

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE

JOSÉ DA ASSUNÇÃO GOMES MENDES

**RECONHECIMENTO DA FALA SUBVOCAL BASEADO EM ELETROMIOGRAFIA
DE SUPERFÍCIE (EMG) UTILIZANDO ANÁLISE DE COMPONENTES
INDEPENDENTES (ICA) E REDE NEURAL MLP**

São Luís
2007

JOSÉ DA ASSUNÇÃO GOMES MENDES

**RECONHECIMENTO DA FALA SUBVOCAL BASEADO EM ELETROMIOGRAFIA
DE SUPERFÍCIE (EMG) UTILIZANDO ANÁLISE DE COMPONENTES
INDEPENDENTES (ICA) E REDE NEURAL MLP**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia de Eletricidade da Universidade Federal do Maranhão, para obtenção do grau de Mestre em Engenharia de Eletricidade, na área de Ciência da Computação.

Orientador: Prof. Dr. Sofiane Labidi
Co-orientador: Prof. Dr. Allan Kardec Barros

São Luís
2007

Mendes, José da Assunção Gomes.

Reconhecimento da fala subvocal baseado em eletromiografia de superfície (EMG) utilizando análise de componentes independentes (ICA) e rede neural MLP / José da Assunção Gomes Mendes. – 2008.

77f.

Impresso por computador (fotocópia).

Orientador: Sofiani Labidi

Dissertação (Mestrado) – Universidade Federal do Maranhão, Curso de Programa de Pós-Graduação em Engenharia de Eletricidade, São Luís, 2008.

1. Processamento da fala. 2. Eletromiografia de superfície. I. Labidi, Sofiani, orient. II. Título.

CDU 004.934

**RECONHECIMENTO DA FALA SUBVOCAL BASEADO EM
ELETROMIOGRAFIA DE SUPERFÍCIE (EMG) UTILIZANDO
ANÁLISE DE COMPONENTES INDEPENDENTES
(ICA) E REDE NEURAL MLP**

José da Assunção Gomes Mendes

Dissertação aprovada em 19 de dezembro de 2007.



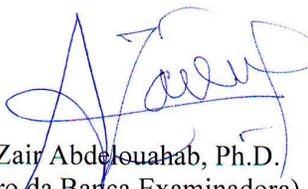
Prof. Sofiane Labidi, Dr.
(Orientador)



Prof. Allan Kardec Duailibe Barros Filho, Ph. D.
(Co-orientador)



Prof. Guilherme de Alencar Barreto, Dr.
(Membro da Banca Examinadora)



Prof. Zair Abdelouahab, Ph.D.
(Membro da Banca Examinadora)

À Deus
A minha família, Valdenia e Lucas
Ao meu pai (*in memoriam*)

AGRADECIMENTOS

A DEUS, pela sabedoria e sapiência concebida para elaboração deste trabalho, e pela força para ter chegado até aqui, com muita dificuldade e esforço.

A minha mãe e meu pai (*in memoriam*) pelo amor e dedicação recebidos durante toda a minha vida, pois, tenho certeza que o sonho, não só deles, mas como de todos os pais é de verem seus filhos conquistando mais essa vitória.

A minha família que esteve e sempre estará presente nos bons e maus momentos.

Ao meu orientador Prof. Dr. Sofiane Labidi pela confiança e orientação deste tema.

Ao meu co-orientador Prof. Dr. Allan Kardec Barros pela grande aceitabilidade, credibilidade e co-orientação deste tema.

Ao meu filho Lucas pela força, idéias, amizade e amor.

A minha esposa Valdenia pelo amor, carinho e paciência.

Ao engenheiro eletricitista Ricardo Robson pela elaboração do sistema de aquisição de dados, utilizado neste trabalho.

Ao meu amigo peruano Juan Mauricio Villanueva pelas contribuições no desenvolvimento deste trabalho

Aos colegas do Laboratório de Processamento Biológico (PIB) e do Laboratório de Sistemas Inteligentes (LSI) que contribuíram para o desenvolvimento deste trabalho.

A todas aquelas pessoas que diretamente me ajudaram, na aquisição dos dados, servindo como objeto de estudo, necessários para validação deste trabalho.

"Fala subvocal é silenciosa, ou fala subaudível, tal como, quando uma pessoa silenciosamente lê ou fala pra si mesma".

Chuck Jorgensen

RESUMO

O desempenho dos sistemas de reconhecimento da fala é comumente degradado por incapacidades relacionadas com a fala ou por através de fatores do mundo real tais como nível de ruído do ambiente e reverberação. Nesta pesquisa, nós propomos um sistema de reconhecimento subvocal da fala. Este sistema é baseado em Eletromiografia de superfície (sinal EMG) para aquisição de dados subvocais, Análise de Componentes Independentes (ICA) para extração das características e Rede Neural MLP para classificação. Nós avaliamos o desempenho do sistema usando um banco de dados dos fonemas das vogais subvocais. De acordo com os resultados obtidos, a metodologia proposta obteve uma taxa de sucesso de 93,99%.

Palavras-chave: Eletromiografia, ICA, Rede Neural, Fala Subvocal

ABSTRACT

The performance of speech recognition systems is commonly degraded by either speech-related disabilities or by real-world factors such as the environment's noise level and reverberation. In this research, we propose a subvocal speech recognition system based on electromyography (EMG signal) for subvocal acquisition, Independent Component Analysis (ICA) for feature extraction and Neural Networks MLP for classification. We have evaluated the system's performance using a subvocal vowel phonemes database. According to the results, the methodology proposed obtained a success rate of 93.99%.

Keywords: Electromyography. ICA. Neural Network. Subvocal Speech

LISTA DE FIGURAS

Figura 1	Representação TBB (Transcrição biunívoca brasileira) entre barras e Representação IPA (International Phonetic Association) entre colchetes.....	19
Figura 2	Cordas Vocais: (a) - Abertas (Respiração); (b) - Fechadas (Fonação).	22
Figura 3	Funcionamento básico do aparelho fonador.....	24
Figura 4	Arquitetura de uma RNA com duas saídas.....	53
Figura 5	Ilustração das direções de dois fluxos básicos de um sinal em uma rede neural MLP: Sentido direto e Sentido reverso.....	57
Figura 6	Sistema de reconhecimento subvocal da fala proposto: (a) – Fase de Aquisição de Dados; (b) - Fase de Aprendizagem dos Dados; (c) -Fase de Classificação dos Dados.....	59
Figura 7	Circuito de Amplificação e Filtragem (Canal 1 (CH1)), (Canal2 (CH2)).....	60
Figura 8	Microcontrolador (PIC 16873P).....	62
Figura 9	Esquema ilustrativo do processo de aprendizagem numa RNA.....	65
Figura 10	Diagrama sistema subvocal (Posicionamento dos eletrodos e Gravação).....	68
Figura 11	Sinais EMG das vogais subvocais: (a) - vogal subvocal [a], (b) - vogal subvocal [ɛ].....	69
Figura 12	Arquitetura da rede neural MLP usada.....	70

LISTA DE SIGLAS

ATM	Articulação Temporomandibular
BSS	Separação cega de fontes
EMG	Eletromiografia
IA	Inteligência Artificial
ICA	Análise de Componentes Independentes
MLP	Multicamadas Perceptron
PCA	Análise de Componentes Principais
RBF	Função de Base Radial
RNA	Rede Neural Artificial
RNA's	Redes Neurais Artificiais
SAD	Sistema de Aquisição de Dados

