



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-12997-PRE/8274

**PROCESSAMENTO DE SINAIS DE VOZ PADRÕES
COMPORTAMENTAIS POR REDES NEURAIS ARTIFICIAIS**

Gustavo Ravanhani Matuck

Relatório Final de Projeto de Iniciação Científica (PIBIC/CNPq/INPE), orientado pelo
Dr. José Demisio Simões da Silva

INPE
São José dos Campos
2005



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

RELATÓRIO FINAL DE PROJETO DE INICIAÇÃO CIENTÍFICA
(PIBIC/CNPq/INPE)

PROCESSAMENTO DE SINAIS DE VOZ NA IDENTIFICAÇÃO DE
PADRÕES COMPORTAMENTAIS POR REDES NEURAIS ARTIFICIAIS

Gustavo Ravanhani Matuck (Bolsista PIBIC/CNPq)

E-mail: gmatuck@hotmail.com

Dr. José Demisio Simões da Silva (LAC/INPE, Orientador)

E-mail: demisio@lac.inpe.br

Junho 2005

SUMÁRIO

1 - INTRODUÇÃO	1
1.1 - APRESENTAÇÃO DO TEMA	1
1.2 - OBJETIVOS	3
1.2.1 - <i>Objetivos Gerais</i>	3
1.2.2 - <i>Objetivos Específicos</i>	3
1.2.3 - <i>Organização do Trabalho</i>	4
2 - VOZ HUMANA	5
2.1 - INTRODUÇÃO	5
2.2 - FORMAÇÃO DA VOZ	5
2.3 - PROPRIEDADES DA VOZ	7
2.4 - TIPOS DE LESÕES NA VOZ	8
2.5 - MANUTENÇÃO DA VOZ	11
3 - PROCESSAMENTO DIGITAL DE SINAL	13
3.1 - INTRODUÇÃO	13
3.2 - BREVE HISTÓRICO	14
3.3 - O PARADIGMA DOS QUATRO UNIVERSOS	15
3.4 - TRANSFORMADA DE FOURIER	16
3.4.1 - <i>Introdução</i>	16
3.4.2 - <i>Histórico</i>	17
3.4.3 - <i>Transformadas</i>	19
3.4.4 - <i>Aplicações</i>	23
3.4.5 - <i>Algoritmo da Transformada de Fourier</i>	23
3.4.6 - <i>Propriedades da Transformada de Fourier</i>	26
3.4.7 - <i>Exemplo Prático</i>	28
3.5 - IMPLEMENTAÇÃO DA TRANSFORMADA DE FOURIER	31
3.6 - APLICAÇÕES DA IMPLEMENTAÇÃO DA DFT	33
3.6.1 - <i>Primeiro Procedimento</i>	34
3.6.2 - <i>Segundo Procedimento</i>	35
4 - RNAs NO RECONHECIMENTO DE STRESS NA VOZ	38
4.1 - APLICAÇÃO	38
4.2 - TREINAMENTO DAS RNAs SEM O PRÉ-PROCESSAMENTO DA DFT	40
4.2.1 - <i>Aplicação</i>	40
4.2.2 - <i>Conclusão</i>	44
4.3 - TREINAMENTO DAS RNAs COM O PRÉ-PROCESSAMENTO DA DFT	45
4.3.1 - <i>Pré-processamento da DFT nas vozes</i>	45
4.3.2 - <i>Processamento das vozes pela RNA</i>	49
4.3.3 - <i>Conclusão</i>	53
5 - CONCLUSÕES E RECOMENDAÇÕES FINAIS	55
6 - REFERÊNCIAS BIBLIOGRÁFICAS	56

Lista de Figuras

Figura 2.1 – Anatomia da garganta.....	4
Figura 2.2 – Nódulo - Sintoma provocado na garganta.....	5
Figura 2.3 – Pólipos na laringe.....	7
Figura 2.4 – Edema provocado na garganta.....	13
Figura 2.5 – Garganta com cisto.....	15
Figura 3.1 – Paradigma dos Quatro Universos.....	16
Figura 3.2 – Jean Baptiste Joseph Fourier (1768 – 1830).....	17
Figura 3.3 – Pierre Simon Laplace.....	18
Figura 3.4 – Analogia entre Continuidade e Periodicidade.....	19
Figura 3.5 – Software MATLAB.....	28
Figura 3.6 – Espectro de potência.....	29
Figura 3.7 – Sinal de voz de 2 segundos.....	30
Figura 3.8 – Espectro de potência do sinal de voz de 2 segundos.....	31
Figura 4.1 – Software Sound Forge.....	32
Figura 4.2 – Ilustração das vozes gravadas.....	33
Figura 4.3 – Ilustração das etapas utilizadas no treinamento da rede MLP sem o pré-processamento da DFT.....	34
Figura 4.4 – Erro encontrado no treinamento da rede MLP.....	35
Figura 4.5 – Filtragem do sinal de voz.....	36
Figura 4.6 – Ilustração das Etapas utilizadas no treinamento da rede MLP com o pré-processamento da DFT.....	37
Figura 4.7 – Erro encontrado no treinamento da rede MLP.....	39

Lista de Tabelas

Tabela 2.1 - Cordas vocais humana	33
Tabela 3.1 – Relações de domínios de tempos dos sinais	43
Tabela 3.2 – Frequências e tempos do sinal a ser analisado	44
Tabela 3.3 – Etapas da Transformada Discreta de Fourier	45
Tabela 3.4 – Etapas do algoritmo da DFT	45
Tabela 3.5 – Comparação entre o sinal original e sua inversa.....	46
Tabela 3.6 – Correlação entre a voz de 2 segundo e sua inversa.....	47
Tabela 4.1 – Disposição das vozes utilizadas na rede MLP	48
Tabela 4.2 – Codificação dos diagnósticos de voz	48
Tabela 4.3 – Porcentagem de acerto das vozes <i>treinadas</i> na rede MLP.....	49
Tabela 4.4 – Porcentagem de acerto das vozes <i>generalizadas</i> pela rede MLP	51
Tabela 4.5 – Comparação entre as vozes calmas e de estresse.....	52
Tabela 4.6 – Cálculo do espectro de potência das vozes	53
Tabela 4.7 – Comparação entre a voz original e a inversa do sinal.....	54
Tabela 4.8 – Disposição das vozes pré-processadas utilizadas na rede MLP.....	54
Tabela 4.9 - Codificação dos diagnósticos das vozes pré-processadas	56
Tabela 4.10 – Porcentagem de acerto das vozes <i>treinadas pela</i> rede MLP.....	57
Tabela 4.11 – Porcentagem de acerto das vozes <i>generalizadas pela</i> rede MLP	58
Tabela 4.12 – Comparação entre as vozes pré-processadas calmas e nervosas	59
Tabela 4.13 – Comparação entre os dois treinamentos realizados	60

RESUMO

Em trabalho anterior (Matuck, 2004), além do estudo conduzido abordando técnicas de Redes Neurais Artificiais (RNAs), foi realizado um processamento para reconhecimento de estados emocionais de falantes através de sinais de vozes captadas digitalmente, sem qualquer pré-processamento. Os resultados mostraram-se inconclusivos sobre a capacidade da rede neural de identificar o estado emocional dos falantes, já que o índice de acerto foi em torno de 50%. A motivação para o trabalho veio da hipótese de que é possível utilizar sinais de voz para análise de níveis de stress de uma pessoa que se encontra em estado de isolamento social (confinamento). Como tentativa de caracterizar melhor os padrões de voz para o reconhecimento, este trabalho tem como foco a aplicação de técnicas de Fourier para o pré-processamento de dados visando a redução de ruídos presentes nos sinais. A melhoria no reconhecimento das vozes mostra que os ruídos originados pelo processo de aquisição das vozes precisam ser removidos para melhorar a eficiência do sistema de reconhecimento.

1 - INTRODUÇÃO

1.1 - Apresentação do tema

Em Matuck (2004) foi abordado o problema das consequências relacionadas ao isolamento social forçado, como exemplo indivíduos que trabalham em locais de difícil acesso e com grupo muito reduzido de pessoas, quando não há possibilidades de se fazer avaliações diretas, no sentido médico-paciente para diagnosticar alterações comportamentais. Como esse confinamento pode produzir mudanças no estado emocional (aumento do nível de stress, por exemplo), este trabalho foca um estudo relacionado à voz humana que possa permitir uma avaliação capaz de identificar, com confiabilidade, padrões na mudança do estado emocional de indivíduos nessas condições. Assim, neste trabalho utiliza-se técnicas de Redes Neurais Artificiais (RNAs - Neural Networks [HAY – 94]) para realizar tal avaliação.

As RNAs são sistemas computacionais com propriedades particulares como: *capacidade de aprendizagem ou de adaptação* aos dados utilizados durante o seu treinamento; *generalização*, ou seja, a capacidade de reconhecer dados nunca vistos anteriormente; e são *tolerantes a falhas*. A capacidade de aprendizagem permite que as RNAs possam diferenciar padrões de comportamentos através da sua exposição de exemplos, que são apresentados várias vezes para as Redes Neurais durante o processo de treinamento, o que faz com que estas extraiam as principais características inerentes dos dados.

Durante boa parte do tempo destinado ao projeto foi conduzido um estudo relacionado à tecnologia de RNAs bem como à implementação de alguns modelos específicos existentes na literatura. Para a realização de tal estudo, foi necessário um estudo prévio relacionado ao funcionamento do sistema nervoso humano, abordando os níveis de organização estrutural do cérebro, a plasticidade neuronal e todo funcionamento de um neurônio biológico.

Para colocar em prática o aprendizado relacionado aos estudos das Redes Neurais Artificiais foram implementados alguns tipos de arquiteturas como a rede Perceptron de Múltiplas Camadas (Multilayer Perceptron), a rede Perceptron de Camada Simples (Singlelayer Perceptron), a rede Kohonen, a rede Adaline e a rede Madaline (extensão do modelo Adaline). Essas redes foram utilizadas em aplicações como, por exemplo, o reconhecimento de caracteres e testes com portas lógicas (portas AND, OR e XOR). Após o treinamento destas aplicações foi verificada a capacidade destas redes de generalizarem aos novos dados nunca vistos anteriormente e também foi realizada a comparação entre estas arquiteturas para verificar qual rede era mais adequada para os problemas citados.

Em outro estudo realizado se verificou a capacidade das RNAs de identificação de padrões em testes psicológicos que são utilizados no reconhecimento na mudança do estado emocional de uma pessoa. Para tal utilizou-se dois testes específicos. O primeiro é o teste

VARK desenvolvido por Fleming e Mills (1992), o qual tem por finalidade investigar as preferências por modos pelos quais as pessoas trabalham com informação. Este teste foi treinado e comparado usando todas as arquiteturas de RNAs citadas, buscando o grau de aprendizado e a mais indicada para a identificação destes padrões.

Em outro experimento, utilizou-se o teste psicológico MPCL (Mooney Problem Check List) desenvolvido por Leonard V. Gordon e Roos L. Mooney (1950), que tem como objetivo ajudar indivíduos a expressarem seus problemas pessoais. Este teste é útil, por exemplo, para aumentar a compreensão do professor sobre seus estudantes onde os questionários são separados por faixa etária e abrangem áreas como: saúde e desenvolvimento, repouso e família, moral e religião, sexo e casamento. Este teste foi treinado nas arquiteturas de RNAs mencionadas para comparação e verificação da rede que produz um melhor diagnóstico para esta avaliação.

Como um último experimento procurou-se analisar sinais de voz para a identificação de níveis de stress. Através de um gravador portátil foi produzido um banco de dados contendo vozes de pessoas em situações calmas e de stress, através da leitura de um texto pré-determinado. Uma parte desse banco de vozes foi treinado pela rede Perceptron de Múltiplas Camadas obtendo 98,78% de reconhecimento pela rede neural. Após o treinamento, foi verificada a capacidade da rede de generalizar vozes nunca vistas anteriormente, apresentando a segunda parte do banco de vozes para o seu reconhecimento constatando um reconhecimento de 50,2%.

Uma desvantagem percebida neste experimento foi durante a formulação do banco de vozes pelo gravador portátil. Junto à gravação percebeu-se a presença de ruídos anexados às vozes como chiado e fatores externos em geral, dados que influenciaram a qualidade de aprendizado pela rede neural.

Este projeto tem o objetivo de produzir uma avaliação com maior confiabilidade na identificação de padrões de stress presentes na voz, pois acredita-se no fato de que a voz pode ser um grande indicador de padrões emocionais presentes no ser humano. Assim, procurou-se realizar estudos de técnicas de pré-processamento de dados, para o treinamento das redes neurais, com o objetivo de reduzir as fontes de ruídos e chiados que podem ser a causa de um baixo desempenho das redes neurais na identificação de padrões de voz.

Para a identificação desses padrões essenciais da voz há alguns mecanismos existentes na literatura que processam o sinal de voz armazenando somente suas características fundamentais, como por exemplo, a Transformada de Wavelet (muito utilizadas na extração de padrões de variação de temperatura), Transformada de Laplace (transformada de variável), Transformada K, Transformada Z, entre outros.[GM – 99]

Em Matuck (2004) o treinamento das RNAs baseava-se no sinal no domínio do tempo, onde o sinal era analisado na sua forma original pela rede. Para a grande parte das aplicações em processamento digital de sinais, a representação destes no domínio do tempo não é a mais adequada. Muitas vezes, as informações cruciais estão contidas no chamado espectro em frequência do sinal, que mostra as frequências existentes no sinal que está sendo analisado. Por exemplo, o espectro em frequência de um sinal de eletrocardiograma é de grande valia para que um cardiologista possa saber se o paciente apresenta alguma patologia ou não. Baseando-se nisso há o interesse de um tratamento do sinal e sua

utilização no domínio da frequência armazenado as características fundamentais específicas de cada voz para utilizá-las no treinamento e reconhecimento das RNAs.

Neste projeto utilizou-se a Transformada de Fourier para o pré-processamento da voz antes do treinamento das Redes Neurais Artificiais. A Transformada de Fourier é utilizada em diversas aplicações como: análise, síntese e codificação de voz; sinais de radar; avaliação de falhas mecânicas (motores); análise de imagens; telecomunicações; entre outras. Acredita-se que com o pré-processamento da voz, o treinamento da rede neural é otimizado e a capacidade de aprendizagem e identificação a novos padrões é mais promissora.

1.2 - Objetivos

1.2.1 - Objetivos Gerais

O principal objetivo desse trabalho é a realização de um processamento de sinais de voz para identificação de padrões comportamentais por Redes Neurais Artificiais.

1.2.2 - Objetivos Específicos

- Adquirir conhecimentos sobre Transformada de Fourier (Fourier Transform);
- Implementação de alguns modelos de Fourier em aplicações existentes atualmente na literatura;
- Utilização de Redes Neurais Artificiais na caracterização de características comportamentais de um falante;
- Complementação de conhecimento sobre Redes Neurais Artificiais;
- Implementação do protótipo básico.

1.2.3 - Organização do Trabalho

Este trabalho está disposto em 5 capítulos. O capítulo 1 trata dos aspectos introdutórios do trabalho, como a apresentação do tema e sua ligação com o trabalho anterior em Matuck (2004).

O capítulo 2 faz uma abordagem geral sobre a voz humana, demonstrando sua estrutura física bem como algumas propriedades inerentes à voz. Também é comentado sobre algumas lesões que podem ocorrer e prejudicar a voz e algumas recomendações para uma manutenção saudável.

No capítulo 3 é feita uma introdução sobre a área de processamento digital de sinal, bem como o seu surgimento para mundo. Também são abordados conceitos da Transformada de Fourier como sua história, alguns tipos de transformadas existentes na literatura e algumas aplicações desta técnica de processamento. Foi realizado neste capítulo todo um estudo abordando técnicas da Transformada de Fourier como o seu algoritmo, algumas propriedades pertinentes a esta metodologia e um exemplo matemático prático. Neste capítulo também se encontra uma implementação da Transformada Discreta de Fourier através do software Matlab e alguns exemplos utilizando este procedimento.

Já no capítulo 4 são referenciados dois procedimentos abordando o reconhecimento de stress da voz humana por técnicas de Redes Neurais Artificiais. O primeiro procedimento mostra o treinamento das vozes pela rede Perceptron de Múltiplas Camadas (MLP – Multilayer Perceptron) sem o pré-processamento pela Transformada Discreta de Fourier. Já o segundo experimento mostra o mesmo treinamento, mas com o pré-processamento das vozes para o domínio da frequência. No término deste capítulo são realizadas comparações sobre estas duas aplicações.

No capítulo 5 são feitas as considerações finais, propostas para trabalhos futuros dando de certa forma uma continuidade ao estudo geral deste projeto e conclusões do trabalho.

2 - VOZ HUMANA

2.1 - Introdução

A voz existe pelo desejo que o ser humano possui de verbalizar pensamentos, emoções e opiniões, sendo parte integrante da nossa identidade. É uma das extensões mais fortes da nossa personalidade, muitas vezes é possível reconhecer alguém só pela voz. A voz transmite carinho, simpatia, ansiedade, tensão, emoção, inclusive a nossa idade e experiência de vida. O seu uso incorreto pode levar a uma *disfonia*, que é um sintoma de dificuldade na emissão normal da voz, podendo gerar alterações nas pregas vocais (hiperemia, edema, nódulo, pólipos, etc).

Para que a produção vocal seja adequada, há necessidade do equilíbrio entre os parâmetros vocais que são os representados pelas qualidades vocais, constituídas por: ressonância, tipo de voz, intensidade, frequência, articulação, velocidade de fala, pronúncia e aspectos referentes à emissão. Em algumas situações este equilíbrio não acontece sendo responsável por aspectos como rouquidão, aspereza, tensão, hiper-nasalidade, dentre outras que podem ser encontradas, tanto na voz do adulto como na infantil.

Desde sempre a espécie humana pretendeu criar máquinas que produzissem e entendessem a voz humana. A utilização de voz para interagir com sistemas automáticos tem um vasto campo de aplicações. A combinação com a rede telefônica permite acesso remoto a bases de dados e a novos serviços como, por exemplo, a consulta de correio eletrônico via telefone de qualquer ponto do mundo e a consulta de horários de vôos sem necessidade de um operador. A utilização destas tecnologias facilita a integração de pessoas com deficiências. Por exemplo, o uso de síntese de voz permite a uma pessoa muda, usando um teclado, usar o telefone para fazer coisas simples como encomendar uma pizza.

