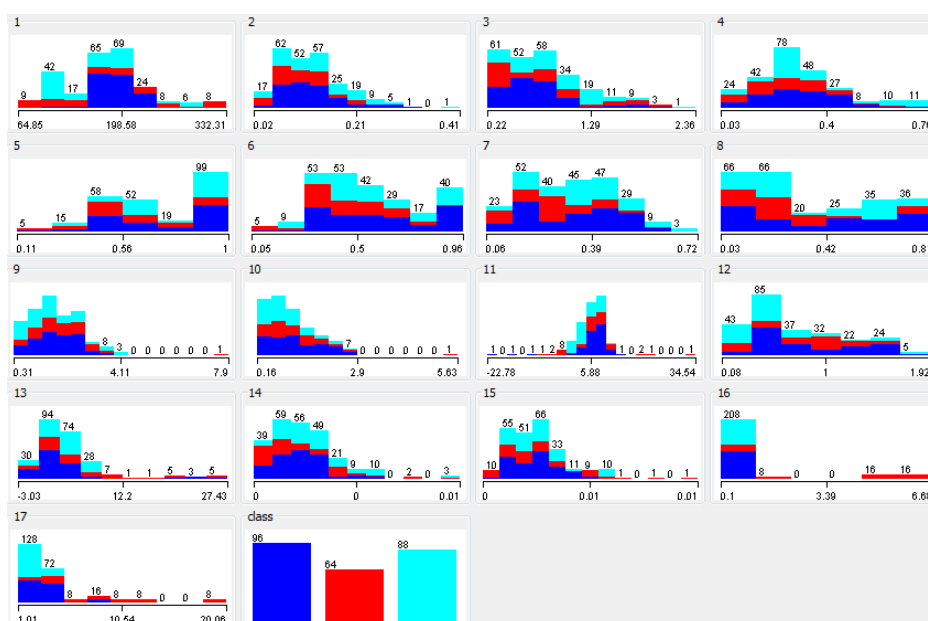


Figura 16 – Distribuição dos dados gerados pelo software WEKA

Como a distribuição se concentra mais em alguns trechos, foi feita a normalização por partes de todos os coeficientes pelo Matlab. A normalização foi feita limitando os valores entre 1 e -1.

7.4 Resultado da classificação dos parâmetros do sinal glotal

A base de dados original foi dividida em três conjuntos: de treinamento (70%), de validação (20%) e de teste (10%). Após inúmeros testes variando o número de processadores, o melhor resultado foi alcançado com 10 processadores na camada escondida.

A rede neural tem três saídas: locutores com nódulos na corda vocal, contendo 96 gravações de voz, locutores com paralisia nas cordas vocais, contendo 64 gravações, e locutores com vozes normais, contendo 88 gravações. Instâncias classificadas corretamente com 10 processadores na camada escondida, alcançou o resultado de 96%, separando 10% da base para o conjunto de teste. A matriz de confusão é mostrada na Tabela 4.

	Nodule	Paralysis	Normal	
Nodule	10	0	1	Nodule
Paralysis	0	6	0	Paralysis
Normal	0	0	8	Normal

Tabela 4 – Matriz de confusão

8 Cronogramas

8.1 Cronograma Realizado

Cronograma de estudo	Mês					
2010	08 2010	09 2010	10 2010	11 2010	12 2010	01 2011
Entendimento do problema						
Verificação bibliográfica						
Conhecimento do Aparat						
Conhecimento do MATLAB						
Conhecimento do MikroC						
Conhecimento do Proteus						
Conhecimento do Audacity						
Conhecimento do Praat						

Cronograma de projeto	Mês					
2010	08 2010	09 2010	10 2010	11 2010	12 2010	01 2011
Base de vozes para patologias vocais						
PIC 18F452						
Download e instalação de Softwares						
Entrega de resultados - orientador						
Entrega de resultados - coordenação						

Cronograma de desenvolvimento	Mês					
2010	08 2010	09 2010	10 2010	11 2010	12 2010	01 2011
Extração das amostras						
Extração dos coeficientes do Aparat						
Extração dos coeficientes do Praat						
Criação da rede no WEKA						

2011	Mês					
	01 2011	02 2011	03 2011	04 2011	05 2011	06 2011
Criação da rede em MATLAB						
Criação da rede em C						
Adaptação da rede para o PIC (PROTEUS)						
Desenvolvimento do protótipo						
Testar rede no PIC						

9 Conclusões e Trabalhos Futuros

9.1 Conclusões

Este trabalho apresentou uma proposta de um novo modelo de classificação de patologias da voz, que é capaz de distinguir uma pessoa saudável de outra com nódulo ou paralisia nas cordas vocais. Para desenvolver um modelo para a classificação feita neste trabalho, criamos um banco de dados de locutores com estas duas doenças, paralisia e nódulo nas cordas vocais e locutores com vozes normais. Os locutores têm diferentes idades e sexo. Os resultados obtidos foram satisfatórios e melhor do que os obtidos em outros trabalhos. No entanto, é importante notar que não é conclusivo que separadamente os parâmetros identificarão corretamente a patologia da voz, mas é conclusiva de que quando eles são combinados, o desempenho é melhorado. O método dos parâmetros do sinal glotal em comparação com o método tradicional MFCC tem melhor desempenho.