LISTA DE TABELAS

Tabela 1	Desempenho das redes neurais MLP.....	71
Tabela 2	Desempenho das redes neurais RBF.....	71
Tabela 3	Classificação das vogais subvocais.....	72
Tabela 4	Proposta das pesquisas anteriores.....	72
Tabela 5	Proposta desta pesquisa.....	73

LISTA DE QUADROS

Quadro 1	Aparelho Digestivo.....	16
Quadro 2	Aparelho Respiratório.....	17

SUMÁRIO

	LISTA DE FIGURAS	09
	LISTA DE SIGLAS.....	10
	LISTA DE TABELAS.....	11
	LISTA DE QUADROS.....	12
1	INTRODUÇÃO	15
1.1	Aparelho fonador.....	16
1.2	Fala humana.....	17
1.3	Reconhecimento da fala	19
1.4	Mecanismo de fonação	21
1.5	Trabalhos relacionados	25
1.6	Organização do trabalho	26
2	ELETROMIOGRAFIA (EMG).....	27
2.1	Aquisição do sinal eletromiográfico.....	28
2.1.1	Freqüência de amostragem.....	28
2.1.2	Eletrodos (Sensores).....	29
2.2	Fisiologia.....	34
2.3	Uso da eletromiografia nas diversas áreas de saúde	34
2.3.1	Fisioterapia	34
2.3.2	Medicina	35
2.3.3	Educação Física	35
2.3.4	Fonoaudiologia	36
2.3.5	Odontologia	37
2.3.6	Terapia ocupacional	37
2.4	Conclusão	42
3	SISTEMA DE RECONHECIMENTO SUBVOCAL	39
3.1	Fala subvocal.....	39
3.2	Reconhecimento subvocal	40
3.3	Conclusão	42
4	ANÁLISE DE COMPONENTES INDEPENDENTES (ICA).....	43

4.1	Definições	44
4.2	Definição de independência	45
4.3	Técnicas de estimação das componentes independentes	46
4.4	Negentropia como medida de não-gaussianidade	48
4.5	Seleção de características mais significantes	50
4.6	Conclusão	51
5	REDES NEURAIS ARTIFICIAIS	52
5.1	Redes neurais perceptron multicamadas (MLP)	55
5.2	Conclusão	58
6	MODELO PROPOSTO	59
6.1	Fase de aquisição de dados	59
6.1.1	Placa de aquisição de dados	60
6.1.2	Eletrodos (Sensores)	63
6.2	Fase de aprendizagem de dados	63
6.3	Fase de classificação dos dados	64
6.4	Medidas de avaliação de um sistema diagnóstico	65
6.5	Conclusão	66
7	RESULTADOS E DISCUSSÃO	67
7.1	Resultados	67
7.1.1	Dados capturados na fase de aquisição	67
7.2	Dados subvocal	67
7.3	Classificação dos dados usando rede neural MLP	70
7.4	Conclusão	66
8	CONCLUSÃO E TRABALHOS FUTUROS	73
	REFERÊNCIAS	75

1 INTRODUÇÃO

Este capítulo tem por objetivo apresentar os conceitos sobre os seguintes tópicos: Aparelho fonador; Fala Humana; Reconhecimento da fala; Mecanismo da fonação e os Trabalhos relacionados com esta pesquisa, visto a importância multidisciplinar do tema em questão que é o reconhecimento subvocal da fala baseado na eletromiografia de superfície (EMG), utilizando análise de componentes independentes (ICA) e rede neural multicamadas perceptron (rede MLP).

A pesquisa tem por objetivo abordar o problema de reconhecimento subvocal da fala na língua portuguesa brasileira (fonemas das vogais orais e abertas) para pessoas com aparelho fonador normal, através da implementação prática de um sistema de reconhecimento da fala subvocal baseado em sinais EMG e um sistema de aquisição de dados, recorrendo à teoria da análise de componentes independentes (ICA) usada na extração de funções bases e de Rede Neural MLP para a classificação dos sinais EMG (vogais subvocais).

O modelo proposto é composto de 03 (três) fases: aquisição dos dados; aprendizagem dos dados e classificação dos dados. Este sistema foi implementado utilizando um sistema de aquisição de dados, scripts Matlab[®] e um de banco de dados de sinais EMG (vogais subvocais) de 10 (dez) pessoas através de um programa C, armazenado no microcontrolador e um script Matlab[®] de aquisição dos sinais EMG. Através da técnica de ICA é construído um subespaço no qual são projetados os sinais subvocais das vogais. Desse modo, qualquer outro sinal que não seja projetado neste subespaço será descartado pelo sistema..

Os resultados alcançados mostram que ICA é uma técnica bastante eficiente na aprendizagem dos sinais EMG (vogais subvocais) e a rede neural MLP na classificação desses sinais, alcançando um desempenho de 95,4%, utilizando a técnica de validação cruzada, com uma configuração de 185:21:5 (185 neurônios na camada de entrada, 21 neurônios na camada escondida e 5 neurônios na camada

de saída), e uma taxa de sucesso de 93,99% no reconhecimento da fala subvocal, baseado em Eletromiografia de superfície (EMG), utilizando Análise de Componentes Independentes (ICA) e Rede neural MLP.

1.1 Aparelho fonador

O ser humano não possui nenhum aparelho destinado exclusivamente à produção do som. O aparelho fonador de que se vale para expressão vocal passa de uma adaptação secundária, de um aproveitamento parasitário de partes do digestivo e respiratório. A laringe aparece na escala animal quando é necessário proteger o aparelho respiratório contra a entrada de sólidos ou líquidos que pudessem causar asfixia.

O aparelho fonador é um conjunto adaptado de outros aparelhos: respiratório e digestivo. O aparelho fonador é composto de: pulmões, traquéia, laringe (onde estão situadas as cordas vocais), faringe, cavidades oral e nasal, sendo estes os responsáveis pela produção da fala (SOUZA, 2000).

No aparelho fonador esses 2 (dois) aparelhos têm a função de produzir sons: voz cantada e voz falada. Nos quadros abaixo, são mostradas as funções biológica e fonatória dos órgãos que compõem o aparelho digestivo e o aparelho respiratório (VIANA, 2007).

ÓRGÃO	FUNÇÃO BIOLÓGICA	FUNÇÃO FONATÓRIA
Lábios	Contém os alimentos na boca	Articulação de sons bilabiais (B,P,M) e labiodentais (F,V)
Dentes	Tritura os alimentos	Escoamento do som
Língua	Joga o alimento para o esôfago	Participa de todos os sons produzidos
Palato duro (céu da boca)	Suporte da língua	Projeção da voz
Faringe	Direciona o ar para os pulmões, e os alimentos para o esôfago.	Caixa de ressonância

Quadro 1: Aparelho Digestivo
Fonte: VIANA, 2007

ÓRGÃO	FUNÇÃO BIOLÓGICA	FUNÇÃO FONATÓRIA
Cavidades Nasais	Filtrar, aquecer e umidificar o ar	Vibração e amortização do som - ressonância nasal
Faringe	Via de passagem do ar	Amplia os sons - caixa de ressonância
Laringe	Via de passagem do ar	Vibrador - contém as cordas vocais
Traquéia	Via de passagem do ar - defesa a via aérea	Suporte para vibração das cordas vocais
Pulmões	Trocas gasosas e respiração vital	Fole e reservatório de ar para vibrar as cordas vocais
Musculatura respiratória	Desencadeia o processo respiratório	Produção de pressão no ar que sai

Quadro 2: Aparelho Respiratório.
Fonte: VIANA, 2007

No que diz respeito ao aparelho fonador, a única parte que se poderia considerar parcialmente original é a parte que pertence ao sistema nervoso. Na verdade, o ato de falar envolve todo o organismo. Com a prática constante, a fala acaba por se tornar um mecanismo automático.