2.2 - Formação da Voz

Como todos os outros sons, a voz é produzida por vibrações. No interior da laringe estende-se um tecido esticado com duas pregas, chamadas *cordas vocais*. São elas que vibram quando falamos. As cordas são fibras elásticas que se distendem ou se relaxam pela ação dos músculos da laringe conforme exibido na Figura 2.1. A frequência natural da voz é determinada, em grande parte, pelo comprimento das cordas vocais. As mulheres têm voz mais aguda que os homens porque suas cordas vocais são mais curtas. Pela mesma razão, as vozes das crianças são mais agudas do que as dos adultos.

A laringe não é o único órgão responsável pela fonação. Os lábios, a língua, os dentes, o véu palatino e a boca concorrem para a formação dos sons. Quando murmuramos

produzimos os sons colocando a boca e a língua em determinadas posições sem fazer vibrar as cordas vocais. O timbre da voz humana depende dos inúmeros espaços que vibram em ressonância com as cordas vocais. Aí se incluem as cavidades ósseas, cavidades nasais, a boca, a garganta, a traquéia e os pulmões, bem como a própria laringe.

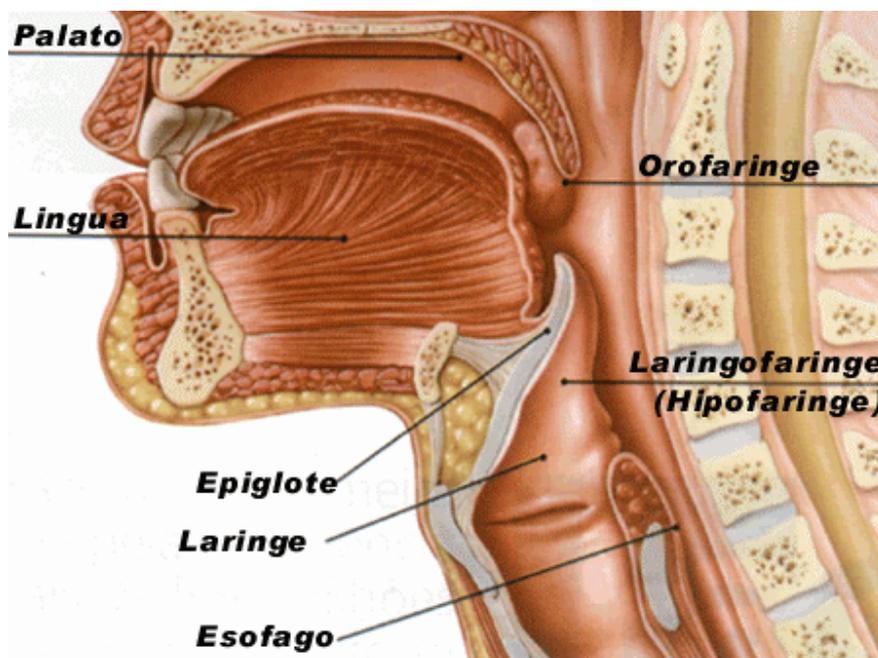


Figura 2.1 – Anatomia da garganta.

Todo o ar inspirado e expirado passa pela laringe onde nos dois lados dela se encontram as cordas vocais relaxadas sem produzir som. Quando uma pessoa fala ou canta, seu cérebro envia mensagens pelos nervos até os músculos que controlam as cordas vocais. Os músculos fazem a aproximação das cordas de modo que fique apenas um espaço estreito entre elas. Quando o diafragma e os músculos do tórax empurram o ar para fora dos pulmões, ele produz a vibração das cordas vocais. O controle da altura do som se faz aumentando-se ou diminuindo-se a tensão das cordas vocais. Os diferentes estados das cordas vocais humana é exibido na tabela 2.1 abaixo.

Cordas vocais afastadas. Silêncio	Cordas vocais encostadas. Produção de som	Cordas vocais entreabertas. Sussurro

Tabela 2.1 Cordas vocais humana.

A mais baixa freqüência que pode dar a audibilidade de um tom é mais ou menos a de **20** vibrações por segundo, enquanto a mais alta se encontra entre **10.000** e **20.000** vibrações por segundo. A freqüência comum de um piano é de **40** a **4.000** vibrações por segundo. A freqüência da voz humana se encontra entre **60** e **1.300** vibrações por segundo

Para emitir corretamente a voz, é necessária uma colocação correta da respiração, sem tensão ou relaxação das pregas vocais. Estas devem ter uma tonicidade "ideal" para que sua movimentação ocorra normalmente. A respiração deverá ter uma coordenação entre inspiração e expiração, coordenação fono-respiratória, diretividade do sopro expiratório, movimentos livres da musculatura da região torácica e do músculo diafragmático. Uma respiração ideal provoca o alargamento da base dos pulmões, permitindo maior entrada de ar e menor contração da musculatura torácica, pescoço e parte inferior da face (mandíbula e língua).

A fonação é um poderoso veículo de comunicação pelos seus aspectos lingüísticos e pela entonação da voz. Para que a fonação seja normal é necessário que, além do aparelho fonador, a laringe funcione adequadamente e em sinergia, que os mecanismos respiratórios, os de ressonância e com o sistema nervoso estejam adaptados à fonação. Mas as qualidades vocais vêm junto com o nascimento do indivíduo. Seria necessário aperfeiçoá-lo, a partir dessas condições inatas. O principal é perceber as virtudes ou defeitos da própria voz, para aproveitar as primeiras e corrigir, na medida do possível, as segundas.

2.3 - Propriedades da Voz

Existem três elementos determinantes da voz humana: o organismo, o ambiente e a personalidade. A parte orgânica condiciona a voz, pois esta depende especialmente da conformidade do aparelho vocal e do estado físico geral. O ambiente, também tem a sua importância, pois muitos hábitos de locução provêm da imitação ou contágio, como por exemplo, o sotaque regional. Por último, a personalidade também influi. A inteligência, a entoação, a dicção, etc, revelam, sem dúvida alguma, o caráter e o próprio eu. A voz humana tem, como qualquer outro som, qualidades próprias:

- **Tom** - é a altura musical da voz. Segundo o tom, as vozes humanas classificam-se em agudas ou graves. A escala de registro e de altura permite classificar as vozes masculinas, geralmente em três categorias: tenor, barítono e baixo. Também existem tipos de vozes intermediárias. Sob o ponto de vista oratório, a melhor voz, é a do barítono;
- **Timbre** - é o matiz pessoal da voz. É um fenômeno complexo e está determinado pelo tom fundamental e pelos seus harmônicos ou secundários. Reconhece-se pelo timbre característico a pessoa com a qual se fala. Há vozes bem timbradas e agradáveis, mas também existem roucas, agudas e chiadas;
- **Quantidade** - é a duração do som. Segundo a quantidade, os sons podem ser longos ou curtos, com toda a gama intermediária de semi-longos, semi-curtos,

etc. A quantidade depende, geralmente, das características de cada língua, dos costumes lingüísticos das regiões ou países, da psicologia do habitante, etc.

- **Intensidade** - é a maior ou menor força com que se produz a voz. Há vozes fracas e vozes fortes.

Em fonética denomina-se sotaque ao conjunto dos elementos anteriores, cuja combinação especial em cada língua e mesmo em cada indivíduo, dá a essa língua uma característica diferente daquela ouvida no inglês ou no francês. Por exemplo, dentro do campo lingüístico do espanhol, há um sotaque argentino, um outro mexicano e tantos sotaques quantos países onde o espanhol é falado. A diferença existe também entre zonas ou estados de um mesmo país. Também podemos dizer que se encontram diferenças em bairros de uma mesma cidade.

Existem na literatura, cinco sistemas básicos de fala em aspectos referentes à emissão da voz:

- **Respiração** - Exerce um papel especial na fala. O fôlego e o controle respiratório são instrumentos vitais para que não ocorram frases interrompidas e freqüentes interrupções durante a fala;
- **Fonação** - Se refere à qualidade e às características da voz produzida pela laringe;
- **Ressonância** - É a modificação seletiva da inflexão na voz quando a corrente de ar passa através da rinofaringe, orofaringe e da boca. Esta modulação ou amplificação cria as características individuais da voz;
- **Articulação** - É definida como a produção dos sons da fala por meio de paradas ou constrições da corrente de ar, vocalizada ou não-vocalizada, pelos movimentos dos lábios, língua, véu palatino ou faringe. Isto exige os movimentos dos lábios, da língua, dos dentes, do palato duro ou mole;
- **Prosódia** - Refere-se à qualidade normal da fala, incluindo velocidade, sincronização, intervalo, regulagem, melodia e ênfase.

Finalmente, os sons têm uma escala diferente de percepção ou alcance. Existem sons que, por natureza própria, conseguem-se ouvir de muito longe, como a vogal "a", enquanto outros são escutados a uma distância mais curta, como a "u". As vozes têm também diferente alcance ou percepção, segundo as pessoas.

2.4 - Tipos de lesões na Voz

Os principais tipos de lesões orgânicas resultantes das disfonia funcionais são: nódulos, pólipos e edemas das pregas vocais. Estas três alterações da mucosa da prega vocal têm como característica comum, o fato de representarem uma resposta inflamatória da túnica mucosa a agentes agressivos, quer sejam de natureza externa, quer sejam decorrente do próprio comportamento vocal. Será abordado agora alguns tipos de lesões provocadas na voz.

- **Nódulos**

Os nódulos (Figura 2.2) resultam de: fatores anatômicos predisponentes (fendas triangulares), personalidade (ansiedade, agressividade, perfeccionismo) e do comportamento vocal inadequado (uso excessivo e abusivo da voz). O tratamento dos nódulos é *fonoterápico*. A indicação cirúrgica, todavia, pode ser feita quando os mesmos apresentam características esbranquiçada, dura e fibrosada, ou ainda quando existe dúvida diagnóstica.

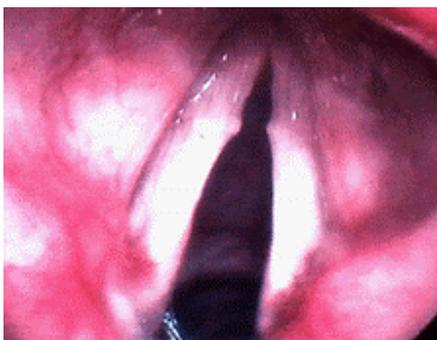


Figura 2.2 – Nódulo – Sintoma provocado na garganta.

- **Pólipos**

Os pólipos (Figura 2.3) são inflamações decorrentes de traumas em camadas mais profundas da lâmina própria da laringe, de aparência vascularizada. O tratamento é *cirúrgico*. A voz típica é rouca. As causas podem estar relacionadas a fatores como abuso da voz ou agentes irritantes, alergias, infecções agudas, etc.

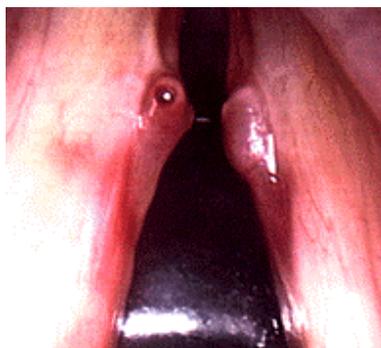


Figura 2.3 – Pólipos na laringe.

- **Edemas das cordas vocais**

Os edemas relacionam-se com o uso da voz, sendo. Normalmente são localizados e agudos (Figura 2.4). O tratamento é medicamentoso ou através de repouso vocal. Os edemas generalizados e bilaterais representam a laringite crônica, denominada *Edema de Reinke*. Este sintoma é encontrado em pessoas expostas a fatores irritantes externos, especialmente o tabagismo (fumo) e o elitismo, sendo o mais importante fator associado ao uso excessivo e abusivo da voz. Quando discretos, os edemas podem ser tratados com medicamentos e fonoterapia, assegurando-se a eliminação de seu fator causal. Os edemas quando volumosos, necessitam de remoção cirúrgica, seguida de reabilitação fonoaudiológica.

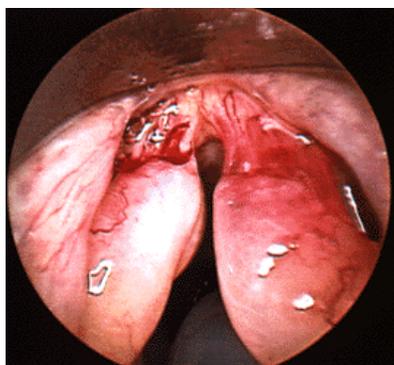


Figura 2.4 –Edema provocado na garganta.

- **Infecções em geral**

Os fatores infecciosos, incluindo as sinusites, diminuem a ressonância e alteram a função respiratória, produzindo modificações na voz. O efeito primário das infecções das vias aéreas superiores têm efeito direto sobre a faringe e a laringe, podendo provocar irritação e edema das pregas vocais. Estes processos infecciosos podem gerar atividades danosas, como o pigarro e a tosse que, por sua vez, podem causar traumatismos nas pregas vocais. Há também fatores imunológicos, endócrinos, auditivos e emocionais, que podem causar transtornos na emissão da voz.

- **Laringite**

A laringite é a inflamação da laringe (onde estão as cordas vocais) e das áreas próximas. É um sintoma que ocorre no resfriado, bronquite, pneumonia, e outras infecções respiratórias. Há duas formas de laringite: aguda e crônica. A laringite aguda acontece de repente e não dura muito tempo. A laringite é chamada crônica quando a rouquidão dura

um longo período. Os sintomas são: rouquidão e tosse, com sensação de corpo estranho na garganta, aumento de secreção, pigarro e, ocasionalmente, dor de garganta.

A laringite aguda é geralmente causada por um vírus, mas pode resultar de uma infecção bacteriana. A laringite crônica pode ser causada por tabagismo (ritmo intenso), uso intenso da voz (quando se fala ou canta em tom muito alto), tosse forte, ou exposição a substâncias irritantes.

O tratamento envolve a eliminação dos fatores que provocam a irritação da laringe (exposição a produtos químicos e tóxicos, nível elevado de ruídos, maus hábitos alimentares, refluxo alimentar devido a gorduras, pigarro crônico, etc.), além da promoção de hábitos que melhoram a higiene vocal, evitando os abusos da voz.

- **Cisto**

O Cisto (Figura 2.5) são lesões benignas que se parecem com pequenas esferas nas margens das pregas vocais. Geralmente são congênitos (nasce com a pessoa). Podem ser decorrentes do bloqueio de um ducto glandular da mucosa cordal, no qual há retenção de muco, principalmente após abusos vocais. O tratamento, geralmente, é cirúrgico. O principal sintoma vocal é a rouquidão.



Figura 2.5 – Garganta com cisto.

2.5 - Manutenção da Voz

É importante a preservação da voz para mantê-la sempre apta a ser usada, principalmente para aqueles que a tem como instrumento de trabalho. Os cuidados apresentados a seguir foram pensados de forma que a pessoa possa preservar a sua voz e conscientizar-se da importância de sua manutenção:

- Beber de sete a oito copos de água por dia;

- Buscar atendimento especializado se a voz for o seu principal instrumento de trabalho;
- Manter uma higiene bucal;
- Pastilhas, sprays ou medicamentos, só indicados por Médicos;
- Evitar auto-medicação e soluções caseiras (gingibre, romã, etc.);
- *Fumo*: Ao tragar, a fumaça agride todo o sistema respiratório e, também, as pregas vocais. O fumo é considerado um dos principais fatores desencadeantes do câncer de laringe e pulmão;
- *Drogas*: O uso de drogas tem ação direta sobre a laringe. A maconha é extremamente lesiva, irritando a mucosa da região do trato vocal;
- *Evitar bebidas alcoólicas*, pois o álcool tem um efeito anestésico, provocando a redução da sensibilidade, é quando na maioria das vezes ocorre um abuso vocal, lesando as pregas vocais;
- Evitar a competição sonora - Falar com um ruído de fundo leva a pessoa à "competir" com o som, ou seja, a pessoa fala mais alto para que possa ser entendida. É fundamental que se mantenha a intensidade vocal em um nível moderado, sem falar nem muito alto e nem sussurrado, pois ambas apresentam um esforço maior do que o necessário;
- Evitar mudanças bruscas de temperatura interna (ingestão sucessiva de alimentos quentes e muito gelados), ou mesmo externa, principalmente quando ao ar livre ou em ambientes aquecidos ou resfriados artificialmente;
- Manter uma postura relaxada ao falar ou cantar.

Nota-se então a importância de se ter um bom cuidado com a voz para que não possibilite o surgimento de lesões abordadas neste tópico. Para uma pessoa que faz da voz um instrumento de trabalho é imprescindível que esta manutenção seja praticada diariamente e sempre antes de utilizá-la deve-se realizar algumas técnicas de aquecimento da voz, entre outros procedimentos.

3 - PROCESSAMENTO DIGITAL DE SINAL

3.1 - Introdução

Deve-se ressaltar que só será abordado neste capítulo apenas alguns conceitos introdutórios relacionados a técnicas de processamento digital de sinal, tendo-se como foco principal o estudo da Transformada de Fourier.

Toda a nossa vida se baseia em sinais, que são medidos, processados, analisados, e dão origem a decisões. O som, a temperatura e a luz são exemplos de sinais que utilizamos no dia a dia. Os ouvidos convertem o som em sinais elétricos, que chegam ao cérebro, e este é capaz de analisar algumas das suas propriedades, tais como amplitude, frequência e fase, determinar a direção em que se encontra a fonte de som, e reconhecê-lo, como música, fala, o ruído de um automóvel, etc. Os nervos colocados nas partes expostas da pele sentem a temperatura e enviam para o cérebro sinais elétricos, que podem originar decisões tais como ligar um aquecedor, abrir uma janela, etc. Os olhos focam as imagens na retina, que converte essas imagens em sinais elétricos e os envia para o cérebro, que, pela análise da cor, da forma, da intensidade, etc., da luz é capaz de reconhecer objetos, medir distâncias, detectar o movimento, etc.