Este trabalho usa parâmetros do sinal glotal, e duas patologias de voz que é uma novidade.

9.2 Trabalhos Futuros

É indicado para trabalhos futuros, que seja feito o sistema completo de entrada do sinal de voz no microcontrolador.

A ideia seria acoplar um microfone ao microcontrolador e programá-lo a extrair, em tempo de execução os coeficientes glotais do sinal de voz entrado, e logo após passá-los como parâmetro para a rede neural - anteriormente treinada - implementada na PIC. Sendo assim, a rede neural classificaria a voz, mostrando o resultado da classificação em seu display.

Outra possibilidade para continuação desse trabalho seria a adição de mais padrões para classificação: tanto de outras doenças vocais, como a separação dos tipos de paralisia e tipos de nódulos vocais.

10 Referências

PULAKKA, H. **Analysis of Human Voice Production Using Inverse Filtering, High-Speed Imaging, and Electroglottography**. University of Technology Helsinki, 2005.

ALKU P., **Glottal wave analysis with Pitch Synchronous Adaptive Inverse Filtering**, Speech Communication, vol. 11, pp. 109-118, 1992.

JULIANO S. M., **Um estudo comparativo entre o sinal electroglotográfico e o sinal de voz**, Dissertação de mestrado em Engenharia de Telecomunicações, UFF, 2008.

ALKU P. and Vilkman. E., **Amplitude Domain Quotient for Characterization of the Glottal Volume Velocity Waveform Estimated by Inverse Filtering**. Speech Communication, 18(2):131-138, 1996.

ALKU P., Bäckström T., and Vilkman E. Normalized Amplitude Quotient for Parameterization of the Glottal Flow, Journal of the Acoustical Society of America, 112(2):701-710, August, 2002.

TITZE and Sundberg J. **Vocal intensity in Locutors and singers**, Journal of the Acoustical Society of America, 2936-2946, May, 1992.

AIRAS M., **Methods and studies of laryngeal voice quality analysis in speech production**, Dissertation for the degree of Doctor Helsinki University of Technology, 2008.

CHILDERS, D. G.; LEE, C. K. **Vocal quality factors: Analysis, synthesis, and perception**. Journal of the Acoustical Society of America, Maio de 1990.

LAUKKANEN, E. Vilkman, P. Alku, **Related to Stress and Emotional State: a Preliminary Study**, Journal of Phonetics, 24(3):313-335, 1996.

CATALDO, E.; RODRIGUES, Fabiana; BRANDÃO, A; LUCERO, Jorge. **Usando Redes Neurais para classificação de padrões de voz**. In: XXVIII CNMAC – Congresso Nacional de Matemática Aplicada e Computacional, 2005, São Paulo. Anais do XXVIII CNMAC, (2005).

GOBL, C., Chasaide. A., Amplitude-based source parameters for measuring voice quality, In

VOQUAL, 151-156, 2003.

BEHLAU, M. S.; PONTES, P. **Avaliação e tratamento das difonias**. São Paulo: Lovise, 1995.

ROSA I. S., **Análise acústica da voz de indivíduos na terceira idade**. Tese de mestrado Universidade de São Carlos, 2005.

ARAÚJO, S. A. **Análise acústica da voz humana**. 1999. 88f. Dissertação (Mestrado em Otorrinolaringologia) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 1999.

MONTAGNOLI, A. N. **Análise acústica do sinal de voz**. 1999. 65f. Relatório – Escola de Engenharia de São Carlos: Universidade de São Paulo, 1999.

LINVILLE, S. E. **The sound of senescence**. Journal of Voice, St. Louis, v.10, n.2, p.190-200, June 1996.

BRAGA, Antônio de P.; CARVALHO, Andre Ponce de Leon F.; LUDERMIR, Teresa Bernarda. **Redes Neurais Artificiais: Teoria e Aplicações**. 2. Ed. Rio de Janeiro: LTC, 2007. 283p.

KENT, R., READ C. **The Acoustic Analysis of Speech**. Singular Publishing Group, 1992.

CATALDO, E.; RUBENS, S.; NICOLATO, L; **Uma discussão sobre Modelos Mecânicos de Laringe para Síntese de Vogais**. ENGEVISTA, V.6, N.1, P 47-57, abril 2004.

VAN DEN BERG, J., **Myoelasticaerodynamic theory of voice production**. Journal of Speech and Hearing Research, vol.1, p 227-244, 1958.

FOLMER, JOHNSON., TORE, N. O., **Oscilações, ondas, acústica**. São Paulo: Nobel, 1968.

SOFTWARE PRAAT, <http://www.fon.hum.uva.nl/praat/>, University of Amsterdam.

SOFTWARE APARAT, <http://aparat.sourceforge.net/index.php/>, Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing.

ALAGAMBIGAI, P., Thangavel, K., **Feature Selection for Visual Clustering**. ARTCom 2009 Kerala, vol 1, pag 498, 2009.

BLOCH, Pedro, **Estudo da Voz Humana**. Rio de Janeiro: BRASILUSO, 1958.

MICROCHIP TECHNOLOGY INCORPORATED. **PICKit™ 2 – Microcontroller programmer: user's guide**. Disponível em <www.microchip.com>. Acesso em 01 nov. 2010.