Ver uma pessoa através de sua fala é uma realidade. É que pela fala pode-se ver realmente todo o indivíduo. Ela nos revela o ser humano nos seus mais diversos aspectos. Ouvindo alguém falar, pode-se deduzir o estado real do aparelho fonador, o estado neuropsíquico do paciente, o estado glandular e seu grau de cultura. A voz, por outro lado, ainda é mais reveladora do que a própria fala, pois se podem distinguir nuances e sutilezas, uma vez que a voz é individual, enquanto a fala é comum de todos (BEUTTENMÜLLER; BEUTTENMÜLLER, 1995).

1.2 Fala humana

Supõe-se que a linguagem tenha começado a desenvolver-se há cerca de 1,5 milhões de anos, com o Australopithecus. Um dos últimos passos na evolução da linguagem ocorreu com o Homo sapiens a mais de 40.000 anos. O desenvolvimento da linguagem permitiu que os conhecimentos adquiridos pelos homens pudessem ser compartilhados e possivelmente passados de uma geração para outra, abrindo novos horizontes à espécie.

A fala é uma das capacidades ou aptidões de comunicação que os seres humanos possuem para manifestar seus pensamentos, opiniões e sentimentos através de vocábulos que transladam textualmente quando necessários. Consiste no principal sinal entre os distintos sinais abordados pela linguagem natural, como por exemplo, ideogramas, gestos, gritos, trejeitos e outros tipos de linguagem corporal.

Os sons da fala não devem ser confundidos com os fonemas da Língua Portuguesa, uma vez que, o som é entendido como uma complexa realidade física acústica de cada unidade sonora da fala, enquanto que os fonemas correspondem à percepção eclética e interpretativa realizada pelo falante e ouvinte, respectivamente. O fonema é a menor unidade sonora ou articuladora da palavra. Os fonemas são produzidos pelo aparelho fonador, que é constituído por um conjunto de órgãos que auxiliam no ato da fala. A fala representa a emissão de sons articulados na linguagem oral enquanto que a produção da voz é a fonação (PAULA, 2000).

A fala é o som resultante da vibração das cordas ou pregas vocais, provocadas pelo ar expelido pelos pulmões e que atravessa a traquéia, a laringe, a faringe e a boca. A fonação é justamente a produção de sons pela ação de vários pequenos músculos localizados na laringe, sendo a prega vocal um deles.

Os sons da fala são classificados desde a Antigüidade em dois grupos principais: vogais e consoantes. Vogais são os sons da fala que ocorrem quando o fluxo de ar percorre o aparelho fonador livre e continuamente, sem causar turbulência ou enfrentar obstrução temporária. Quando o fluxo de ar provoca turbulência no aparelho fonador ou encontra uma obstrução temporária, teremos a emissão de consoantes.

As vogais são também os sons musicais produzidos pela vibração das cordas vocais. São chamados fonemas silábicos, pois constituem o fonema central de toda sílaba.

A emissão de uma vogal específica requer o posicionamento de várias

partes do aparelho fonador. A língua se move dentro da boca para frente e para trás, para cima e para baixo. Os lábios podem ficar distensos ou arredondados. O véu que fica no fundo da boca pode ficar elevado ou rebaixado, com isso permite que parte do fluxo de ar saia pela cavidade nasal. A combinação de todas essas possibilidades de posicionamento permite a produção um número indefinido de vogais. (GRAMÁTICA DESCRITIVA, 2007)

Vale ressaltar que nesta pesquisa serão usados os fonemas das 5 (cinco) vogais orais abertas ([a], [ɛ], [i], [ɔ], [u]) conforme classificação quanto à função das cavidades bucal e nasal e quanto ao timbre, para formar o banco de dados das vogais subvocais.

A Figura 1 mostra a representação das vogais orais abertas, conforme Representação TBB (Transcrição biunívoca brasileira) e Representação IPA (International Phonetic Association).

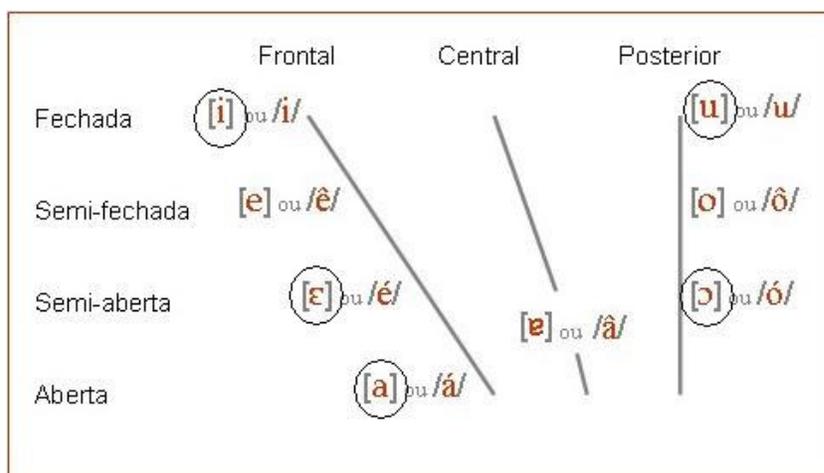


Figura 1: Representação TBB (Transcrição biunívoca brasileira) entre barras e Representação IPA (International Phonetic Association) entre colchetes.
Fonte: GRAMÁTICA, 2007

1.3 Reconhecimento da fala

Desde os primórdios da revolução industrial, o homem sonhava em desenvolver autômatos, posteriormente chamado de robôs, dotados de certa inteligência e capacidade de compreensão, para poder desempenhar tarefas de

forma autônoma, a fim de substituir o trabalho humano, como por exemplo, através de comandos da fala. A tecnologia existente na época não era suficiente para que tais avanços pudessem ser alcançados. As primeiras conquistas foram obtidas através da descoberta de novos conhecimentos relacionados ao processamento de informação, que proporcionaram o surgimento de computadores com um maior poder de processamento e armazenamento, e o surgimento da Inteligência Artificial.

O campo da Inteligência Artificial (IA) cujo nome foi proposto em 1956 por MacCarthy é composto por um agregado de métodos que visa solucionar determinados tipos de problemas, na área da engenharia e computação. Estes métodos, mais conhecidos como algoritmos computacionais ou programas, tentam se espelhar na maneira de resolução de problemas adotados pelo homem.

As redes neurais artificiais (RNA's), um dos nichos da IA, propõe uma metodologia para solucionar alguns problemas relacionados à área de inteligência artificial, modelando sistemas através de conexões que possam simular o sistema nervoso humano, abrangendo a capacidade que o mesmo possui de aprender e agir perante as mais adversas situações apresentadas.