Os fantásticos desenvolvimentos nos últimos anos presenciados na área da microeletrônica tornaram possível pôr em prática este pensamento de uma forma efetiva, e está na origem do Processamento Digital de Sinal (PDS), disciplina que hoje ocupa um papel preponderante em várias áreas da ciência. Do ponto de vista da engenharia, *sinais* são funções ou seqüências que servem para transportar informação de uma fonte de mensagens a um destinatário. As características específicas dos sinais dependem do canal de comunicações utilizado para este transporte. Estes sinais são processados no lado do transmissor com a finalidade de produzi-los e configurá-los, e no lado receptor para extrair a informação neles contida, se possível com a máxima eficiência.

Em sistemas eletrônicos a fonte geradora de informação, o canal de comunicação e o destinatário são elementos pré-definidos com características em geral muito bem definidas, como por exemplo, em sistemas de transmissão telefônica, de irradiação, de sinais de rádio ou de televisão ou de detecção de alvos com radar sonar. Em outras situações, como nos processos de medição em investigação científica, a fonte de mensagens e o canal de comunicações poderão estar apenas parcialmente caracterizados. Sinais bioelétricos como o eletrocardiograma, eletroencefalograma, entre outros, são estudados há décadas com a finalidade de se extrair informação sobre estados patológicos de órgãos, sem que se tenha muitas vezes a mínima certeza de que tal informação é de fato transportada por estes sinais.

Os sinais que constituem a voz humana codificam uma variedade de informações como, por exemplo, sobre a semântica do que está sendo dito, sobre a identidade do locutor e até sobre o seu estado de espírito. As redes neurais biológicas do sistema nervoso humano têm um desempenho de invejável eficiência na extração desta informação.

3.2 - Breve Histórico

É por volta do século XIX que apareceram os modelos matemáticos básicos dos sinais e sistemas contínuos, com as Transformadas de Laplace e de Fourier onde seus trabalhos contribuíram para muitas áreas da matemática, ciência e engenharia. Pierre Simon, Marquês de Laplace, o maior astrônomo teórico depois de Newton, nascido vinte anos antes de Fourier, aplicou os seus conhecimentos matemáticos ao estudo dos movimentos planetários, dando origem à hoje designada transformada de Laplace, que cedo encontrou aplicação em muitas outras áreas científicas.

Também De Moivre, que em 1730 introduziu a hoje chamada transformada em z , deve ser creditado como um dos precursores do Processamento Digital de Sinal. No entanto, é ao advento dos computadores digitais, verificado nos anos 40, que se deve o nascimento do PDS como disciplina. Nos anos 50, engenheiros e cientistas como Shannon e Bode nos Bell Telephone Laboratories e Linville no MIT foram certamente dos primeiros a equacionar a utilização de computadores de sinal em processamento de sinal. No início dos anos 60, Kaiser, nos laboratórios Bell, apresentou importantes contribuições para a análise e a síntese de filtros digitais.

Já em 1965, James W. Cooley e Jhon Tukey publicaram um artigo em que apresentavam um algoritmo que reduz consideravelmente o tempo de cálculo da Transformada Discreta de Fourier. O algoritmo veio a ser conhecido como Transformada Rápida de Fourier ou simplesmente FFT (Fast Fourier Transform) e é considerado o mais avançado, deste século, no campo da análise numérica. Posteriormente, encontraram-se na literatura trabalhos anteriores aos de Cooley e Tukey em que são apresentados algoritmos semelhantes a FFT. Os trabalhos de Runge, por exemplo, foram publicados em 1903 e 1905 [BRI-74]. Talvez esses trabalhos anteriores ao de Cooley e Tukey não tenham tido grande repercussão porque na época não havia computadores suficientemente rápidos.

Na década de 1970 surgiram microcomputadores menores, mais rápidos e mais baratos que os computadores mais antigos. Esse fato, juntamente com a repercussão do artigo de Cooley e Tukey, resultou num renovado interesse na Transformada Rápida de Fourier. Atualmente, o cálculo da FFT não está mais limitado a funções expressas analiticamente e isso, por sua vez, possibilitou avanços tecnológicos em diversas áreas como, por exemplo, processamento de sinais e espectroscopia atômica e molecular.

Atualmente, o PDS emergiu das aplicações militares onde nasceu e desempenha um papel chave em produtos de consumo, industriais e de telecomunicações. Microprocessadores de sinal de baixo custo são componentes essenciais de jogos eletrônicos, telefones celulares, brinquedos, leitores de CDs, discos de computadores, modems, impressoras, sistemas de reconhecimento de voz e de conferência vídeo, e muitos outros produtos familiares. Cada vez mais aplicações tradicionalmente do domínio dos sistemas analógicos estão a encontrar soluções digitais mais baratas e mais confiáveis.

3.3 - O Paradigma dos Quatro Universos

Para se entender melhor este conceito que se chama processamento digital de sinais deve-se abordar o processo que é sair do mundo real onde os sinais são contínuos, e ir para o mundo do computador onde tudo é discreto. Este paradigma é conhecido como o *Paradigma dos Quatro Universos* (Figura 3.1) e é estruturado da seguinte maneira:

- **O Universo Físico**, onde estão localizados os objetos do mundo real onde será focado o estudo em questão. Estes objetos se caracterizam através da variação de uma determinada grandeza, podendo ocorrer em relação ao espaço ou tempo;
- **O Universo Matemático**, onde são formuladas descrições abstratas desses objetos. Contém uma descrição matemática formal dos objetos do mundo real, conhecido também como a *conceitualização* dos objetos do universo do mundo real;
- **O Universo de Representação**, que vai permitir trazer essas descrições abstratas para o mundo digital, e é onde se dará a discretização dos sinais contínuos. Este universo é constituído por descrições simbólicas e finitas associadas a componentes do universo matemático. Neste nível é realizado o mapeamento dos objetos conceituais para representações geométricas.
- **O Universo de Implementação** é um conjunto de regras específicas à implementação do modelo dentro de um ambiente computacional. É nesta arquitetura que acontece a *codificação* do sinal discretizado na memória do computador através de uma estrutura de dados. Várias estruturas de dados podem ser implementadas para um mesmo modelo conceitual, levando-se em conta alguns requisitos como, por exemplo, o desempenho, capacidade do equipamento, o volume de dados a ser processado, entre outros.

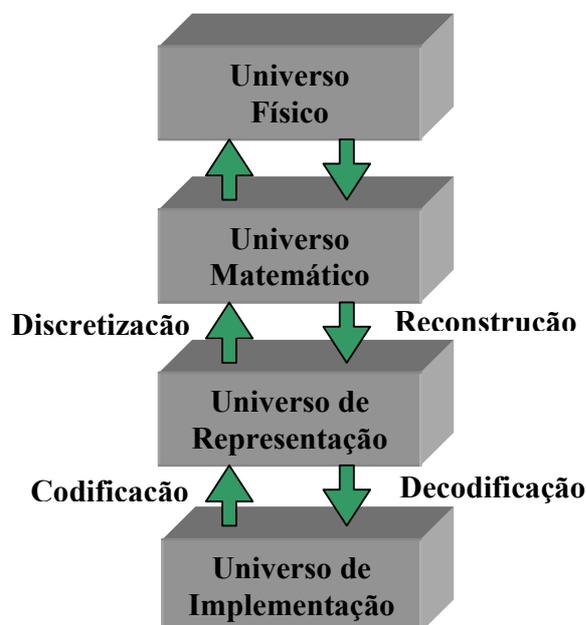


Figura 3.1 – Paradigma dos Quatro Universos.

Portanto, para estudar um determinado fenômeno, ou objeto da natureza, no computador, nós precisamos associá-lo a um modelo matemático e então encontrar uma representação discreta para esse modelo que pode ser implementada no computador. A codificação dos dados pode ser decodificada e depois reconstruída para podermos voltar com o fenômeno estudado para o mundo real.

3.4 - Transformada de Fourier

3.4.1 - Introdução

Novas tecnologias em sistema informatizados têm ganhado cada vez mais mercado e espaço dentre usuários de computadores do tipo PC (Personal Computer). Muitas dessas tecnologias têm como fundamento básico a aquisição e o processamento lógico de sinais. Não obstante, técnicas de processamento de sinais se fazem presentes com o intuito de melhorar e apurar melhor as informações obtidas de forma analógica.

Uma dessas técnicas é a transformação ou mapeamento de coordenadas. Usualmente, essa técnica específica é fruto da utilização de uma ferramenta matemática poderosa, a Transformada. A Transformada é um procedimento matemático que tem como principal finalidade a mudança ou o mapeamento de um conjunto de coordenadas em outro conjunto.

Uma ferramenta muito utilizada neste processamento é a transformada de Fourier, a qual nos permite ter uma visão do sinal a ser analisado no domínio da frequência, facilitando sobremaneira esta análise e o seu processamento, normalmente, aplicando-se técnicas de filtragem digital.

Um problema computacional surge quando se quer implementar essa ferramenta, o tempo de processamento ou conversão de um sinal em um sistema de coordenada para um outro em um sistema de interesse é considerável. Na prática esse problema pode ser solucionado com a utilização de algoritmos mais rápidos de transformadas que juntamente com os teoremas de convolução e da correlação permite, de maneira simplificada, a implementação das técnicas de filtragens para eliminação de ruídos e interferências dos sinais em análise.

Para entrarmos mais a fundo sobre estes conceitos de processamento de sinais e abordarmos a Transformada de Fourier é necessário antes saber quem idealizou esta técnica que é muito utilizada até hoje em vários ramos da ciência.

3.4.2 - Histórico

Jean-Baptiste Joseph Fourier, nasceu em 21 de Março de 1768, em Auxerre na França (Figura 3.2). Aos 12 anos, Fourier começou a mostrar parte do seu talento, redigindo sermões para sacerdotes de várias cidades. Dois anos mais tarde iniciou seus estudos de Matemática, conseguindo grande destaque. Considerado menino-prodígio, foi convidado a ingressar na ordem dos beneditinos, mas antes de ordenar-se, chegou a Revolução de 1789.

Fourier que sempre desejara ser militar aderiu com entusiasmo à causa da Revolução. Com a criação da Escola Normal e da Escola Politécnica, das quais foi conferencista, Fourier começou a desenvolver os trabalhos que o immortalizaram como matemático. Data dessa época sua teoria para calcular raízes irracionais das equações algébricas, cujo estudo Newton iniciara. Tendo acompanhado Napoleão no Egito, Fourier desenvolveu ali estudos de arqueologia, tornando-se especialista em egiptologia. Fourier trabalhou nessa época como engenheiro, dirigindo uma fábrica de armamentos do exército francês no Egito.



Figura 3.2 - Jean Baptiste Joseph Fourier (1768 - 1830).

Em 1802, Fourier lançou sua obra mais notável, a “Teoria Analítica do Calor” (*Théorie Analytique de la Chaleur*). Em seu livro, ele dedica toda uma seção à solução do problema de desenvolvimento de uma função qualquer em série de senos e co-senos de arcos múltiplos. Generalizou o procedimento, partindo de um caso específico para empregá-lo em qualquer caso.

Voltando à França em 1812, Fourier desenvolveu, na sua obra "Memorial", uma teoria sobre a condução do calor, tornando-se precursor da Física-Matemática. Neste último estudo, o matemático francês foi levado a criar um novo tipo de desenvolvimento em série,

diferente do método de Taylor por empregar funções periódicas em vez de potências, e que recebeu seu nome.

Antes de 1930 iniciou o estudo de wavelet (funções que satisfazem a certos requisitos matemáticos e são usadas na representação de dados ou de outras funções) com suas teorias de análise de frequência. Fourier afirmava que a análise em wavelets não é feita segundo a frequência, mas sim segundo a escala. Os algoritmos wavelet processam dados em diferentes escalas e resoluções, permitindo que sejam vistos tanto o global quanto os detalhes.

Fourier deu um passo decisivo, ao usar indiferentemente os símbolos de integração e o de somatória infinita, que conduziu às chamadas séries de Fourier. Coube a Fourier o mérito de haver criado esse instrumento matemático, de extraordinária fecundidade, com o qual as funções periódicas descontínuas pudessem ser apresentadas através de funções contínuas. Com o objetivo de definir a propriedades de condutividade térmica entre os materiais através de observações experimentais, Fourier definiu a lei básica de *condução de calor* entre sólidos avançando em muito as pesquisas nessa área. Em 1830 Fourier morreu; vítima de um aneurisma cerebral.

- **Condução de Calor**

Condução de calor é o processo de transferência de energia térmica de uma região mais quente para outra mais fria, através das partículas presentes no meio que as separa sem que haja deslocamento de matéria. Essa transferência de energia ocorre pois, a região de maior temperatura possui moléculas vibrando com intensidade maior (maior energia cinética). Com uma vibração mais acentuada cada molécula transmite energia para a molécula vizinha que passa a vibrar mais intensamente e assim sucessivamente.

Evidentemente, há materiais com propriedades de condução de calor mais eficientes que outros. Portanto diz-se que a transferência de calor é mais eficaz quanto melhor condutor for o material (o aço é um ótimo condutor). Essa eficiência de condução de calor é medida pela constante de proporcionalidade denominada *condutividade térmica*. Devido a maior aproximação de suas moléculas, os sólidos tendem a apresentar maiores valores de condutividade térmica em comparação com líquidos e gases. Como se percebe a condução de calor é um fenômeno que exige a presença de um meio material e que, conseqüentemente, não ocorre no vácuo.

De acordo com Fourier, a quantidade de calor que atravessa um condutor, por unidade de tempo, chama-se fluxo de calor (ϕ) e é calculado por:

$$\phi = \frac{Q}{\Delta t} \quad (3.1)$$

A lei de Fourier diz que, conforme exibido na Equação 3.2, a quantidade de calor Q que atravessa uma parede, sob uma diferença de temperatura constante, é diretamente

proporcional à área da secção transversal A , à diferença de temperatura (θ) entre as regiões separadas pela parede e ao tempo (Δt) de transmissão e inversamente proporcional à extensão atravessada, ou espessura (e) da parede.

$$Q = \frac{k.A.\Delta\theta.\Delta t}{e} \quad (3.2)$$

Como o calor se propaga de partícula para partícula, corpos mais densos, com maior número de partícula por unidade de volume, especialmente partículas livres, são bons condutores. Isto explica por que os metais são bons condutores. Pelo mesmo motivo, os líquidos e gases não são bons condutores de calor. Os materiais em que a condução térmica praticamente não ocorre são chamados isolantes térmicos, por exemplo, a madeira e o isopor.

3.4.3 - Transformadas

Sinais são quantidades físicas ou variáveis detectáveis por meio das quais informações podem ser transmitidas. Existem duas formas tradicionais de se representar sinais físicos:

- **Representação no domínio do tempo** - onde a amplitude do sinal é representada como uma função do tempo;
- **Representação no domínio da frequência** - onde a função que representa o sinal mostra a amplitude de cada frequência que o compõe.

Diversos estudos vêm sendo feitos na área de processamento digital de sinais com o intuito de melhorar a aquisição dos sinais que registram informações, necessárias para uma avaliação confiável. Na literatura existem vários tipos de transformadas que auxiliam na captura e interpretação da informação de um sinal, métodos que definem a melhor forma de sinal a ser utilizada, por exemplo, como entrada de uma Rede Neural Artificial. A seguir serão citadas algumas transformadas que são muito utilizadas em processamento digital de sinais.

- **Transformada Z**

A Transformada Z é uma generalização da Transformada de Fourier de sinais discretos desenvolvido por De Moivre em 1730. Tal como a Transformada de Laplace, a Transformada Z permite que sistemas com função de transferência racional sejam caracterizados pelo seu mapa de pólos e zeros.

A Transformada Z desempenha para os sinais discretos o mesmo papel que a Transformada de Laplace para os sinais contínuos. Para uma seqüência $x(n)$, definido para todo n , a Transformada Z de $x(n)$ é exibida pela Equação 3.3, onde z é uma variável complexa e $X(z)$ é a função desta variável complexa.

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \quad (3.3)$$

A localização dos pólos e da região de convergência permite determinar características como a causalidade e a estabilidade. O mapa de pólos e zeros permite esboçar geometricamente a Transformada de Fourier à parte um fator de escala.

- **Transformada de Laplace**

Esta técnica está presente quase sempre em estudos da dinâmica e do controle de processos. É usada em sistemas lineares e permite converter equações diferenciais comuns em equações algébricas. A transformada de Laplace tem seu nome em homenagem ao matemático francês Pierre Simon Laplace (Figura 3.3).

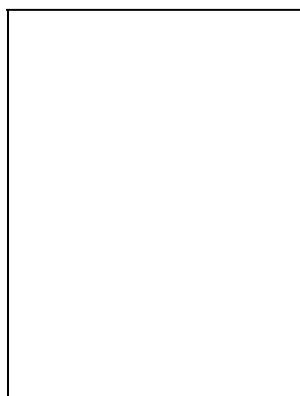


Figura 3.3 - Pierre Simon Laplace.

Em Matemática, e em particular na análise funcional, a transformada de Laplace de uma função $f(t)$ definida para todo número real $t \geq 0$ é a função $F(s)$, definida abaixo pela Equação 3.4.

$$F(s) = \mathcal{L}\{f(t)\} = \int_0^{\infty} e^{-s \cdot t} f(t) dt \quad (3.4)$$

As propriedades desta transformada a tornam útil para a análise de sistemas dinâmicos lineares. A vantagem mais interessante desta transformada é que a integração e a derivação tornam-se multiplicações e divisões, da mesma maneira que o logaritmo transforma a multiplicação em adição. Ela permite levar a resolução de equações diferenciais à resolução de equações polinomiais, que são muito mais simples de resolver.

- **Transformada de Wavelet**

As *wavelets* são funções matemáticas que surgiram em meados da década de 80 e que se destacaram em diversas áreas da ciência e da engenharia. A principal motivação atrás do desenvolvimento das *wavelets* foi à busca por um algoritmo rápido que representasse funções e conjuntos de dados de forma compacta.

As funções da Transformada Discreta de Wavelet (*Discrete Wavelet Transform – DWT*) são operações lineares em um vetor cujo tamanho é uma potência de dois, transformando-o em outro vetor, numericamente diferente e de mesmo tamanho.