Uma das aplicações das RNA's é o reconhecimento de palavras faladas, que, por sua vez, auxilia e principalmente agiliza algumas tarefas desempenhadas manualmente tais como a digitação de um texto, controle de eletroeletrônicos, comandos de voz para robôs para a realização de algumas funções que podem ou não colocar em risco a vida humana. Cabe também ressaltar o considerável auxílio que esta tecnologia pode proporcionar às pessoas portadoras de deficiência física.

O reconhecimento da fala consiste em identificar fonemas, sílabas, palavras para formar a mensagem original, ou uma informação onde existe uma seleção mais direta da resposta, sem interpretação dela. Assim, uma ação pode ser executada diretamente quando um padrão falado é reconhecido.

1.4 Mecanismo de fonação

A fonação é uma função neurofisiológica inata, mas a voz vai se formando ao longo da vida, de acordo com as características anatomofuncionais do indivíduo e os aspectos emocionais de sua história pessoal (BELHAU, 2004).

Percebe-se que o mecanismo de fonação envolve os centros de controle específicos da fala no córtex cerebral, funções mecânicas da produção da voz e o controle do som para produzir um fonema definido. O aparelho fonador está constituído por regiões definidas dos aparelhos digestivo e respiratório, sendo auxiliado pelo mecanismo de audição. Sua principal função é produzir os sons, que podem ser a voz cantada ou a voz falada. O aparelho fonador pode ser dividido em seis partes: 1. *Produtor*; 2. *Condutor*; 3. *Vibrador*; 4. *Ressoador*; 5. *Articulador*; 6. *Sensor/Coordenador*.

O *produtor* é formado pelos pulmões, músculos abdominais, diafragma, músculos intercostais e músculos extensores da coluna. Os pulmões se apresentam como uma massa esponjosa e elástica, que ocupa toda a cavidade torácica e estão protegidos pela caixa de mesmo nome, que é formada por diversos músculos e ossos. Além de serem responsáveis pelos fenômenos respiratórios (inspiração e expiração) e pela troca gasosa, os pulmões juntamente com as vias respiratórias superiores servem de reservatório de ar que possibilitam a vibração das cordas vocais.

O *condutor* é caracterizado pelo conjunto de tubos e túbulos que compõe o aparelho fonador sendo responsáveis pela produção da coluna de ar que

pressiona a laringe fazendo vibrar as cordas vocais e produzindo o som.

O *vibrador* é composto pelas cordas vocais e a laringe. A laringe, como mencionada anteriormente, é considerada o órgão da fala, pois nela estão localizadas as cordas vocais. As cordas são formadas por duas membranas que se localizam nas bordas da glote e que vibram à passagem do ar. Logo, quando uma pessoa está respirando normalmente, as cordas vocais estão separadas uma da outra, formando uma abertura em forma de “V”. A produção do som se dá quando os músculos da laringe tornam rijas as cordas vocais, produzindo o som quando o ar as atravessa.

Na Figura 2 são mostradas as posições que cordas vocais se encontram em processo de *respiração* (cordas vocais abertas) e de *fonação* (cordas vocais fechadas).

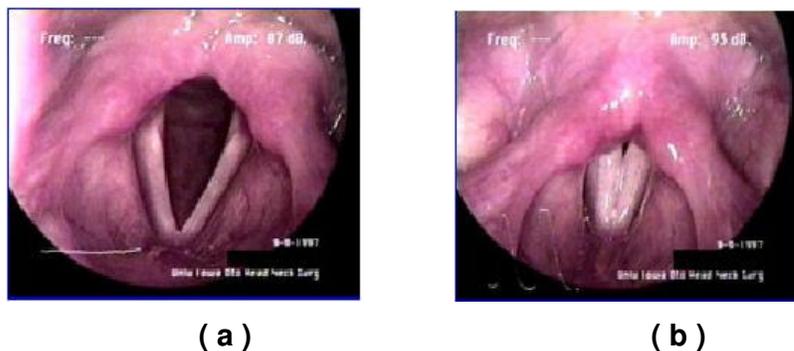


Figura 2: Cordas Vocais: (a) - Abertas (Respiração); (b) - Fechadas (Fonação)
Fonte: FÉLIX, 2007

O homem adulto tem as cordas vocais maiores e muito espessas, por isso numa emissão vocal, em um período de 1 segundo, suas pregas vibram, em média, 113 vezes. Espera-se que a frequência fundamental para o homem adulto esteja ao redor de 80 a 150 Hz. A mulher adulta tem as pregas vocais menores e menos espessas. Numa emissão, em um período de 1 segundo, suas pregas vocais vibram em média 204 vezes. Espera-se que a frequência fundamental da mulher adulta esteja ao redor de 150 a 250 Hz. Para crianças até a puberdade, as pregas são menores e mais delgadas (leves). Em um período de 1 segundo, suas pregas vocais

vibram em média 260 vezes, sendo que a frequência fundamental esteja ao redor dos 235 Hz, podendo variar de acordo com a idade (COIMBRA, 2007)

O *ressoador* é formado pela cavidade nasal, a faringe e a boca, e tem como função ampliar o som. As fossas nasais são duas cavidades existentes no nariz, onde o ar entra para o organismo. Elas são separadas por uma estrutura chamada de septo nasal que se comunica com o exterior através das narinas, com a boca e a faringe pelas coanas (aberturas de comunicação das narinas com a cavidade bucal). Sua função biológica é de filtrar, aquecer e umidificar o ar. Como contribuição fonatória, visa à vibração e amortização do som, resultando na ressonância nasal.

O *articulador* constituiu-se pelos componentes responsáveis pela articulação dos sons, transformando-os em orais e nasais. O sentido dado ao som é outra função desempenhada por este componente. O articulador compreende os órgãos localizados na boca como lábios, língua, palato, maxilar, dentes, úvula e véu palatino. O principal órgão articulador é a língua, que participa de todos os sons produzidos e interfere diretamente na formação das vogais e consoantes. Os lábios são dobras membranosas localizadas na parte externa da boca e que se dividem em superior e inferior, os quais têm como função fonatória produzir os sons bilabiais da língua falada (Língua Portuguesa Brasileira). Os dentes são órgãos que guarnecem as maxilas e não só auxiliam na mastigação, mas também fazem parte do sistema fonador, contribuindo para o escoamento do som produzido.

O *Palato* se divide em duas partes: o palato duro, céu da boca, e o palato mole onde se encontram a úvula. O palato duro como articulador é responsável pela projeção da voz. Já o palato mole (úvula), permite ou não a passagem do som produzido pelo *ressoador*. O véu palatino, conjunto de músculos ligado ao palato, caracteriza a distinção entre os fonemas produzidos. Se a coluna de ar encontrar o véu levantado, fechando a passagem pelas fossas nasais, haverá sons/fonemas orais, caso contrário, nasais.

A Figura 3 mostra o funcionamento básico do aparelho fonador, onde os pulmões forçam o ar em direção à glote, que atingindo as cordas vocais faz com que elas vibrem. A boca e o nariz contribuem para a modulação da voz.