Estudos na área de processamento de sinais destacaram as *wavelets* como uma transformada capaz de capturar a essência de um conjunto de dados com apenas um pequeno conjunto de coeficientes, de forma superior às tradicionais transformadas, o que contribui para a compactação dos áudios digitais. Existem algumas famílias de wavelets, citadas abaixo:

- **Haar** - É a primeira e a mais simples de todas. É descontínua e equivale a Daubechies citada a seguir;
- **Daubechies** - Compactly-supported orthonormal wavelets;
- **Biortogonal** - Apresenta a propriedade de fase linear, que é necessária na reconstrução de sinais e imagens. Utiliza duas wavelets, uma para decomposição e outra para reconstrução, o que gera propriedades interessantes;
- **Coiflets** - A função wavelet possui $2N$ momentos iguais a zero e a função escala tem $2N-1$ momentos iguais a zero;
- **Symlets** - São wavelets simétricas. Foi proposta como uma modificação da família Daubechies pela própria, possuindo características similares as desta família;
- **Morlet** - Não possui função escala e é explícita;
- **Mexican Hat** - Também não possui função escala, mas não é explícita;
- **Meyer** - A wavelet e a função de escala estão definidas no domínio de frequência.

- **Transformada de Mellin**

Em matemática, a Transformada de Mellin é uma transformada integral que pode ser vista como uma versão multiplicativa da Transformada de Laplace. Esta transformada integral extremamente conectada com as Séries de Dirichlet é freqüentemente usada na teoria numérica e a teoria de expansões assintóticas, sendo muito relacionada com a Transformada de Laplace e a Transformada de Fourier. A transformada de Mellin de uma função F é exibida na Equação 3.5 abaixo:

$$F(s) = M\{f(x)\} = \int_0^{\infty} x^{s-1} f(x) dx \quad (3.5)$$

- **Transformada de Hilbert**

Em processamento digital de sinais é freqüentemente necessário olhar as relações entre as partes reais e imaginárias de um sinal complexo. Essas relações são geralmente descritas pelas Transformadas de Hilbert. Esta metodologia é utilizada em análises complexas para gerar funções analíticas de valores complexos de funções reais. A transformada de Hilbert é usada para gerar funções o quais os componentes são conjugados harmônicos.

A transformada de Hilbert de um sinal $x(t)$ é definido a ser um sinal o quais componentes da freqüência são todas divididas em fases por $-\left(\pi / 2\right)$ radianos. A transformada de Hilbert de um sinal real é dada pela Equação 3.6 abaixo:

$$H[s(t)] = \frac{1}{\pi} \int \frac{s \tau}{t - \tau} dt \quad (3.6)$$

Com a transformada de Hilbert, pode-se calcular diretamente a fase acústica através da freqüência de resposta. Ou seja, dada uma freqüência de resposta de magnitude acústica de um sistema de caixas de auto-falante, por exemplo, pode-se transformar os dados em freqüência de resposta de fase acústica, através da Transformada de Hilbert.

Além dessas transformadas abordadas, há outras também muito utilizadas em processamento digital de sinais como, por exemplo, a Transformada K que é a revisão das transformadas de Fourier, Mellin e Hilbert o qual propiciou a fundação do projeto para o desenvolvimento desta nova transformada.

3.4.4 - Aplicações

A Transformada de Fourier possui muitas aplicações em disciplinas científicas. Como exemplo pode-se citar:

- Física;
- Teoria dos números;
- Análise combinatória;
- Processamento de sinais;
- Teoria da probabilidade;
- Estatística;
- Criptografia;
- Acústica;
- Oceanografia;
- Óptica;
- Geometria;

Nos campos relacionados com o processamento digital de sinal, a Transformada de Fourier é tipicamente utilizada para decompor sinais nos seus componentes em frequência e sua amplitudes. Atualmente existem diversas ferramentas matemáticas que provêm à transformação de uma função (sinal) do domínio do tempo para o domínio da frequência.

Em muitas aplicações é necessário conhecer não apenas quais as frequências que compõem um sinal, mas a localização desta frequência no domínio do tempo (exemplo: a frequência de 50 Hertz existe no intervalo de tempo de 100ms a 125 ms). Como exemplo, pode-se fazer uma analogia com o processamento de um radar onde a existência de determinada frequência detecta a presença de um objeto e a localização dessa frequência permite determinar a posição deste objeto.

3.4.5 - Algoritmo da Transformada de Fourier

A *Transformada de Fourier* é uma transformada integral que expressa uma função em termos de funções de base senoidal, como soma ou integral de funções senoidais multiplicadas por coeficientes (amplitudes). Existem várias variações diretamente relacionadas desta transformada, dependendo do tipo de função a transformar.

A Transformada de Fourier é normalmente apresentada em sua forma contínua. No entanto a sua implementação computacional é inviável já que os computadores só realizam cálculos discretos. Daí a importância do entendimento da Transformada Discreta de Fourier (DFT – Discrete Fourier Transform).

De acordo com Fourier, toda função real periódica de período 2π , pode ser escrita como uma soma infinita de senos e co-senos, conforme mostra a Equação 3.6 abaixo:

$$f(x) = a_0 + \sum_{k=1}^{\infty} (a_k \cos(k.x) + b_k \text{sen}(k.x)) \quad (3.6)$$

Onde os coeficientes a_0 , a_k e b_k são calculados segundo as equações a seguir:

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x) dx \quad (3.7)$$

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(k.x) dx \quad (3.8)$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \text{sen}(k.x) dx \quad (3.9)$$

A Tabela 3.1 abaixo apresenta a correspondência entre a função no domínio do tempo e a função correspondente no domínio da frequência, para vários tipos de função [BRI - 74]. A transformada de Fourier é construída a partir da série de Fourier dada pela Equação 3.6 e sua forma discreta é descrita conforme a Equação 3.10 abaixo:

$$F(\omega_p) = \frac{1}{2N} \sum_{p=0}^{2N-1} f(t_k) e^{i\omega_p t_k} = \frac{1}{2N} \sum_{p=0}^{2N-1} f(t_k) [\cos(\omega_p t_k) + i \text{sen}(\omega_p t_k)] \quad (3.10)$$

Esta transformada recebe um sinal de entrada no domínio do tempo, $f(t_k)$, e retorna um sinal no domínio da frequência, $F(\omega_p)$.

Domínio do Tempo	Domínio da Freqüência
Real	Parte real par Parte imaginária ímpar
Imaginária	Parte real ímpar Parte imaginária par
Parte real par Parte imaginária ímpar	Real
Parte real ímpar Parte imaginária par	Imaginária
Real e par	Real e par
Real e ímpar	Imaginária e ímpar
Imaginária e par	Imaginária e par
Imaginária e ímpar	Real e ímpar

Tabela 3.1 – Relações de domínios de tempos dos sinais

Para se realizar o cálculo da Transformada de Fourier de um sinal, é necessário saber qual será a sua *taxa de amostragem*. A taxa de amostragem indica quantas vezes num segundo se podem tirar amostras de voz ou música analógica. A taxa de amostragem exprime-se em hertz, isto é, número de amostras por segundo onde as mais comuns são 5.500, 11.025, 22.050 e 44.100 Hertz. Um método prático mostra que a taxa de amostragem tem de ter uma freqüência pelo menos duas vezes mais elevada que o som que se está a registrar. Por outras palavras, se quisermos converter um som numa representação digital utilizando toda a capacidade de audição humana (20 - 20.000 Hertz), as amostras do sinal têm de ser registradas a uma taxa igual à utilizada pelos leitores de CD - 44.100 Hertz.

A precisão da amostragem indica a quantidade de valores numéricos diferentes que uma amostra pode conter. Tanto a taxa como a precisão da amostragem afeta a qualidade do som e o número de bits necessário para os representar. Quanto mais elevada for a taxa e a precisão da amostragem, melhor será a qualidade do som. Contudo, um maior número de bits requer mais espaço de armazenagem e maior capacidade de transmissão de dados.

3.4.6 - Propriedades da Transformada de Fourier

Sabe-se então que há duas maneiras de representar uma mesma função ou sinal: uma representação no domínio do tempo ou do espaço e outra no domínio da frequência. A representação de um sinal no domínio do tempo está presente, naturalmente, no nosso dia a dia. Contudo, certas operações, principalmente na engenharia, tornam-se muito mais simples e esclarecedoras se trabalharmos no domínio da frequência, domínio este, conseguido através das Transformadas de Fourier. É muito importante observar o que ocorre em um domínio, quando efetuamos certas operações no outro domínio.

A seguir serão citadas apenas algumas propriedades da Transformada de Fourier. Desde já se deve enfatizar que as fórmulas e variáveis apresentadas a seguir não serão provadas matematicamente, mas somente citadas de acordo com a propriedade em questão, pois se tem como objetivo apenas mostrar algumas propriedades desta metodologia.

- **Separabilidade**

Esta propriedade nos mostra que o par de transformadas $\mathbf{F}(\mathbf{u}, \mathbf{v})$ e $\mathbf{f}(\mathbf{x}, \mathbf{y})$ pode ser obtido em dois passos separados. Em outras palavras, a função $\mathbf{F}(\mathbf{u}, \mathbf{v})$ é obtida pela transformação em cada linha de $\mathbf{f}(\mathbf{x}, \mathbf{y})$ e o resultado é multiplicado pelo número total das mesmas, \mathbf{M} , obtendo-se $\mathbf{F}(\mathbf{u}, \mathbf{v})$. $\mathbf{F}(\mathbf{u}, \mathbf{v})$ é obtida, agora, transformando-se $\mathbf{F}(\mathbf{x}, \mathbf{v})$ coluna por coluna. A Equação 3.11 abaixo mostra esta propriedade.

$$F(u, v) = \frac{1}{M} \sum_{x=0}^{M-1} e^{-j2\pi ux/M} \frac{1}{N} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi yv/N} \quad (3.11)$$

- **Translação**

Esta propriedade mostra, conforme exibido pela Equação 3.12, que a multiplicação de $\mathbf{f}(\mathbf{x}, \mathbf{y})$ pela exponencial $e^{j2\pi(u_0 x/M + v_0 y/N)}$ resulta num deslocamento na frequência para o ponto $(\mathbf{u}_0, \mathbf{v}_0)$. De maneira análoga, se multiplicarmos a transformada $\mathbf{F}(\mathbf{u}, \mathbf{v})$ pela mesma exponencial e tomarmos a transformada inversa, efetuamos um deslocamento espacial da origem para o ponto $(\mathbf{x}_0, \mathbf{y}_0)$.

$$\mathfrak{I}[f(x, y) e^{j2\pi(u_0 x/M + v_0 y/N)}] = F(u - u_0, v - v_0) \quad (3.12)$$

- **Periodicidade e Continuidade das Funções**

A periodicidade mostra que se $\mathbf{f}(\mathbf{x}, \mathbf{y})$ é periódica, somente um período é necessário para especificar completamente $\mathbf{F}(\mathbf{u}, \mathbf{v})$ no domínio da frequência. O mesmo se aplica a $\mathbf{f}(\mathbf{x}, \mathbf{y})$ no domínio espacial – Equação 3.13.

$$\mathfrak{T}[f(x + M, y + N)] = F(u + M, v + N) = F(u, v) \quad (3.13)$$

Verifica-se, através da Figura 3.4, a analogia entre *continuidade* e *periodicidade* que esta propriedade está relacionada com a natureza do sinal, da mesma maneira que o tipo de representação de Fourier está relacionado com a natureza do sinal. Esta propriedade permite se ter uma análise qualitativa da função, seja qual for o domínio, evitando assim um trabalho quantitativo exaustivo e, às vezes, desnecessário.

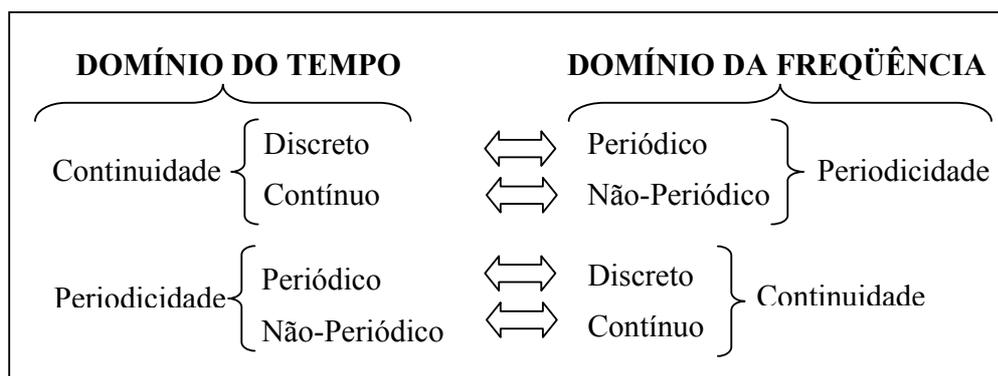


Figura 3.4 - Analogia entre Continuidade e Periodicidade.

- **Rotação**

Esta propriedade nos mostra que uma rotação em $\mathbf{f}(\mathbf{x}, \mathbf{y})$ por ângulo α , produz a mesma rotação em $\mathbf{F}(\mathbf{u}, \mathbf{v})$ e vice-versa – Equação 3.14.

$$\mathfrak{T}\{f[r \cos(\theta + \alpha)]\} = F[\omega \cos(\varphi + \alpha)] \quad (3.14)$$

- **Teorema da Convolução**

O teorema da convolução é, provavelmente, uma das ferramentas mais eficazes na análise em frequência. A importância da convolução no domínio da frequência consiste no fato que se $\mathbf{f}(\mathbf{x})$ tem a transformada de Fourier $\mathbf{F}(\mathbf{u})$ e $\mathbf{g}(\mathbf{x})$ tem sua transformada de Fourier $\mathbf{G}(\mathbf{u})$ então $\mathbf{f}(\mathbf{x}) * \mathbf{g}(\mathbf{x})$ tem $\mathbf{F}(\mathbf{u})\mathbf{G}(\mathbf{u})$ como transformada, ou seja $\mathbf{f}(\mathbf{x}) * \mathbf{g}(\mathbf{x}) \leftrightarrow \mathbf{F}(\mathbf{u})\mathbf{G}(\mathbf{u})$ como exibido na Equação 3.15 abaixo.

$$\mathfrak{T}[f(x) * g(x)] = \int_{-\infty}^{\infty} f(\alpha) e^{-j2\pi u \alpha} d\alpha G(u) = F(u)G(u) \quad (3.15)$$

- **Linearidade**

Esta propriedade é utilizada na solução de funções complexas, quando estas são fracionadas em sub-funções mais simples. Este método é conhecido como fatoração em frações parciais. Deve-se observar, entretanto, que no caso de funções periódicas, o pré-suposto básico é de que as sub-funções possuam o mesmo período fundamental. A técnica de fatoração de uma função no domínio da frequência é realizada, normalmente, com base na natureza das raízes do polinômio-denominador desta função.

3.4.7 - Exemplo Prático

Antes de aplicar na prática esta metodologia num computador, teve-se como interesse a resolução de um teste de mesa para comprovar matematicamente o funcionamento da Transformada Discreta de Fourier. A partir de um sinal $f(t_k)$ com apenas quatro amplitudes foram transformadas do domínio do tempo para o domínio da frequência, conforme ilustrado abaixo. Para este exemplo só foram capturadas quatro frequências (0, 1, 2 e 3) do sinal apenas como demonstração da veracidade da aplicação.

$$\text{Sinal} = \left[0, -1, 1, \frac{1}{2} \right] = f(t_k)$$

Para o cálculo dos coeficientes e utilizada a fórmula da Transformada Discreta de Fourier, conforme exibido pela Equação 3.16:

$$F(w_p) = \frac{1}{2N} \sum_{k=0}^3 f(t_k) * [\cos(w_p \cdot t_k) + i \cdot \text{sen}(w_p \cdot t_k)] \quad (3.16)$$

- Onde as frequências, tempos e o período são definidos a seguir:
 - A variável w_p é o cálculo da frequência no período (T) em que se quer analisar o sinal:

$$\blacksquare w_p = \frac{2 \cdot \pi \cdot p}{T}$$

- A variável t_k é o tempo que se pega do sinal de entrada:

$$\blacksquare t_k = \frac{K \cdot T}{2 \cdot N}$$

- A variável T é a determinação da taxa de amostragem:

$$\blacksquare T = 2\pi$$

- Calculando as frequências e os tempos do sinal de acordo com a tabela 3.2:

Frequências		Tempos	
$w_0 = \frac{2 \cdot \pi \cdot 0}{2\pi} = 0$	$w_2 = \frac{2 \cdot \pi \cdot 2}{2\pi} = 2$	$t_0 = \frac{0 \cdot 2\pi}{2 \cdot 2} = 0$	$t_2 = \frac{2 \cdot 2\pi}{2 \cdot 2} = \pi$
$w_1 = \frac{2 \cdot \pi \cdot 1}{2\pi} = 1$	$w_3 = \frac{2 \cdot \pi \cdot 3}{2\pi} = 3$	$t_1 = \frac{1 \cdot 2\pi}{2 \cdot 2} = \frac{\pi}{2}$	$t_3 = \frac{3 \cdot 2\pi}{2 \cdot 2} = \frac{3\pi}{2}$

Tabela 3.2 – Frequências e tempos do sinal a ser analisado.