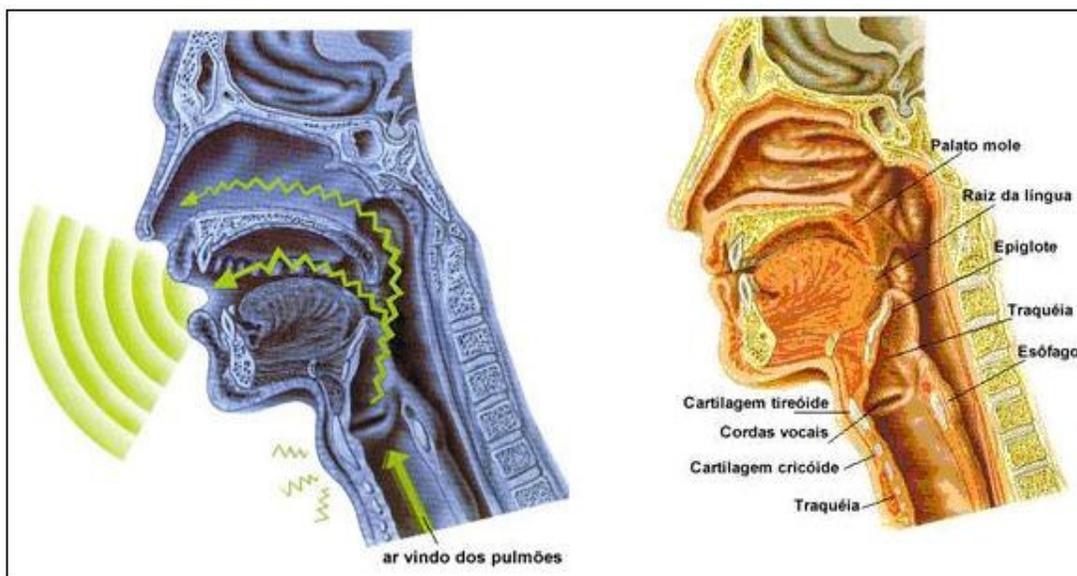


Figura 3: Funcionamento básico do aparelho fonador
Fonte: D'AMBROSIO, 2007

A última parte que integra o mecanismo de fonação é o *ouvido* (mecanismo de audição como feedback) que tem como função captar, selecionar e interpretar o som. O ouvido humano é um órgão sensível que possibilita captar, perceber e interpretar ondas sonoras numa faixa de frequência que está compreendida entre 20 e 20 KHz (PAULA, 2000)

Vale ressaltar que através desta pesquisa foi possível observar que existem 3 (três) tipos de cordas vocais: corda vocal do homem, corda vocal da mulher e corda vocal da criança. Contudo informa-se que não é meta desta pesquisa a análise da distinção entre ambas. E também ressaltar que as pessoas que participaram desta pesquisa como objetos de estudo possuem o aparelho fonador normal e a independência de saber se era homem, mulher ou criança relacionado às suas cordas vocais.

1.5 Trabalhos relacionados

Alguns trabalhos apresentam o uso de sensores EMG, no Reconhecimento da Fala Subvocal e Reconhecimento da Fala Subaudível, a saber:

Primeiramente, tem-se Chuck Jorgensen e Kim Binsted (JORGENSEN; BINSTED, 2000) que propuseram uma classificação do sinal do eletromiograma (EMG) subvocal para controlar e modificar uma interface de um browser de página de internet.

Em seguida, tem-se Chuck Jorgensen, D. Diana Lee e Shane Agabon (JORGENSEN; LEE; AGABON, 2003) que propuseram uma classificação do sinal do palatograma/eletromiograma (EMG/EPG), como um método para o reconhecimento da fala silenciosa.

Kim Binsted e Charles Jorgensen (BINSTED; JORGENSEN, 2003), que propuseram um reconhecimento da fala subauditiva, usando sensores eletromiograma (EMG) para uso de interface em ambientes ruidosos, para comunicações discretas ou seguras e para usuários com as deficiências relacionadas com a fala.

Por último, J. Bradley Betts e Charles Jorgensen (BETTS; JORGENSEN, 2006) que propuseram um Sistema Reconhecimento da Fala baseado em EMG para um pequeno vocabulário de 15 (quinze) palavras da língua inglesa.

Os trabalhos anteriormente citados foram desenvolvidos usando a língua inglesa (dígitos, fonemas da vogal, fonemas consoantes e palavras).

Nesta pesquisa é apresentado um sistema de reconhecimento da fala subvocal baseado em Eletromiografia de superfície (EMG), utilizando Análise de Componentes Independentes (ICA) e Rede Neural perceptron MLP, usando os

fonemas das vogais orais e abertas da língua portuguesa brasileira. O sistema proposto é constituído de 3 (três fases): aquisição, treinamento e classificação.

1.6 Organização do trabalho

Este trabalho está dividido em oito capítulos:

No capítulo 2, são abordados os conceitos sobre eletromiografia, aquisição do sinal eletromiográfico (frequência de amostragem, eletrodos (sensores)), fisiologia e o uso da eletromiografia nas diversas áreas de saúde.

Por sua vez, o capítulo 3 aborda os conceitos sobre a fala subvocal e descrição do reconhecimento subvocal. O capítulo 4 traz o histórico e a fundamentação teórica da análise de componentes independentes (ICA) usada para extrair as funções bases das amostras das 5 (cinco) vogais subvocais.

Apresentam-se as definições sobre redes neurais artificiais e especificamente uma abordagem sobre redes neurais perceptron multicamadas (MLP) no capítulo 5. O capítulo 6 apresenta os resultados alcançados com a implementação da técnica de ICA e Rede Neural na extração de funções bases, classificação e reconhecimento da fala subvocal.

O capítulo 7 traz os resultados dos experimentos que foram descritos nesta pesquisa. Por fim, o capítulo 8, apresenta as conclusões e sugestões para trabalhos futuros.

2 ELETROMIOGRAFIA (EMG)

A eletromiografia é a ciência que estuda os fenômenos bioelétricos que ocorrem nas membranas celulares das fibras musculares esqueléticas durante o repouso, o esforço mínimo quando se tenta isolar a unidade motora, ou ainda ver várias unidades motoras trabalhando isoladas e o esforço contrátil máximo. Unidades motoras se compõem de um neurônio do corno anterior, um axônio, suas junções neuromusculares, e todas as fibras musculares inervadas por este axônio. O axônio simples conduz um impulso para todas as suas fibras musculares, fazendo com que sofram despolarização de modo relativamente simultâneo. A despolarização produz atividade elétrica, que se manifesta como potencial de ação da unidade motora que é graficamente registrada como eletromiograma no eletromiógrafo.

A eletromiografia utiliza a técnica de monitoramento da atividade elétrica das membranas excitáveis, representando a medida dos potenciais de ação do sarcolema, como efeito de voltagem em função do tempo. O sinal eletromiográfico (EMG) é a soma algébrica de todos os sinais detectados em certa área, podendo ser afetado por propriedades musculares, anatômicas e fisiológicas, assim como pelo controle do sistema nervoso periférico e a instrumentação utilizada para a aquisição dos sinais (MARCHETTI; DUARTE apud ENOKA, 2000).

O eletromiógrafo é um equipamento que registra a atividade elétrica presente no músculo em contração, a qual é decorrente da ativação neuromuscular em condições normais. O registro eletromiográfico é importante, pois permite observar o comportamento eletrofisiológico do músculo em diferentes condições fisiológicas, analisando-se a influência da temperatura corpórea, idade, sexo, esforço (intensidade do exercício, frequência com que se pratica o mesmo e quantidade de contrações em um determinado tempo) e do percentual de fibras musculares rápidas e lentas.