- A seguir são apresentados todos os cálculos para a obtenção dos coeficientes:

$$F(w_0) = \frac{1}{2 \cdot N} \cdot \left\{ \begin{array}{l} f(t_0) \cdot [\cos(w_0 \cdot t_0) + i \cdot \text{sen}(w_0 \cdot t_0)] + f(t_1) \cdot [\cos(w_0 \cdot t_1) + i \cdot \text{sen}(w_0 \cdot t_1)] \\ + \\ f(t_2) \cdot [\cos(w_0 \cdot t_2) + i \cdot \text{sen}(w_0 \cdot t_2)] + f(t_3) \cdot [\cos(w_0 \cdot t_3) + i \cdot \text{sen}(w_0 \cdot t_3)] \end{array} \right\} = \dots = \frac{1}{8} = 0,125$$

$$F(w_1) = \frac{1}{2 \cdot N} \cdot \left\{ \begin{array}{l} f(t_0) \cdot [\cos(w_1 \cdot t_0) + i \cdot \text{sen}(w_1 \cdot t_0)] + f(t_1) \cdot [\cos(w_1 \cdot t_1) + i \cdot \text{sen}(w_1 \cdot t_1)] \\ + \\ f(t_2) \cdot [\cos(w_1 \cdot t_2) + i \cdot \text{sen}(w_1 \cdot t_2)] + f(t_3) \cdot [\cos(w_1 \cdot t_3) + i \cdot \text{sen}(w_1 \cdot t_3)] \end{array} \right\} = \dots = -\frac{1}{4} - \frac{3i}{8} = -0,25 - 0,375i$$

$$F(w_2) = \frac{1}{2 \cdot N} \cdot \left\{ \begin{array}{l} f(t_0) \cdot [\cos(w_2 \cdot t_0) + i \cdot \text{sen}(w_2 \cdot t_0)] + f(t_1) \cdot [\cos(w_2 \cdot t_1) + i \cdot \text{sen}(w_2 \cdot t_1)] \\ + \\ f(t_2) \cdot [\cos(w_2 \cdot t_2) + i \cdot \text{sen}(w_2 \cdot t_2)] + f(t_3) \cdot [\cos(w_2 \cdot t_3) + i \cdot \text{sen}(w_2 \cdot t_3)] \end{array} \right\} = \dots = \frac{3}{8} = 0,375$$

$$F(w_3) = \frac{1}{2 \cdot N} \cdot \left\{ \begin{array}{l} f(t_0) \cdot [\cos(w_3 \cdot t_0) + i \cdot \text{sen}(w_3 \cdot t_0)] + f(t_1) \cdot [\cos(w_3 \cdot t_1) + i \cdot \text{sen}(w_3 \cdot t_1)] \\ + \\ f(t_2) \cdot [\cos(w_3 \cdot t_2) + i \cdot \text{sen}(w_3 \cdot t_2)] + f(t_3) \cdot [\cos(w_3 \cdot t_3) + i \cdot \text{sen}(w_3 \cdot t_3)] \end{array} \right\} = \dots = -\frac{1}{4} + \frac{3i}{8} = 0,25 + 0,375i$$

A seguir são calculados os espectros de potência para os quatro coeficientes calculados:

- Frequência 0: $\frac{1}{8} \rightarrow \sqrt{\left(\frac{1}{8}\right)^2 + 0} = 0,125$

- Frequência 1: $-\frac{1}{4} - \frac{3i}{8} \rightarrow \sqrt{\left(-\frac{1}{4}\right)^2 + \left(-\frac{3}{8}\right)^2} = 0,450$

- Frequência 2: $\frac{3}{8} \rightarrow \sqrt{\left(\frac{3}{8}\right)^2} = 0,375$

- Freqüência 3: $-\frac{1}{4} + \frac{3i}{8} \rightarrow \sqrt{\left(\frac{-1}{4}\right)^2 + \left(\frac{3}{8}\right)^2} = 0,450$

Para a verificação da confiabilidade do sinal obtido no domínio da freqüência torna-se necessário à realização do processo inverso para a obtenção do sinal original novamente (domínio do tempo). Para isso é utilizado a Transformada Discreta Inversa de Fourier (IDFT – Inverse Discrete Fourier Transform), conforme ilustrado pela Equação 3.17 abaixo:

$$F(w_p) = \sum_{p=0}^3 f(t_k) * [\cos(w_p \cdot t_k) - i \cdot \text{sen}(w_p \cdot t_k)] \quad (3.17)$$

- A seguir são apresentados todos os cálculos para o retorno ao sinal original:

$$F(t_0) = \left\{ \begin{array}{l} f(w_0) \cdot [\cos(w_0 \cdot t_0) - i \cdot \text{sen}(w_0 \cdot t_0)] + f(w_1) \cdot [\cos(w_1 \cdot t_0) - i \cdot \text{sen}(w_1 \cdot t_0)] \\ + \\ f(w_2) \cdot [\cos(w_2 \cdot t_0) - i \cdot \text{sen}(w_2 \cdot t_0)] + f(w_3) \cdot [\cos(w_3 \cdot t_0) - i \cdot \text{sen}(w_3 \cdot t_0)] \end{array} \right\} = \dots = 0$$

$$F(t_1) = \left\{ \begin{array}{l} f(w_0) \cdot [\cos(w_0 \cdot t_1) - i \cdot \text{sen}(w_0 \cdot t_1)] + f(w_1) \cdot [\cos(w_1 \cdot t_1) - i \cdot \text{sen}(w_1 \cdot t_1)] \\ + \\ f(w_2) \cdot [\cos(w_2 \cdot t_1) - i \cdot \text{sen}(w_2 \cdot t_1)] + f(w_3) \cdot [\cos(w_3 \cdot t_1) - i \cdot \text{sen}(w_3 \cdot t_1)] \end{array} \right\} = \dots = -1$$

$$F(t_2) = \left\{ \begin{array}{l} f(w_0) \cdot [\cos(w_0 \cdot t_2) - i \cdot \text{sen}(w_0 \cdot t_2)] + f(w_1) \cdot [\cos(w_1 \cdot t_2) - i \cdot \text{sen}(w_1 \cdot t_2)] \\ + \\ f(w_2) \cdot [\cos(w_2 \cdot t_2) - i \cdot \text{sen}(w_2 \cdot t_2)] + f(w_3) \cdot [\cos(w_3 \cdot t_2) - i \cdot \text{sen}(w_3 \cdot t_2)] \end{array} \right\} = \dots = 1$$

$$F(t_3) = \left\{ \begin{array}{l} f(w_0) \cdot [\cos(w_0 \cdot t_3) - i \cdot \text{sen}(w_0 \cdot t_3)] + f(w_1) \cdot [\cos(w_1 \cdot t_3) - i \cdot \text{sen}(w_1 \cdot t_3)] \\ + \\ f(w_2) \cdot [\cos(w_2 \cdot t_3) - i \cdot \text{sen}(w_2 \cdot t_3)] + f(w_3) \cdot [\cos(w_3 \cdot t_3) - i \cdot \text{sen}(w_3 \cdot t_3)] \end{array} \right\} = \dots = \frac{1}{2}$$

Verificou-se então que este processamento funciona realmente, já que dado um sinal qualquer no domínio do tempo, foi possível calcular os coeficientes do sinal (DFT), o seu espectro de potência, e o retorno para o sinal original através da IDFT. A tabela 3.3 mostra as etapas desta aplicação.

Processos da Transformada Discreta de Fourier				
Sinal Original	0	-1	1	$\frac{1}{2}$
Cálculo da DFT	$\frac{1}{8}$	$-\frac{1}{4} - \frac{3i}{8}$	$\frac{3}{8}$	$-\frac{1}{4} + \frac{3i}{8}$
Espectro de Potência	0,125	0,450	0,375	0,450
Cálculo da IDFT	0	-1	1	$\frac{1}{2}$

Tabela 3.3 – Etapas da Transformada Discreta de Fourier

3.5 - Implementação da Transformada de Fourier

Para colocar em prática todos os estudos relacionados a técnicas de Fourier, foi realizada uma implementação da Transformada Discreta de Fourier e nela utilizadas algumas aplicações existentes na literatura. Para isso, foi utilizando a ferramenta **MATLAB 7.0** como mostra a Figura 3.5 abaixo. Foi escolhido este aplicativo, pois o MATLAB (MATrix LABoratory) é um software interativo de alta performance voltado para o cálculo numérico e científico. O MATLAB integra análise numérica, cálculo com matrizes, processamento de sinais e construção de gráficos em ambiente fácil de usar onde problemas e soluções podem ser expressos como eles são escritos na matemática ou na forma de uma linguagem de programação.

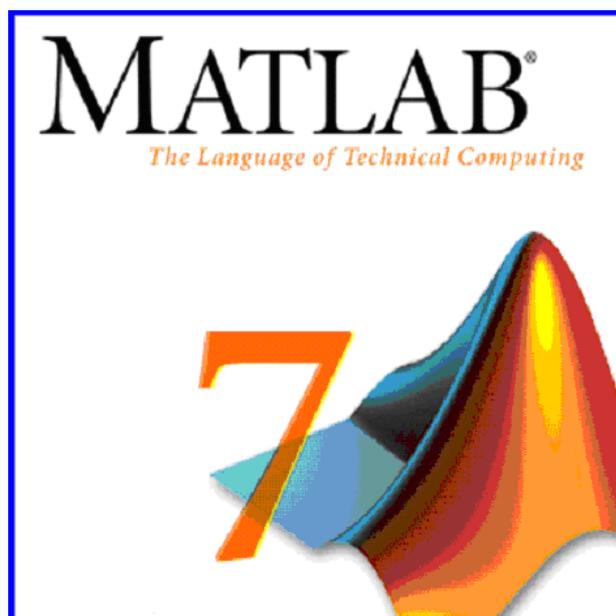


Figura 3.5 – Software MATLAB.

A implementação da Transformada Discreta de Fourier abaixo mostra todo o processo de cálculos dos coeficientes do sinal, seu espectro de potência, o cálculo da Transformada Inversa de Fourier bem como a comparação desta com o sinal original. Pode-se perceber que esta implementação realizada no MATLAB comparada com outras arquiteturas é muito mais simples, pois o elemento básico de informação é uma matriz que não requer dimensionamento permitindo a resolução de muitos problemas numéricos em apenas uma fração do tempo que se gastaria para escrever um programa semelhante em linguagem Fortran, Basic ou C.

Com base neste software, foi produzido um algoritmo no que realizasse a Transformada Discreta de Fourier que é ilustrada a seguir. Neste algoritmo são executados procedimentos divididos por etapas conforme a tabela 3.4 abaixo.

Algoritmo da Transformada Discreta de Fourier	
Etapas	Procedimento
I	Leitura do sinal que se deseja processar
II	Configuração das frequências que serão obtidas a partir do sinal
III	Cálculo da Transformada Discreta de Fourier
IV	Cálculo do espectro de potência
V	Cálculo da Inversa da Transformada Discreta de Fourier
VI	Verificando a correlação entre o sinal original e sua inversa

Tabela 3.4 – Etapas do algoritmo da DFT.

ALGORITMO DA TRANSFORMADA DISCRETA DE FOURIER

I – Variável que irá receber o sinal que será analisado:

x = sinal que será processado.

II - Configurando os dados para o cálculo a partir do sinal (**x**):

N = size(x,2);

w= freqüências que serão capturadas do sinal;

M = size(w,2);

soma=0;

cont2=1:N;

tempo= ((cont2-1)*2*pi)/N;

III - Cálculo da DFT do sinal:

for cont=1:M

coeficientes(cont) = sum(x.*(cos(w(cont)*tempo) + i*sin(w(cont)*tempo)))/N;

end

IV - Cálculo do espectro do sinal:

espectro = sqrt(real(fw).^2 + imag(fw).^2);

V - Cálculo da IDFT:

for cont=1:N

inversa(cont) = sum(fw.*(cos(w*tempo(cont)) - i*sin(w*tempo(cont))));

end

VI - Verificando a correlação da inversa do sinal com o sinal original:

corrcoef(x,inversa)

3.6 - Aplicações da implementação da DFT

Como aplicações para teste teve-se como interesse abordar alguns tipos de sinais para verificar se a implementação da Transformada de Fourier no MatLab obtém os resultados desejados no domínio da freqüência.

3.6.1 - Primeiro Procedimento

Neste procedimento teve-se como interesse utilizar o sinal do exemplo prático do tópico 3.4.7 para verificar se esta implementação da Transformada de Fourier obtém os mesmos resultados. Dado a implementação do tópico 3.5, algumas etapas para esta aplicação são alteradas como ilustradas a seguir.

- **Etapa I** – Variável que irá receber o sinal que será analisado:
 - $x = [0 \ 1 \ -1 \ 0.5]$;
- **Etapa II** - Configurando as frequências que serão adquiridas:
 - $w = 0:3$;

A seguir são exibidos os cálculos dos coeficientes através da variável chamada “coeficientes” da etapa III:

- $F(w_0) = \frac{1}{8} = 0,125$
- $F(w_1) = -\frac{1}{4} - \frac{3i}{8} = -0,25 - 0,375i$
- $F(w_2) = \frac{3}{8} = 0,375$
- $F(w_3) = -\frac{1}{4} + \frac{3i}{8} = 0,25 + 0,375i$

A seguir são calculados os espectros de potência para os quatro coeficientes calculados através da variável chamada “espectro” da etapa IV, onde o resultado pode ser exibido pela Figura 3.6:

- $F(w_0) = 0,125 = \dots = 0,125$
- $F(w_1) = -0,25 - 0,375i = \dots = 0,450$
- $F(w_2) = 0,375 = \dots = 0,375$
- $F(w_3) = 0,25 + 0,375i = \dots = 0,450$

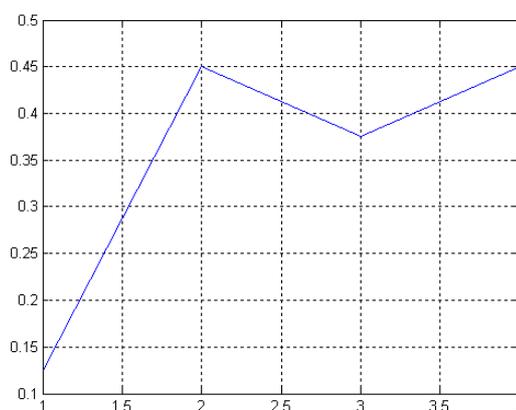


Figura 3.6 – Espectro de potência

A seguir foi realizado o cálculo da Inversa da Transformada Discreta de Fourier. Através da variável “inversa” da implementação da etapa V, o sinal é retornado do domínio da frequência para o domínio do tempo conforme ilustrado pela Tabela 3.5. Pela etapa VI da implementação da DFT, foi comparado o sinal original com a sua transformada inversa, obtendo uma correlação de **100%** de acerto.

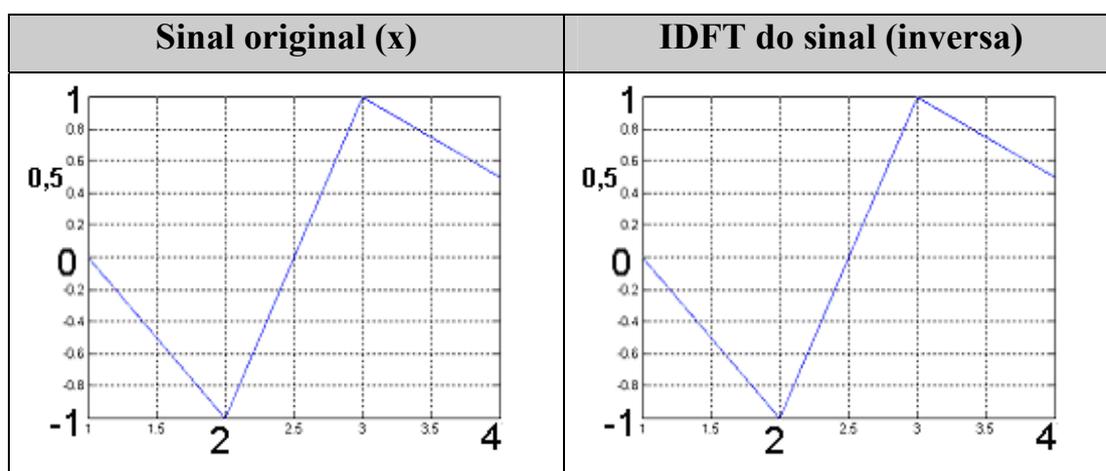


Tabela 3.5 Comparação entre o sinal original e sua inversa

3.6.2 - Segundo Procedimento

Para o próximo teste da implementação da DFT foi utilizado um sinal de voz de 2 segundos (88.200×1), conforme ilustrado na Figura 3.7.

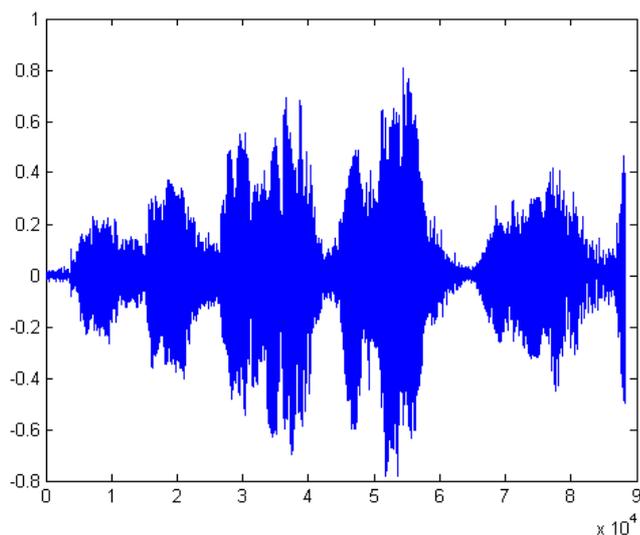


Figura 3.7 - Sinal de voz de 2 segundos.

Dado a implementação da DFT do tópico 3.5, segue abaixo algumas etapas que serão alteradas como ilustradas a seguir.

- **Etapa I** – Variável que irá receber o sinal de voz que será analisado:
 - `x = wavread('sinal_2_segundos');`
- **Etapa II** - Configurando as frequências que serão adquiridas:
 - `w = 0:0.5:22050;`

Após o cálculo dos coeficientes do sinal, foi produzido o espectro de potência da voz exibido pela Figura 3.8.

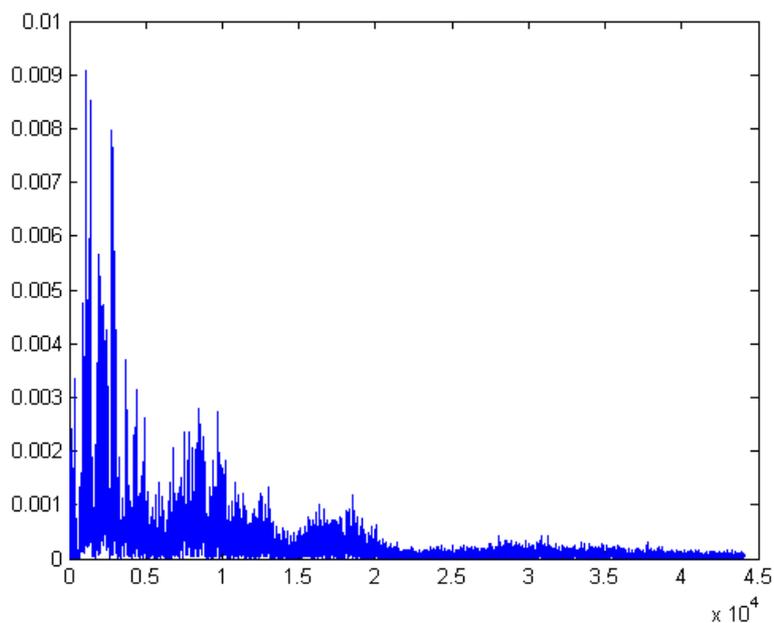


Figura 3.8 – Espectro de potência do sinal de voz de 2 segundos.