A eletromiografia é freqüentemente utilizada para o estudo do movimento humano. Através desta pesquisa foi possível observar que os profissionais da área de saúde (médicos, odontólogos, fisioterapeutas e profissionais esportivos), que se utilizam do exame e relatam que o mesmo oferece um bom diagnóstico das contrações musculares. Este estudo da função de músculos como estrutura viva baseia-se num princípio estabelecido há mais de 200 anos pelo engenheiro eletricitista Luigi Galvani, segundo o qual um músculo esquelético se estimulado eletricamente se contrai e, por outro lado, produz corrente elétrica quando contraído voluntariamente.

2.1 Aquisição do sinal eletromiográfico

O sinal EMG é adquirido por um eletromiógrafo que tipicamente está acoplado a um computador. O sinal EMG captado no corpo humano é um sinal analógico (um sinal contínuo no tempo), que então deve ser convertido para sinal digital (um sinal discreto, que é definido somente para certos intervalos de tempo), para poder ser registrado pelo computador. Para tanto, certos parâmetros devem ser ajustados na aquisição do sinal EMG, dependendo da tarefa e objetivos para posterior análise. Os principais parâmetros são: freqüência de amostragem, componente como eletrodo, amplificadores, filtro, conversor analógico/digital, além do equipamento de armazenagem dos dados (computador).

2.1.1 Freqüência de amostragem

O SENIAM (acrônimo para o consórcio europeu - Surface EMG for the Non-Invasive Assessment of Muscles) define amostragem como sendo a leitura de um valor do sinal em certo instante específico. A definição da freqüência de amostragem possui um papel crítico para a correta reprodução digital do sinal EMG analógico. Um sinal analógico, reproduzido digitalmente com uma baixa freqüência de amostragem, pode não conter todas as informações relevantes.

O teorema de Nyquist ou Shannon propõe que para a correta reconstrução digital do sinal EMG deve-se utilizar uma frequência de amostragem de, no mínimo, o dobro de sua maior frequência. O sinal EMG de superfície pode ter frequências de até cerca de *400 a 500 Hz* (a frequência máxima é afetada por fatores como tipo de unidade motora e contração, tamanho do eletrodo e distância entre eletrodos e músculos, entre outros fatores). Portanto, considera-se como frequência de amostragem mínima para o sinal EMG superficial da ordem de 1000 Hz ou mais (MARCHETTI; DUARTE apud DELSYS, 2006; HERMENS; FRERIKS; DISSELHORST-KLUG; RAU, 2006).

2.1.2 Eletrodos (Sensores)

Os eletrodos são dispositivos de entrada e saída de corrente em um sistema elétrico. O eletrodo estabelece a conexão entre o corpo e o sistema de aquisição, devendo ser colocado próximo o bastante do músculo para que este possa captar sua corrente iônica. A área da interface eletrodo-tecido é chamada de superfície de detecção, comportando-se como um filtro passa-baixa cujas características dependem do tipo de eletrodo e do eletrólito utilizado (MARCHETTI; DUARTE apud DE LUCA, 2006).

Existem diversos tipos de eletrodos, delineados para diferentes tipos de aquisição, tarefa, natureza da pesquisa e músculo específico. No caso dos músculos profundos ou pequenos, utilizam-se eletrodos de fio ou de agulha, pois esses possuem pequena área de detecção e são limitados nos estudos de unidades motoras. Para análise das unidades motoras, utilizam-se eletrodos de agulha, pois possuem menor área de detecção, embora sejam críticos em atividades de contração forçada ou por influência considerável de dor (MARCHETTI; DUARTE apud DE LUCA, 2006).

Muitos outros tipos de eletrodos têm sido desenvolvidos para diferentes propostas, como eletrodos de malha (array), utilizados para aquisição das características de propagação dos potenciais de ação das fibras musculares.

Entretanto, neste texto serão enfatizados apenas os eletrodos superficiais, por sua aplicabilidade ao problema em questão.

Os eletrodos de superfície são fixados na pele, constituindo um sistema sensor que capta a corrente dos estímulos musculares na pele através da interface pele-eletrodo. São geralmente compostos por um sistema Ag/AgCl associado a um gel condutor (eletrólito). Contudo, qualquer combinação metal/gel que permita reação eletrolítica pode servir (MARCHETTI; DUARTE apud DE LUCA, 2006).

O sinal EMG pode ser adquirido, não necessariamente por um simples eletrodo, mas pode resultar de uma combinação dos sinais advindos de vários detectores, podendo, deste modo, ser classificado como monopolar, bipolar e sistema multipolar (MARCHETTI; DUARTE apud DUCHENE; GOUBEL, 2006). A configuração monopolar é grandemente utilizada quando se adquire um sinal simples e associada a este, faz-se necessária a utilização de um eletrodo de referência, longe o bastante do eletrodo ativo para evitar perturbações no campo elétrico das vizinhanças do local de aquisição. A razão sinal/ruído torna-se pobre e a seletividade dramaticamente diminuída, portanto possuem sérias implicações, principalmente em mensurações de velocidade de propagação. Este tipo de configuração monopolar é freqüentemente utilizado em análises gerais do sinal EMG, quando o objetivo é comparar a morfologia interna e sinais externos ou mesmo, quando o músculo é muito pequeno ou estreito para configurações bipolares (MARCHETTI; DUARTE apud DUCHENE; GOUBEL, 2006).

Eletrodos com configuração bipolar são os mais utilizados em estudos que envolvem exercícios de contração voluntária, ou sob condições de estimulação elétrica. O principal interesse em tal configuração está relacionado aos benefícios de uma alta taxa de rejeição de modo comum, para um moderno sistema de amplificação diferencial. A detecção diferencial é empregada para eliminar potencialmente grandes ruídos das linhas de força, sendo sua premissa simples. O sinal é detectado em dois locais, onde a circuitaria eletrônica subtrai e então amplifica sua diferença. Como resultado, qualquer sinal que é comum a ambos os

locais de detecção será removido, e os sinais que são diferentes nos dois locais serão amplificados (MARCHETTI; DUARTE apud DE LUCA, 2006).

Normalmente são utilizados eletrodos superficiais passivos, isto é, que não possuem amplificação no próprio eletrodo. Este tipo de eletrodo apenas detecta o sinal EMG e o envia ao condicionador (amplificador associado ao filtro analógico). Entretanto, em atividades dinâmicas, onde se inclui o ruído advindo do movimento dos cabos, faz-se interessante o uso de eletrodos ativos, que realizam a amplificação do sinal detectado antes de ser enviado ao condicionador. Esses eletrodos possuem um pré-amplificador diferencial que subtrai e amplifica o sinal EMG, tornando, desta forma, o movimento dos cabos menos significativo.

Dois pontos principais devem ser considerados quanto à escolha do eletrodo, sendo o primeiro relacionado à distância entre as superfícies de detecção. Esta distância intereletrodos é definida como a distância de centro a centro entre as áreas condutivas dos mesmos, afetando a largura de banda das frequências e a amplitude do sinal EMG. Uma pequena distância altera a largura de banda para altas frequências e diminui a amplitude do sinal EMG. Por esta razão, a distância pode ser fixada para comparações quantitativas entre aquisições feitas dentro ou entre músculos. Preferivelmente, a superfície de detecção pode ser montada em uma plataforma fixa, não sendo necessário separar as superfícies de detecção por grandes espaços (MARCHETTI; DUARTE apud DELSYS, 2006). A recomendação da distância entre eletrodos, proposta pelo SENIAM (MARCHETTI; DUARTE apud HERMENS, 2006), é de 20 mm (de centro a centro).