Para verificação da confiabilidade do espectro de potência do sinal de voz, foi calculada a Transformada Discreta Inversa de Fourier e comparado com o sinal original (Tabela 3.6), obtendo uma correlação de **100%**.

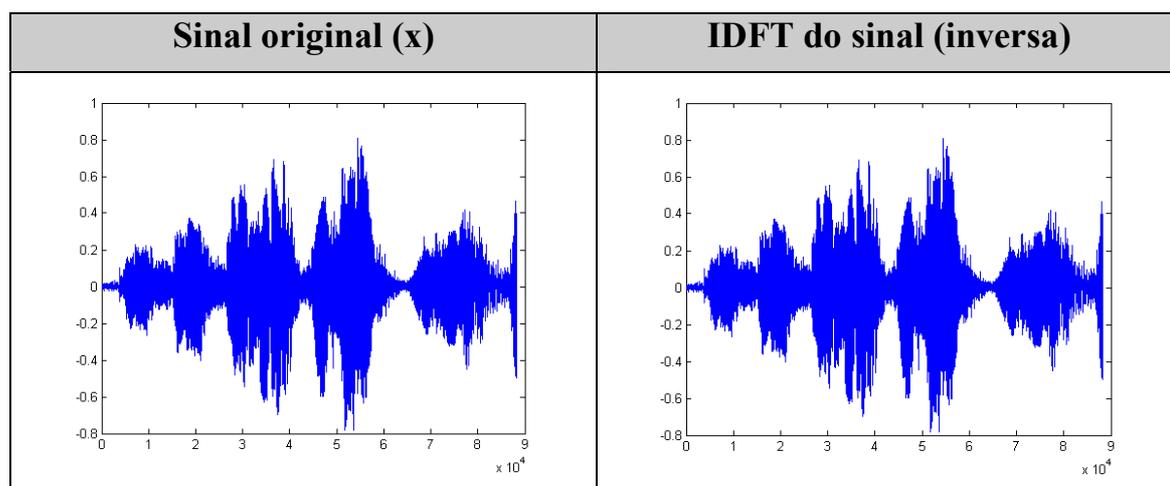


Tabela 3.6 Correlação entre a voz de 2 segundo e sua inversa.

Com os resultados mostrados pela Tabela 3.6, está claro que a transformação do sinal do domínio do tempo para o domínio da frequência é confiável já que o retorno desta para tempo contínuo mostra sua confiabilidade nas informações contidas no sinal.

Com esta aplicação torna-se possível e principalmente confiável o desenvolvimento do protótipo computacional proposto neste projeto que é a utilização das vozes pré-processadas pela DFT (espectro de potência das vozes) que serão utilizadas no treinamento de Redes Neurais Artificiais para o reconhecimento de padrões emocionais na voz.

4 - RNAS NO RECONHECIMENTO DE STRESS NA VOZ

4.1 - Aplicação

O objetivo desta aplicação é verificar se, com o pré-processamento da voz humana através da aplicação da Transformada Discreta de Fourier, o treinamento realizado pela Rede Neural Artificial se torna mais otimizado, possibilitando um aprendizado maior pela rede comparado com um treinamento sem este pré-processamento.

Em Matuck (2004) foi produzido um banco de dados contendo vozes de 14 pessoas, todas geradas com a leitura de um texto pré-determinado. As vozes foram capturadas através de um gravador de mão portátil, o qual foram observados vários fatores externos que estavam presentes junto à voz como chiados, ruídos, entre outros. Parte deste banco de dados (10 vozes calmas e nervosas) foi treinado pela rede Perceptron de Múltiplas Camadas e a segunda parte (4 vozes calmas e nervosas) foi utilizada no processo de generalização da rede o qual obteve um reconhecimento de **50,2%** das vozes nunca vistas anteriormente pela rede neural.

Para este projeto teve-se como interesse melhorar o modo de captura das vozes calmas e nervosas bem como a melhorar a capacidade de aprendizado pela RNAs. Para tal, foi formulado um banco de dados contendo falas de diversas pessoas em estado de estresse e normal. Para aumentar a qualidade do banco de dados das vozes, foram capturadas vozes da televisão, onde o único som emitido por esta era a voz de uma pessoa (estressada ou calma).

Todas as vozes foram gravadas da televisão para o computador no formato “WAV” utilizando, para isto, o software Sound Forge 4.5 (Figura 4.1) considerado um dos melhores editores de áudio existentes no mercado atualmente. Neste programa foi realizada a edição de todas as vozes para utilização no reconhecimento pelas RNAs, onde o tempo de cada voz foi estipulado para dez segundos (10s).

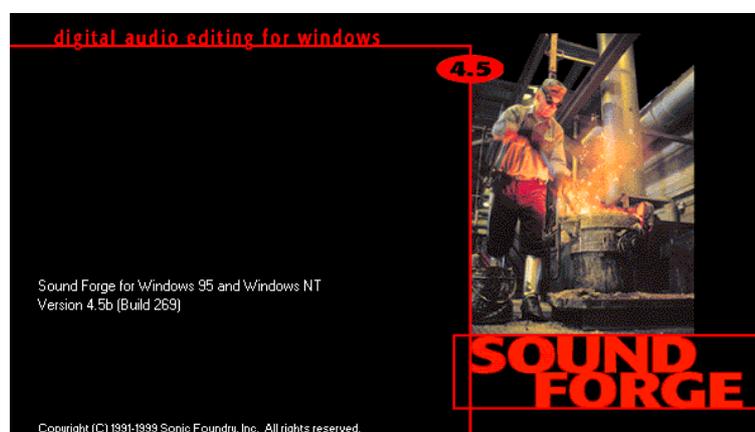


Figura 4.1 – Software Sound Forge.

No total, foram capturadas 35 vozes calmas e 35 vozes nervosas, retiradas de diversos programas como, novelas, filmes, jornais de notícias, programas esportivos, entre outros. Abaixo são ilustradas duas vozes de cada tipo.

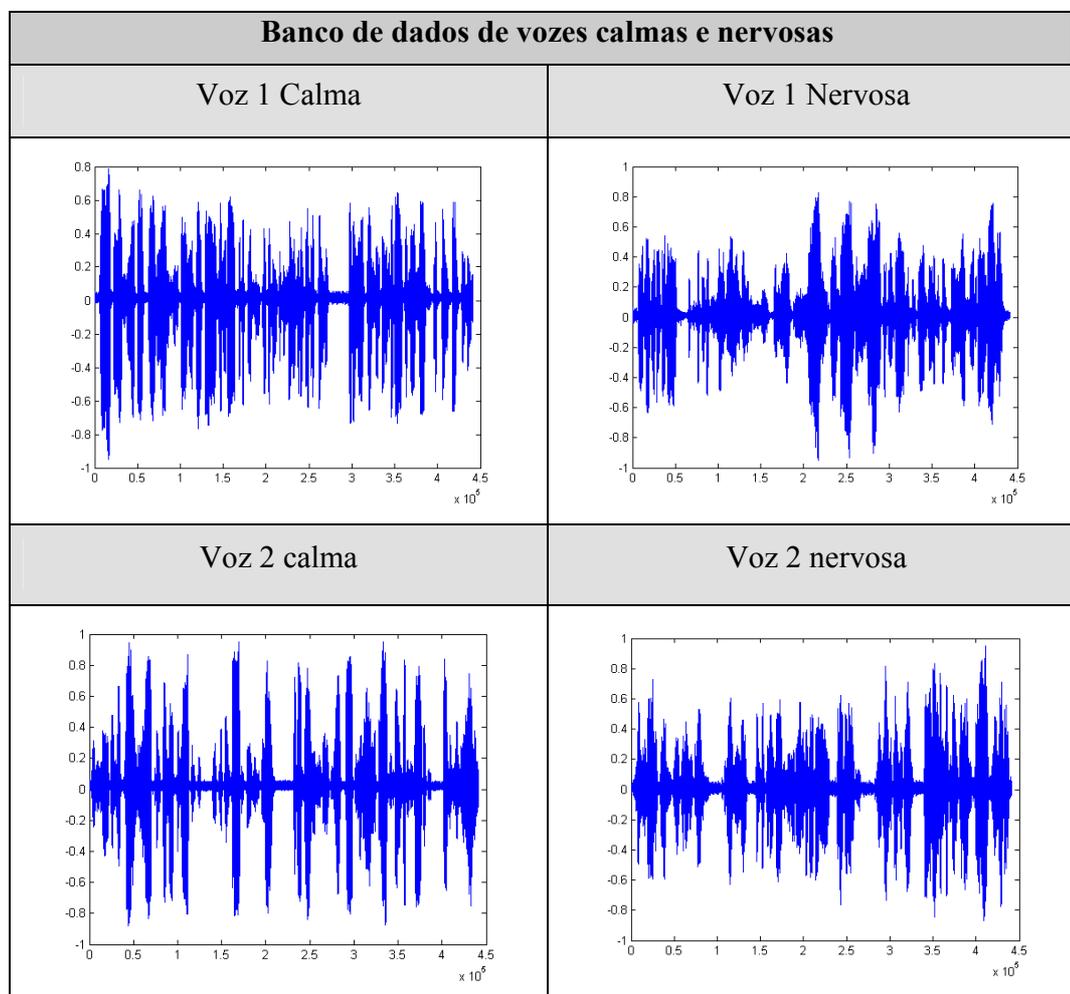


Figura 4.2 - Ilustração das vozes gravadas.

Este banco de dados de voz foram bases para a aplicação em dois procedimentos específicos em Redes Neurais Artificiais envolvendo a rede *Perceptron de Múltiplas Camadas* (MLP – Multilayer Perceptron). A primeira aplicação tinha como objetivo o treinamento das vozes pelas RNAs sem o pré-processamento destas pela Transformada Discreta de Fourier, ou seja, no seu estado original de como foram capturadas e digitalizadas.

Já o segundo procedimento consistiu em pré-processar as vozes pela Transformada Discreta de Fourier, transformando-as para o domínio da frequência. Em seguida, estas vozes foram também utilizadas no treinamento das RNAs pela rede Perceptron de Múltiplas Camadas. Após os dois procedimentos foi realizada uma comparação entre as

duas aplicações para identificar qual a técnica é a mais qualificada para o problema em questão.

4.2 - Treinamento das RNAs sem o pré-processamento da DFT

4.2.1 - Aplicação

As trinta e seis vozes gravadas foram divididas e utilizadas no treinamento e generalização da rede Perceptron de Múltiplas Camadas (Back-Propagation) conforme mostra a tabela 4.1.

Vozes	Diagnósticos	Treinamento	Generalização
1 a 25	Calmas e Nervosas	X	
26 a 36	Calmas e Nervosas		X

Tabela 4.1 - Disposição das vozes utilizadas na rede MLP

Como a rede Perceptron de Múltiplas Camadas possui um aprendizado supervisionado, há a necessidade de um professor externo auxiliando o seu treinamento, o que chamamos de saída desejada para cada padrão que é fornecido a rede. A saída desejada para essas vozes pode ser vista codificada na tabela 4.2 onde a voz só pode ser diagnosticada como “calma” ou “nervosa”.

Diagnóstico	Codificação
Voz calma	0 1
Voz nervosa	1 0

Tabela 4.2 – Codificação dos diagnósticos de voz

Como cada vetor de voz possui uma dimensão muito grande (441000 x 1), este foi dividido em blocos de mil dados (441 blocos com o tamanho: 1000 x 1) e assim colocados seqüencialmente cada bloco na camada de entrada da rede MLP. A Figura 4.3 ilustra esse processo.

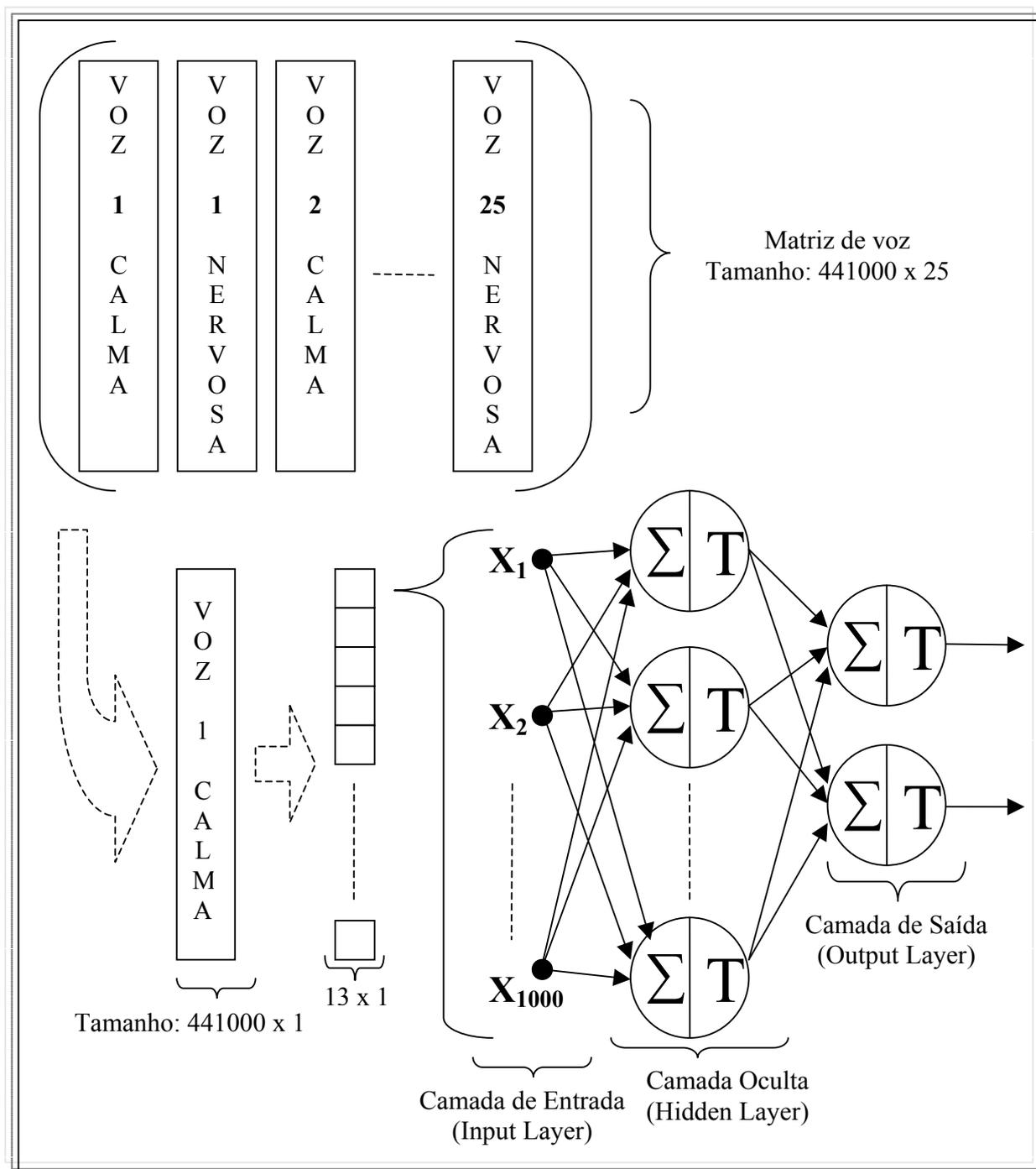


Figura 4.3 – Ilustração das etapas utilizadas no treinamento da rede MLP sem o pré-processamento da DFT

Para o treinamento da rede foi utilizada uma taxa de aprendizado (velocidade com que a rede irá aprender) variada, onde a cada época (período em que todos os exemplos de voz são apresentados para a rede) a taxa era adaptada de acordo com o erro encontrado na saída da rede. Se o erro da época atual é menor que o erro da época anterior, a taxa de aprendizado é aumentada, caso contrário é diminuído.

A função de ativação utilizada na rede foi a função logística sigmoidal, com apenas uma camada oculta (hidden layer) com **100** neurônios, ou seja, o número de neurônios é equivalente a apenas **10%** da camada de entrada da rede, que possui **1000** entradas. A rede MLP foi processada durante **13.754** épocas, ou seja, as **50** vezes (entre calmas e nervosas) foram apresentadas **13.754** vezes com o objetivo de diminuir o erro quadrático médio da rede adaptando os pesos sinápticos da melhor maneira possível.

Neste treinamento o menor erro encontrado pela rede foi de **26.8855**, ou seja, este erro é definido pela soma dos erros de todas as entradas da rede que é o produto das **50** vezes por **441.000** (número de entrada de cada voz) dando um total de **22.050.000** dados. A Figura 4,4 mostra o erro encontrado na rede em função do número de épocas

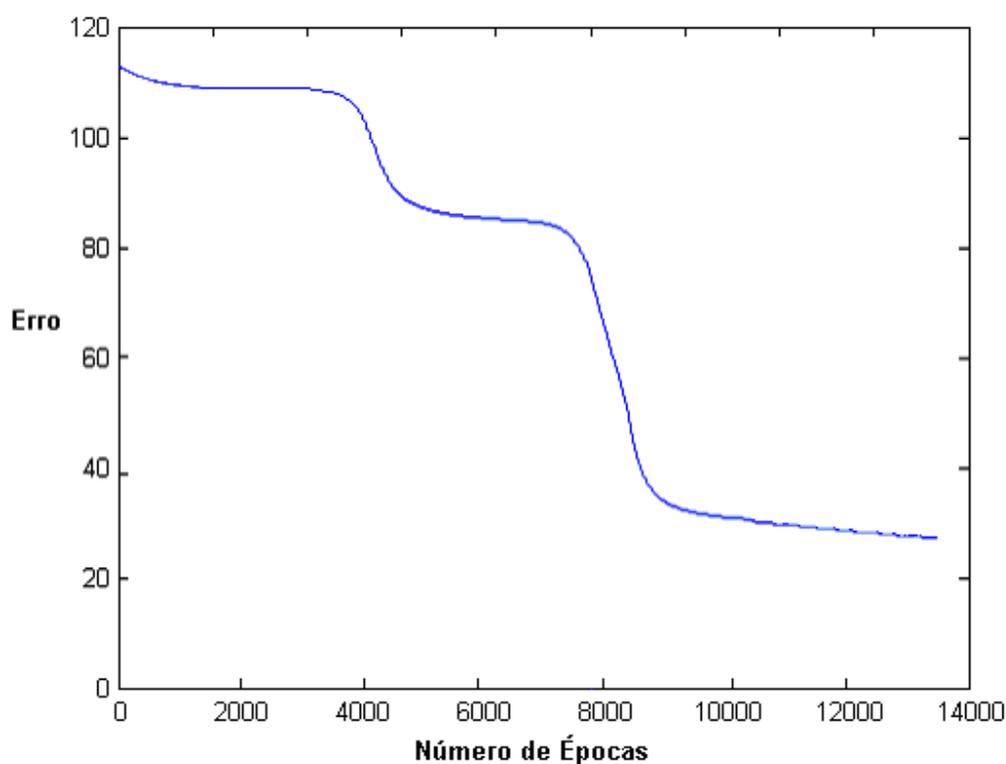


Figura 4.4 - Erro encontrado no treinamento da rede MLP.