O segundo ponto a ser considerado, é o tamanho e a forma das superfícies de detecção. Quanto maior o tamanho da superfície de detecção, maior a amplitude do sinal EMG detectado e menor o ruído elétrico que será gerado na interface entre a pele e a superfície de detecção, entretanto, este deve ser pequeno

o bastante para evitar o cross-talk¹ de outros músculos (MARCHETTI; DUARTE apud DELSYS; HERMENS, 2006).

Em se tratando da qualidade de aquisição do sinal EMG por parte dos eletrodos, faz-se necessário minimizar a influência da impedância pele/eletrodo. Deste modo, certos cuidados devem ser tomados, tais como limpeza da pele, remoção dos pêlos e leve abrasão para remoção de células mortas (MARCHETTI; DUARTE apud WINTER; HERMENS, 2006).

O local e posicionamento dos eletrodos podem causar interferência na qualidade do sinal EMG. Portanto, faz-se relevante a discussão de tais aspectos, considerados fundamentais:

- **Localização do eletrodo com relação ao ponto motor.** Define-se ponto motor como o local no músculo onde a introdução de mínima corrente elétrica causa um perceptível estímulo nas fibras musculares superficiais. Este ponto, usualmente, mas não sempre, corresponde a parte da zona de inervação em um músculo, possuindo grande densidade neural, dependendo da anisotropia do músculo nesta região. Sob o ponto de vista da estabilidade do sinal EMG, um ponto motor fornece um péssimo local para a detecção do sinal EMG (para eletrodos diferenciais), pois nesta região os potenciais de ação viajam em ambas as direções, assim, as fases positivas e negativas dos potenciais de ação podem ser subtraídos, cancelando-se. O SENIAM (MARCHETTI; DUARTE apud HERMENS, 2000) propõe que o eletrodo

¹ Interferência de um sinal por outros sinais que emanam de músculos adjacentes

seja colocado entre o ponto motor e o tendão distal² do músculo avaliado.

- **Formas de Interferência do sinal EMG.** Existem três formas principais de interferência do sinal EMG, que estão relacionadas ao batimento cardíaco, aquisição do sinal EMG de músculos vizinhos (cross-talk) e artefatos eletromecânicos (movimentos do equipamento/cabo e influência da rede elétrica).
- **Direção do eletrodo em relação às fibras musculares.** Como o potencial de ação possui trajetória no mesmo sentido das fibras musculares, o eletrodo deve ser alinhado em tal sentido para a melhor obtenção do sinal EMG (MARCHETTI; DUARTE apud HERMENS, 2006). Quando assim arranjadas, ambas as superfícies de detecção interacionam-se com a maioria dessas mesmas fibras musculares.
- **Eletrodo de referência.** Quanto ao eletrodo de referência, o SENIAM (MARCHETTI; DUARTE apud HERMENS, 2006) recomenda que sejam utilizadas, dependendo dos músculos analisados, as regiões do punho, tornozelo ou processo espinhal C7.

Vale ressaltar nesta pesquisa, que o “eletrodo de referência” introduzido pelos experimentos foi utilizado atrás da orelha direita, seguindo os casos realizados e validados em pesquisa apresentadas nos artigos sobre reconhecimento da fala subvocal ou subaudível, já citados no capítulo 1.

² Um dos componentes do músculo estriado esquelético

2.2 Fisiologia

Do ponto de vista fisiológico, durante uma contração muscular voluntária, a atividade gerada no córtex cerebral ativa o neurônio motor periférico localizado no tronco cerebral (núcleos de pares cranianos motores) ou na medula espinhal e este faz contato sináptico com os músculos, despolarizando a placa motora (local de sinapse neuromuscular), gerando a contração simultânea das várias fibras musculares que compõem uma unidade motora.

O registro eletromiográfico que o eletrodo vai detectar no interior do músculo é a soma dos potenciais de ação de todas as fibras musculares da unidade motora que entram conjuntamente em atividade porque são inervados pelo mesmo neurônio motor.

A eletromiografia é o estudo desses potenciais de ação, analisando suas diferentes características fisiológicas, assim como suas alterações patológicas.

2.3 Uso da eletromiografia nas diversas áreas de saúde

Apresentamos a seguir a importância da eletromiografia em cada área da saúde: Fisioterapia, Medicina, Educação Física, Fonoaudiologia, Odontologia, Terapia Ocupacional, tendo como objetivo explicar a importância desse estudo para fins de análise dos fenômenos bioelétricos.

2.3.1 Fisioterapia

Nesta área de saúde, a eletromiografia pode ser utilizada para os seguintes fins:

- Avaliação da capacidade dos exercícios em facilitar ou inibir a atividade muscular durante o procedimento terapêutico.

- Exame da função muscular; padrões de resposta muscular, início e fim da atividade e o nível de resposta muscular em relação ao esforço, tipo de contração muscular rápida ou lenta.
- Avaliação da atividade muscular durante a marcha, se a instrumentação permite um acompanhamento simultâneo e coordenado da atividade; por exemplo, filme, vídeo tape, eletrogoniômetros, ou interruptores de contato e de apoio.

2.3.2 Medicina

Nesta área de saúde, a eletromiografia pode ser utilizada para:

- Permitir um melhor diagnóstico diferencial entre uma contração fisiológica e uma patológica.
- Orientar também no estudo de patologias musculares de origem degenerativa e inflamatória e no reconhecimento de distúrbios do músculo associado a doenças metabólicas.

Vale ressaltar que na medicina o exame mais usado é o eletroneuromiograma que serve para diferenciar os distúrbios do nervo, da junção neuromuscular e do músculo, sendo que este não diagnostica qual o motivo da lesão, apenas mostra que ela existe.

2.3.3 Educação Física

Nesta área de saúde, a eletromiografia pode ser aplicada para os seguintes propósitos:

- Avaliar a eficácia de equipamentos de musculação, indicando as melhores posições para se trabalhar um determinado músculo;
- Comparar de acordo com um caminhar normal, se um indivíduo está ou não com uma falha mecânica na execução do movimento;
- Analisar o deslocamento da frequência de disparo das fibras musculares que, com o aparecimento da fadiga, disparam em frequências mais baixas.
- Diagnosticar qual é o ponto de fadiga da pessoa, para poder aproveitar o máximo da sua capacidade.
- Fazer o acompanhamento melhor da evolução de um atleta, verificando o crescimento do número de unidades motoras recrutadas.
- Pesquisar a atividade de outros músculos, que atuam em um determinado movimento, além do músculo principal.
- Em avaliações físicas auxiliar na análise de predominância de fibras fásicas ou tônicas. Dessa maneira, pode-se encaminhar o atleta para o esporte que ele, por sua condição genética, desempenhará melhor.

2.3.4 Fonoaudiologia

Nesta área de saúde, a eletromiografia pode ser utilizada para avaliar melhor as condições iniciais do paciente e a progressão do mesmo em relação aos músculos da expressão facial, deglutição, respiração e mastigação.