O próximo passo foi fazer a generalização das dez vezes que foram utilizadas no treinamento, para saber se a rede Perceptron de Múltiplas Camadas aprendeu os padrões o qual treinou. Na ativação da rede com as vozes que foram treinadas, observou um acerto de **96,15%**. Cada voz utilizada no treinamento conseguiu determinada taxa de acerto conforme mostra a tabela 4.3.

Voz	Diagnóstico	Acerto	Voz	Diagnóstico	Acerto
1	Calmo	95.01 %	14	Calmo	90.48 %
	Nervoso	20.41 %		Nervoso	70.29 %
2	Calmo	95.92 %	15	Calmo	96.60 %
	Nervoso	27.21 %		Nervoso	94.78 %
3	Calmo	95.01 %	16	Calmo	74.38 %
	Nervoso	35.37 %		Nervoso	96.37 %
4	Calmo	85.94 %	17	Calmo	92.97 %
	Nervoso	19.05 %		Nervoso	50.57 %
5	Calmo	97.73 %	18	Calmo	95.46 %
	Nervoso	90.48 %		Nervoso	28.12 %
6	Calmo	90.02 %	19	Calmo	90.25 %
	Nervoso	16.55 %		Nervoso	56.92 %
7	Calmo	63.95 %	20	Calmo	90.48 %
	Nervoso	88.89 %		Nervoso	93.88 %
8	Calmo	90.70 %	21	Calmo	92.74 %
	Nervoso	88.21 %		Nervoso	20.41 %
9	Calmo	92.97 %	22	Calmo	96.37 %
	Nervoso	92.52 %		Nervoso	94.56 %
10	Calmo	91.16 %	23	Calmo	96.60 %
	Nervoso	93.20 %		Nervoso	93.65 %
11	Calmo	91.16 %	24	Calmo	93.88 %
	Nervoso	36.73 %		Nervoso	60.32 %
12	Calmo	92.97 %	25	Calmo	96.37 %
	Nervoso	63.95 %		Nervoso	24.49 %
13	Calmo	93.20 %			
	Nervoso	75.06 %			

Tabela 4.3 – Porcentagem de acerto das vozes *treinadas* na rede MLP

Num segundo experimento, foram mostradas à rede as 22 vozes (entre calmas e nervosas) que não foram utilizadas no treinamento, com o intuito de verificar a porcentagem de acerto com padrões de vozes nunca vistos anteriormente pela rede MLP. Verificou-se que na ativação a rede conseguiu obter um acerto de **56,56%** com um erro

quadrático médio encontrado na generalização de **18.365**. A porcentagem de acerto dos diagnósticos de cada voz pode ser visto na tabela 4.4 abaixo.

Voz	Diagnóstico	Acerto	Voz	Diagnóstico	Acerto
1	Calmo	86.39 %	7	Calmo	56.91 %
	Nervoso	83.90 %		Nervoso	69.84 %
2	Calmo	89.57 %	8	Calmo	84.35 %
	Nervoso	78.91 %		Nervoso	49.89 %
3	Calmo	85.71 %	9	Calmo	72.79 %
	Nervoso	91.16 %		Nervoso	51.32 %
4	Calmo	89.34 %	10	Calmo	84.13 %
	Nervoso	92.97 %		Nervoso	41.50 %
5	Calmo	83.90 %	11	Calmo	87,98 %
	Nervoso	95.01 %		Nervoso	75,96 %
6	Calmo	78.00 %			
	Nervoso	36.96 %			

Tabela 4.4 – Porcentagem de acerto das vozes *generalizadas* na rede MLP

4.2.2 - Conclusão

Através desta aplicação percebe-se que com a formulação de um banco de vozes de maior qualidade quando comparadas com os resultados de Matuck (2004), a capacidade de aprendizado dos padrões de voz pela rede neural é mais eficiente. O reconhecimento das vozes calmas e nervosas tanto no treinamento como na generalização pode ser vista na Tabela 4.5 abaixo.

Rede Perceptron de Múltiplas Camadas - MLP			
Treinamento da rede		Generalização da rede	
Vozes calmas	91,29 %	Vozes calmas	80,13 %
Vozes nervosas	61,28 %	Vozes nervosas	32,98 %

Tabela 4.5 – Comparação entre vozes calmas e de estresse.

Torna-se necessário ressaltar algumas desvantagens percebidas durante o processo de treinamento da rede neural nesta aplicação. O tempo demandado para o treinamento de cada época foi estipulado em torno de **8** minutos, o que necessitou o treinamento da rede durante um período de quatro meses direto (24 horas de processamento por dia) para chegar na época **13.754**. O ideal para este procedimento seria deixar a rede treinando por um período de tempo maior para que a mesma pudesse aprender com uma maior eficiência os padrões de voz, o que não foi possível devido ao tempo destinado para este projeto.

Outro fator foi à memória do computador necessária para a realização deste treinamento no Matlab que tinha de ser maior do que **256** megabytes de RAM. Pois o computador não aceitou uma rede com **100** neurônios na camada interna (hidden layer) e com **441** blocos de mil dados na camada de entrada cada bloco, tornando-se necessário um computador com no mínimo **512** megabytes de RAM para realizar este procedimento.

4.3 - *Treinamento das RNAs com o pré-processamento da DFT*

4.3.1 - Pré-processamento da DFT nas vozes

Antes do treinamento das vozes pelas RNAs, todas as vozes foram transformadas do domínio do tempo para o domínio da frequência através da Transformada Discreta de Fourier (algoritmo de implementação do tópico 3.5). Como cada voz tem a duração de dez segundos com **441.000** pontos, tem-se que a taxa de amostragem para este sinal é:

- **Taxa amostral do sinal de voz** $\rightarrow \frac{Sinal}{tempo} = \frac{441.000}{10} = 44.100$

Em seguida foram determinados quais as frequências que serão capturadas do sinal que é exatamente a metade da taxa de amostragem desta voz, ou seja, irá da frequência zero até a frequência **22050**. A seguir são ilustradas algumas etapas da implementação da DFT do tópico 3.5 que são configuradas para este processo.

- **Etapa I** – Variável que irá receber o sinal que será analisado – A voz 1 calma, por exemplo:
 - `x = wavread("voz_01_calma.wav");`
- **Etapa II** - Configurando as frequências que serão adquiridas – Neste caso irá da frequência 0 até a 22050, indo de 0,5 em 0,5:
 - `w = 0:0,5:22050;`

Para exemplificar este procedimento, são ilustradas duas vozes (uma calma e uma nervosa – tabela 4.6) no seu estado original e no domínio da frequência após realizado o cálculo da Transformada Discreta de Fourier (etapa III da implementação do tópico 3.5) e do seu espectro de potência (etapa IV da implementação do tópico 3.5).

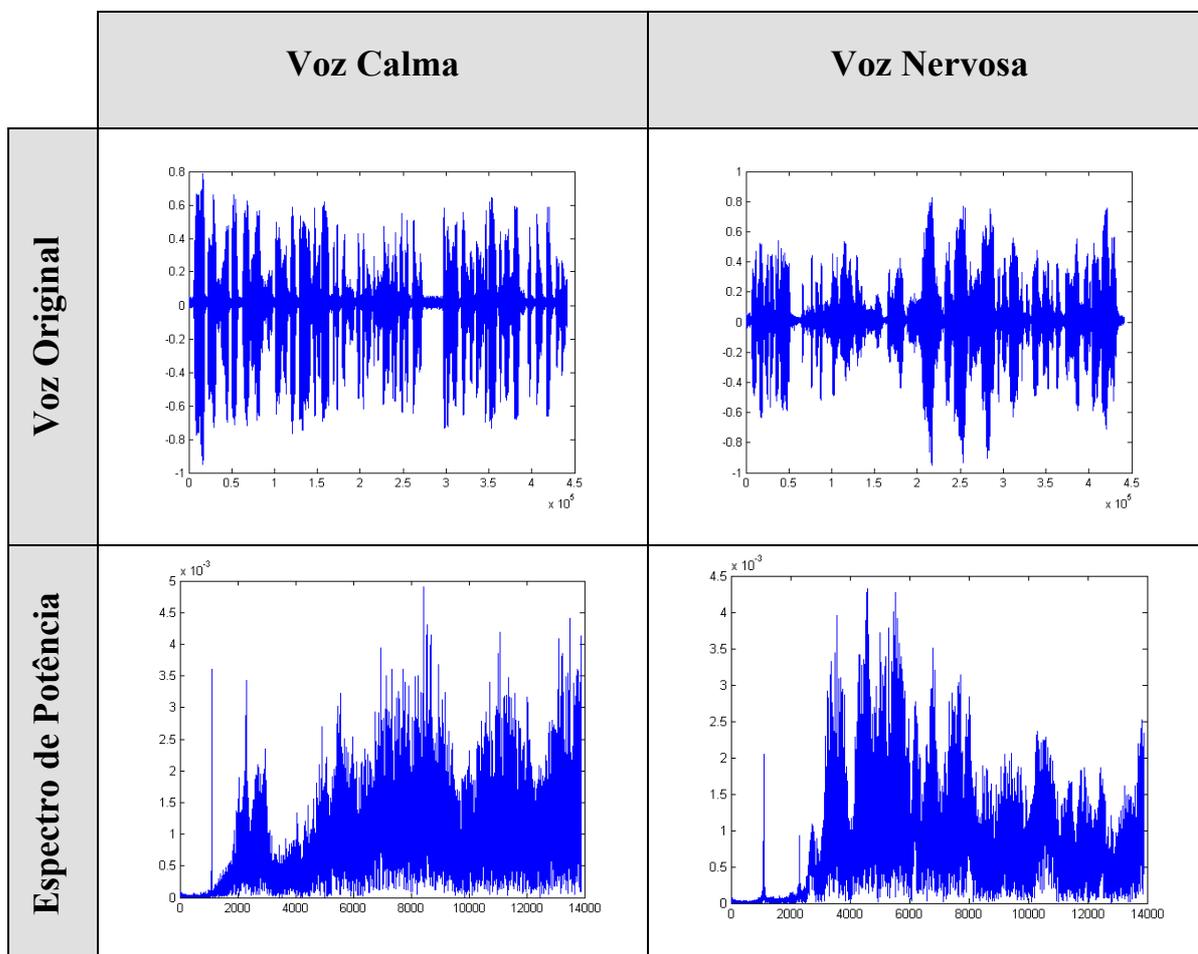


Tabela 4.6 – Cálculo do espectro de potência das vozes.

Após o cálculo do espectro de potência do sinal, foi realizado o procedimento de cálculo da Transformada Discreta Inversa de Fourier, exibida na tabela 4.7, através da etapa V do tópico 3.5 para a verificação da correlação entre o sinal original e o sinal da IDFT. Esta comparação mostrou que a correlação entre os dois sinais, para todas as vozes que foram pré-processadas pela DFT, ficou em torno de **96%** a **98%**, o que é considerado uma aplicação com sucesso na área de processamento digital de sinais.

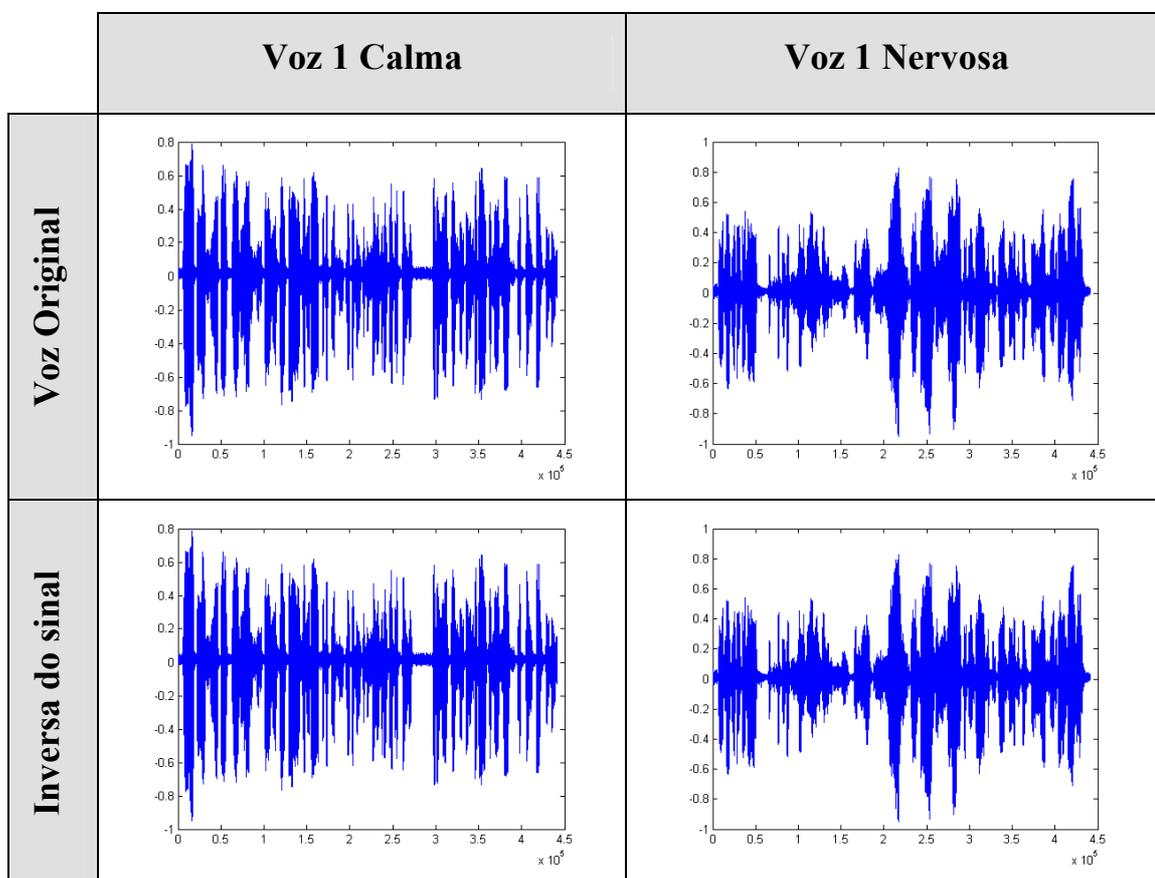


Tabela 4.7 – Comparação entre a voz original e a inversa do sinal

Sabendo a confiabilidade no processo de pré-processamento das vozes pela Transformada Discreta de Fourier através da tabela 4.7, o espectro de potência produzido de cada voz realmente representa as frequências que estão presentes no sinal durante os 10 segundos.

Para melhorar os dados de voz, foi realizado um processo de *filtragem do sinal*. Este procedimento consiste na eliminação das frequências abaixo de **20%** da frequência média do sinal. No caso da voz exibida na Figura 4.5 abaixo, a frequência média encontrada foi de **$0,923 \times 10^{-3}$** exibida por uma linha vermelha.

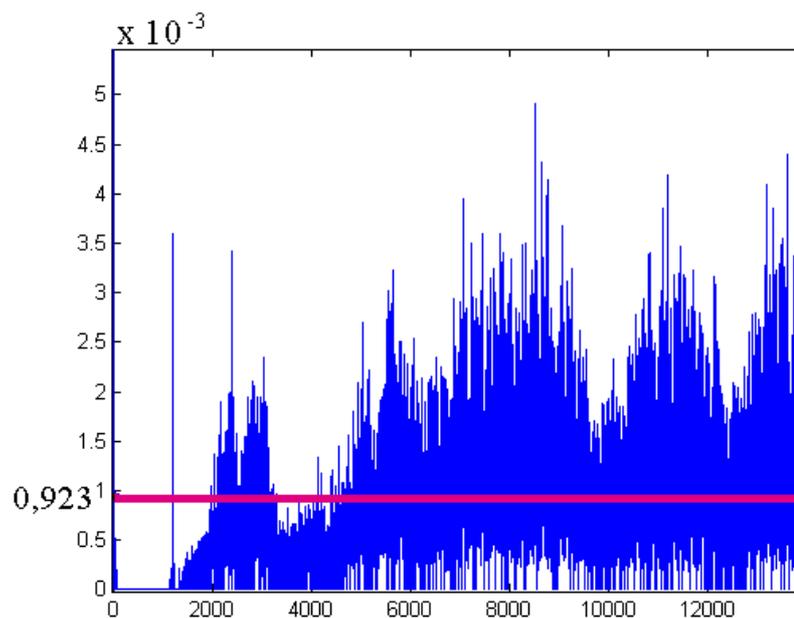


Figura 4.5 – Filtragem do sinal de voz.

Após a realização de todas as etapas do procedimento da Transformada Discreta de Fourier, algumas etapas foram adicionadas para a realização da filtragem do sinal que pode ser vista a seguir:

- **Etapa VII** – Calculando a frequência média do sinal:
 - $\text{freq_media} = \text{sum}(\text{espectro}) / \text{size}(\text{espectro}, 2);$
- **Etapa VIII** – Filtrando o sinal eliminando as frequências abaixo de 20% da frequência média:
 - $\text{sinal_filtrado} = \text{espectro} .* (\text{espectro} \geq 0.2 * \text{freq_media});$

Como para o treinamento da rede neural artificial os dados da camada de entrada da rede são inseridas em blocos de mil pontos, teve-se como interesse a utilização apenas das frequências nas faixas de **50 a 6549,5** Hertz, totalizando **13** blocos com mil pontos cada.

4.3.2 - Processamento das vozes pela RNA

Após a realização do pré-processamento das vozes pela DFT as trinta e seis vozes foram divididas, igualmente a aplicação anterior, e utilizadas no treinamento e generalização da rede Perceptron de Múltiplas Camadas (Back-Propagation) conforme mostra a tabela 4.8.

Vozes	Diagnósticos	Treinamento	Generalização
1 a 25	Calmas e Nervosas	X	
26 a 36	Calmas e Nervosas		X

Tabela 4.8 - Disposição das vozes pré-processadas utilizadas na rede MLP

Através da Tabela 4.9 abaixo são ilustradas as respostas desejadas para cada padrão de voz que é fornecida na camada de entrada da rede Perceptron de Múltiplas Camadas.

Diagnóstico	Codificação
Voz calma	0 1
Voz nervosa	1 0

Tabela 4.9 – Codificação dos diagnósticos das vozes pré-processadas

Também para este procedimento cada vetor de voz (13.000 x 1) foi dividido em blocos de mil (13 blocos com o tamanho: 1000 x 1) e inserido na camada de entrada da rede MLP. A seguir são ilustradas todas as etapas que serão utilizadas para o reconhecimento de níveis de stress na voz humana com vozes pré-processadas pela Transformada Discreta de Fourier.

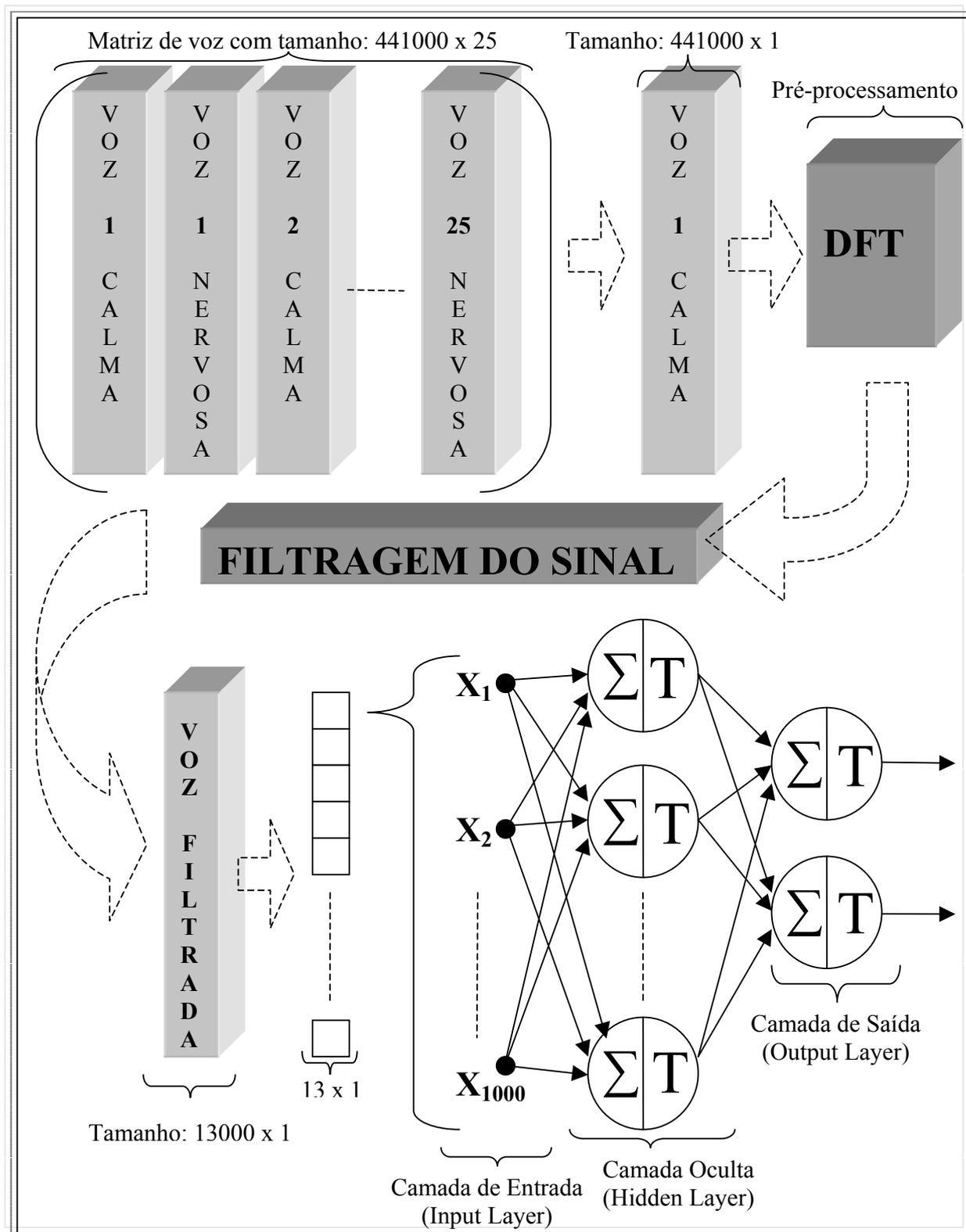


Figura 4.6 – Ilustração das Etapas utilizadas no treinamento da rede MLP com o pré-processamento da DFT

Para o treinamento da rede, como na aplicação anterior, foi utilizada uma taxa de aprendizado variada, onde a cada época a taxa era adaptada de acordo com o erro encontrado na saída da rede. A função logística utilizada na rede foi a função de ativação sigmoidal, com apenas uma camada oculta (hidden layer) com 100 neurônios.

A rede MLP foi processada durante **56.622** épocas, ou seja, as 50 vezes (entre calmas e nervosas) foram apresentadas **56.622** vezes com o objetivo de diminuir o erro quadrático médio da rede adaptando os pesos sinápticos da melhor maneira possível. Neste experimento o menor erro encontrado pela rede foi de **2.128**, ou seja, este erro é a soma dos erros de todas as entradas que são apresentadas à rede. A Figura 4.7 ilustra a trajetória do erro em função do número de épocas no processo de treinamento.

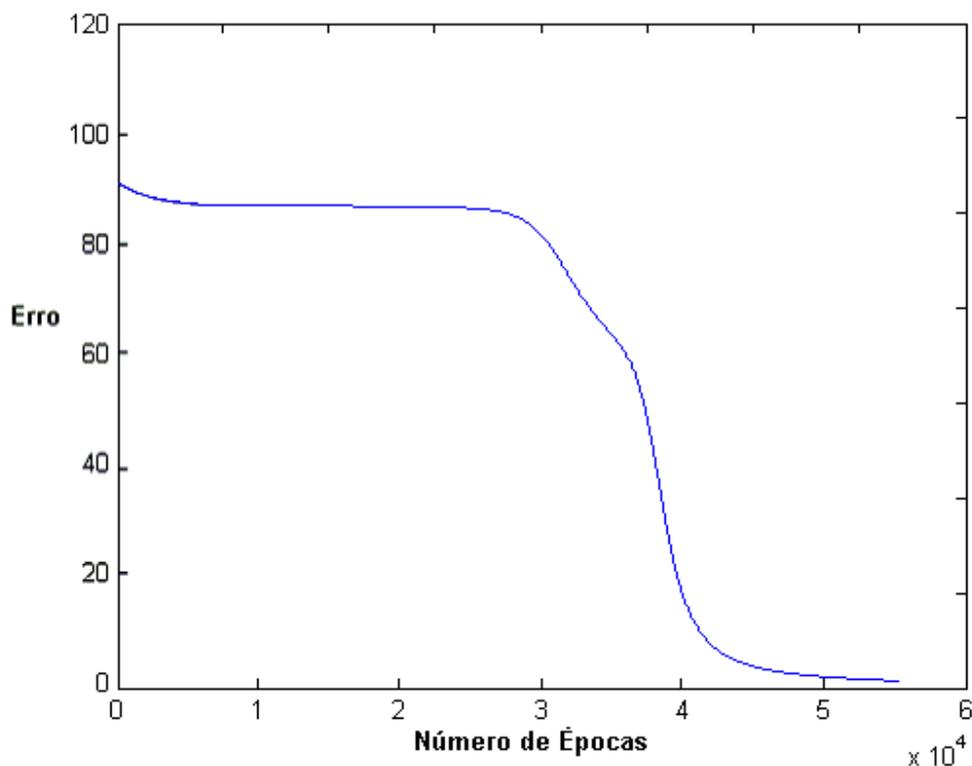


Figura 4.7 - Erro encontrado no treinamento da rede MLP.

O procedimento seguinte consistiu em fazer a generalização das dez vezes que foram utilizadas no treinamento, para saber se a rede Perceptron de Múltiplas Camadas aprendeu os padrões o qual treinou. Na ativação (generalização) da rede com as vezes que foram treinadas, observou um acerto de **90%** com um erro quadrático médio de **0.04**. Cada vez

utilizada no treinamento conseguiu determinada taxa de acerto conforme mostra a tabela 4.10, com destaque para algumas vozes nervosas que obtiveram um acerto de **100%**.

Voz	Diagnóstico	Acerto	Voz	Diagnóstico	Acerto
1	Calmo	92.31 %	14	Calmo	92.31 %
	Nervoso	84.62 %		Nervoso	84.62 %
2	Calmo	84.62 %	15	Calmo	92.31 %
	Nervoso	100.00 %		Nervoso	92.31 %
3	Calmo	92.31 %	16	Calmo	92.31 %
	Nervoso	92.31 %		Nervoso	100.00 %
4	Calmo	92.31 %	17	Calmo	92.31 %
	Nervoso	100.00 %		Nervoso	92.31 %
5	Calmo	92.31 %	18	Calmo	84.62 %
	Nervoso	92.31 %		Nervoso	100.00 %
6	Calmo	76.92 %	19	Calmo	76.92 %
	Nervoso	100.00 %		Nervoso	92.31 %
7	Calmo	69.23 %	20	Calmo	92.31 %
	Nervoso	100.00 %		Nervoso	100.00 %
8	Calmo	92.31 %	21	Calmo	84.62 %
	Nervoso	100.00 %		Nervoso	76.92 %
9	Calmo	84.62 %	22	Calmo	92.31 %
	Nervoso	92.31 %		Nervoso	92.31 %
10	Calmo	84.62 %	23	Calmo	84.62 %
	Nervoso	100.00 %		Nervoso	100.00 %
11	Calmo	84.62 %	24	Calmo	92.31 %
	Nervoso	92.31 %		Nervoso	100.00 %
12	Calmo	84.62 %	25	Calmo	69.23 %
	Nervoso	53.85 %		Nervoso	100.00 %
13	Calmo	84.62 %			
	Nervoso	69.23 %			

Tabela 4.10 – Porcentagem de acerto das vozes *treinadas* pela rede MLP.

Num segundo experimento, foram mostradas à rede as vinte e duas vozes (entre calmas e nervosas) que não foram utilizadas no treinamento, com o intuito de verificar a

porcentagem de acerto com padrões de vozes nunca vistos anteriormente pela rede MLP. Verificou-se que na ativação a rede conseguiu obter um acerto de **71,33%** com um erro quadrático médio encontrado na generalização de **0.266461**. A porcentagem de acerto dos diagnósticos de cada voz pode ser visto na tabela 4.11 abaixo.

Voz	Diagnóstico	Acerto	Voz	Diagnóstico	Acerto
1	Calmo	69.23 %	7	Calmo	15.38 %
	Nervoso	69.23 %		Nervoso	61.54 %
2	Calmo	92.31 %	8	Calmo	61.54 %
	Nervoso	100.00 %		Nervoso	92.31 %
3	Calmo	38.46 %	9	Calmo	84.62 %
	Nervoso	76.92 %		Nervoso	100.00 %
4	Calmo	76.92 %	10	Calmo	69.23 %
	Nervoso	84.62 %		Nervoso	69.23 %
5	Calmo	76.92 %	11	Calmo	76.92 %
	Nervoso	84.62 %		Nervoso	61.54 %
6	Calmo	61.54 %			
	Nervoso	84.62 %			

Tabela 4.11 Porcentagem de acerto das vozes *generalizadas* pela rede MLP.

4.3.3 - Conclusão

Com a realização deste procedimento concluiu-se que, com o pré-processamento das vozes pela aplicação da Transformada Discreta de Fourier, o treinamento destas pela RNA foi muito mais promissor do que com as vozes em seu estado original. O tempo demandado para o treinamento da rede MLP foi muito menor, onde se verificou que a rede obteve um reconhecimento melhor para as vozes nervosas do que para as calmas com destaque para as duas vozes (2 e 9) nervosas demarcadas na Tabela 4.11 que obtiveram **100%** de acerto pela rede neural. O acerto no diagnóstico de cada voz tanto no treinamento quanto na generalização da rede neural pode ser visualizada na tabela 4.12.

Rede Neural MLP com pré-processamento pela DFT			
Treinamento da rede		Generalização da rede	
Vozes calmas	89,95 %	Vozes calmas	67,83 %
Vozes nervosas	90,15 %	Vozes nervosas	74,83 %

Tabela 4.12 – Comparação entre vozes pré-processadas calmas e de estresse.

A desvantagem decorrida desta aplicação foi o tempo de processamento das vozes pela Transformada Discreta de Fourier, que durou cerca de 3 horas para cada voz. Deve-se ressaltar que esta implementação da DFT foi realizada apenas com o intuito de aprendizado e de mostrar que o pré-processamento da voz melhora o aprendizado de padrões pela rede neural.

Treinamento da RNA	COM a DFT	SEM a DFT
Duração de cada época	15 segundos	8 minutos
Tempo de treinamento (Épocas)	56.622 mil	13.754 mil
Duração total do treinamento	Cerca de 3 semanas	Cerca de 4 meses
Erro encontrado no treinamento da rede MLP	2,128	26,8855
Acerto no Treinamento das vozes calmas e nervosas	90 %	96,15 %
Acerto das vozes calmas	89,95 %	91,29 %
Acerto das vozes nervosas	90,15 %	61,07 %
Acerto na Generalização das vozes calmas e nervosas	71,33 %	56,56 %
Acerto das vozes calmas	67,83 %	80,13 %
Acerto das vozes nervosas	74,83 %	32,98 %

Figura 4.13 - Comparação entre os dois treinamentos realizados

Na tabela 4.13 acima, foi realizada uma comparação entre o aprendizado da RNA com o pré-processamento da voz pela DFT e o aprendizado sem este procedimento. Torna-se claro então que para a realização de diagnósticos de voz numa rede neural artificial o pré-processamento do sinal de voz para o domínio da frequência é muito mais eficiente, diminuindo consideravelmente o tempo de processamento e com um reconhecimento muito maior por parte das Redes Neurais Artificiais.

5 - CONCLUSÕES E RECOMENDAÇÕES FINAIS

O principal objetivo deste trabalho foi mostrar que o pré-processamento do sinal de voz pode propiciar uma melhora no reconhecimento de padrões de voz pelas Redes Neurais Artificiais. Para isto foi realizado um breve estudo sobre os conceitos de processamento de sinais e das características da voz humana.

O estudo realizado sobre técnicas de Fourier (Transformada Discreta de Fourier) propiciou o conhecimento e a implementação de alguns exemplos práticos para demonstrar a confiabilidade na transformação de um sinal no domínio do tempo para o domínio da frequência. O processamento digital dos sinais de voz é realizado no domínio da frequência.

A realização do pré-processamento pela Transformada Discreta de Fourier mostrou que uma capacidade melhor de aprendizado de padrões de vozes por uma rede neural do tipo Perceptron de Múltiplas Camadas, obtendo-se um desempenho de mais promissor (73,07%) do que o aprendizado utilizando os sinais de voz no estado original (56,5% - domínio do tempo). Outro fator relevante foi o tempo de treinamento da RNA realizado com as vozes pré-processadas que correspondeu à cerca de 97% mais rápido do que o treinamento sem o pré-processamento pela DFT.

Acredita-se na possibilidade de que, com um banco de dados com um número maior de vozes pré-processadas pela Transformada de Fourier e um tempo maior destinado ao treinamento da RNAs, os diagnósticos de índices de stress humano podem atingir índices de confiabilidade maiores, permitindo que se faça uma avaliação eficiente de pessoas que se encontram em estado de isolamento social.

Os resultados apresentados neste trabalho poderiam ser melhores. Mas muito tempo teve de ser dedicado ao estudo das técnicas de Transformada de Fourier, em detrimento do treinamento dos padrões de vozes.

Para a continuidade deste trabalho, tem-se como sugestão a utilização de outras técnicas em conjunto com a de processamento de sinais de voz, como por exemplo, a captura de frequências fundamentais características de uma voz em seu estado calmo ou de estresse. Outros mecanismos podem ser utilizados também para melhorar a identificação de uma pessoa em estado de isolamento social como, por exemplo, a utilização de testes psicológicos específicos na identificação de uma pessoa em confinamento, a pressão com que uma pessoa responde um questionário através de um teclado, a expressão facial, dispositivo de medição de batimentos cardíacos, da sudorese humana, impulsos elétricos emitidos pelo cérebro entre outros.

6 - REFERÊNCIAS BIBLIOGRÁFICAS

- [HAY-94] HAYKIN, Simon. **Neural Networks: A Comprehensive Foundation**
- [FAU-94] FAUSET, Laurene. Fundamentals of Neural Networks: Architectures, Algorithms, and Applications.
- [BRI-74] BRIGHAM, R., 1974, The Fast Fourier Transform: Time sequence Analysis in Geophysics.
- [KOV-96] KOVÁCS, Zsolt L. Redes Neurais Artificiais: Fundamentos e Aplicações.
- [GUE-77] GUEDES, Djairo de Fiquereido. Análise de Fourier e Equações Diferenciais Parciais
- [ALB-XX] ALBERTO, Luiz Vieira Dias, Marcio Magini. Introdução à Análise de Dados.
- [SED-95] SEDGEWICK, Robert. Algorithms in C. Princeton University - Cap.41 - The Fast Fourier Transform p.583.
- [BOY-96] BOYCE, Willian E. e Richard C. DiPrima. Equações Diferenciais Elementares e Problemas de Valores de Contorno.
- [GOL-97] GOLD, Ben. e Nelson Morgan. Speech and Audio Signal Processing: Processing and Perception of speech and music.
- [ALB-67] ALBERTO, Carlos Loprete. Introdução a Oratória Moderna: Voz Humana.