2.3.5 Odontologia

Nesta área de saúde, a eletromiografia pode ser utilizada nas seguintes tarefas:

- Na investigação da etiologia da dor em síndromes das articulações têmporo-mandibulares, quando se suspeita da existência de um tensionamento muscular. Ela capta a ativação do músculo, elaborando um gráfico dessa ação, registrando as alterações intrínsecas dos potenciais elétricos.
- Na avaliação, se os músculos da ATM (Articulação Temporomandibular) estão trabalhando no momento certo e com a devida intensidade.
- Na função de diagnosticar, o exame eletromiográfico funciona como bio-feedback, educando a pessoa para autocuidados³.

2.3.6 Terapia ocupacional

Nesta área de saúde, a eletromiografia pode auxiliar o(a) terapeuta ocupacional a trabalhar com músculos agonistas e antagonistas de forma dinâmica, para estudar os movimentos cognitivos. Ajuda tanto em uma avaliação como na verificação do resultado de um tratamento.

Os profissionais da área de saúde precisam usar sabiamente a eletromiografia, reconhecendo, contudo, suas limitações como instrumento de mensuração.

³ São os cuidados que devem ter para manter uma boa saúde bucal

A eletromiografia fornece informações durante o tratamento ou avaliação que podem aumentar a eficácia destes, atuando como um mecanismo de feedback em várias situações fisiológicas e patológicas (ARAUJO, 2007).

2.4 Conclusão

Este capítulo apresentou conceitos sobre eletromiografia, como é realizada a aquisição do sinal eletromiográfico (sinal EMG), enfocando a frequência de amostragem dos eletrodos de superfície ponto importante na aquisição dos sinais EMG e os tipos e aplicação dos eletrodos (sensores). Apresenta uma visão fisiológica da eletromiografia e o uso da mesma nas diversas áreas da saúde. O próximo capítulo aborda sobre o sistema de reconhecimento subvocal.

3 SISTEMA DE RECONHECIMENTO SUBVOCAL

O sistema de reconhecimento subvocal é composto da fala subvocal e do reconhecimento subvocal, que é o foco desta pesquisa.

3.1 Fala subvocal

Fala subvocal é caracterizado pelo movimento dos lábios ou de outros órgãos sem produzir sons audíveis. Falar sem movimentar a boca é chamado fala subvocal; pode-se fazer inconscientemente quando ler ou pensar.

“A fala subvocal é silenciosa, ou fala subaudível, tal como quando uma pessoa lê silenciosamente ou fala para ela mesma. Sinais biológicos chegam quando lemos ou falamos com ou sem o atual movimento labial ou facial. Uma pessoa usando o sistema subvocal pensa em frases e fala para ele mesmo tão silenciosamente, que não podemos ser ouvidos, mas a língua e as cordas vocais recebem os sinais da fala do cérebro” (JORGENSEN, 2004).

Fala subvocal é resultado do uso de sinais eletromiográficos (sinal EMG) captados nas superfícies da laringe e das áreas lingual e reproduzidos por programa de computador.

Fala subvocal, silenciosa ou fala subaudível é falar tal como quando uma pessoa lê ou fala silenciosamente para ela mesma, ou mesmo quando lendo e falando para si mesmo, com ou sem o movimento real dos lábios ou da face, onde os sinais biológicos surgem pelos movimentos dos músculos provocados pela articulação da boca sem produzir som.

Fala subvocal é também a interpretação direta “não audível” dos sinais dos sistemas nervosos enviados pelos músculos do trato vocal (i.e., eletromiografia ou sinais EMG). É medido através da superfície de contato dos sensores e os sinais elétricos que são transformados em padrões reconhecidos através dos classificadores como palavra ou componentes das palavras.

O som do silêncio (The sound of silence), a música que imortalizou a dupla Simon e Garfunkel, pode deixar de ser uma licença poética graças a um trabalho de cientistas da NASA. Eles descobriram que pequenos sensores, colocados sob o queixo e dos dois lados do pomo de Adão, podem captar sinais dos nervos e enviá-los para processamento por um programa especializado que consegue traduzi-los em palavras “O que é analisado é o silêncio, ou o som subaudível, tal como quando uma pessoa lê ou fala silenciosamente consigo mesma”, explica Chuck Jorgensen, um dos cientistas da equipe que fez a descoberta. “Sinais biológicos são gerados quando se lê ou fala consigo mesmo com ou sem movimento real dos lábios ou da face”.

3.2 Reconhecimento subvocal

O reconhecimento subvocal é a arte de falar de subvocalização e conversão dos resultados detectados para uma saída digital. É similar ao reconhecimento de voz, exceto pela subvocalização do silêncio que está sendo detectado.

Um conjunto de eletrodos é fixado sobre a pele abaixo da garganta e, sem abrir a boca ou expressar um som, as palavras são reconhecidas por um programa de computador.

O reconhecimento subvocal da fala trata dos eletromiogramas (sinais EMG), que são diferentes para cada pessoa. Conseqüentemente, a consistência do resultado pode ser perdida em função da posição de um eletrodo. Observa-se que

para melhorar a exatidão, alguns pesquisadores neste campo estão confiando nos modelos estatísticos que são obtidos através de suas pesquisas com os eletrodos.

Subvocalização, ou fala silenciosa, é definido como a fala interna feita ao ler uma palavra, assim permitindo que o leitor imagine o som da palavra como ele é lida. Este é um processo natural quando lido, pois ajuda reduzir a carga cognitiva, além de ajudar a mente a acessar significados que permitem a pessoa compreender e recordar o que é lido. Embora algumas pessoas associem a subvocalização com um movimento dos lábios, o termo real remete primeiramente ao movimento dos músculos associados com a fala, sem o movimento literal dos lábios. A maioria das subvocalizações são não detectável (sem auxílio de máquinas), mesmo pela pessoa que faz a subvocalização.

Não há evidência que a subvocalização normal, não-observável, afetará negativamente algum processo de leitura. Nas mais eficientes taxas (memorização, aprendizagem e leitura para compreensão), a subvocalização é detectável pelo leitor. Nas menos eficientes taxas de leitura, scanning e skimming, a subvocalização é menos detectável. Para leitores capacitados, a subvocalização, até certo ponto, mesmo para em taxas de scanning, é normal.

Pode ser impossível eliminar totalmente a subvocalização, porque as pessoas aprendem a ler, associando a visão das palavras com seus sons falados. Associações de sons para palavras são indelevelmente gravadas no sistema nervoso, até mesmo de uma pessoa surda, uma vez que ela terá as palavras associadas a um mecanismo para produzir um som ou uma determinada linguagem de sinais.

A subvocalização é uma parte inerente da leitura e da compreensão de uma palavra, e os testes do micro-músculo sugerem que subvocalização é impossível de eliminar. Tentar parar a subvocalização é potencialmente prejudicial à compreensão, à aprendizagem e à memorização. Nas variáveis de leitura mais eficientes (100-300 palavras por minuto), a subvocalização pode ser usada para

melhorar a compreensão. Constata-se, ainda, que a subvocalização envolve movimentos reais da língua e das cordas vocais que podem ser interpretados por sensores eletromagnéticos.

3.3 Conclusão

Este capítulo apresentou conceitos sobre a fala subvocal e como é realizado o reconhecimento da fala subvocal que compõem o sistema de reconhecimento subvocal, que é foco desta pesquisa. O próximo capítulo aborda sobre a análise de componentes independentes (ICA), uma técnica usada na extração de características básicas (funções bases).

4 ANÁLISE DE COMPONENTES INDEPENDENTES (ICA)

A Análise de Componentes Independentes (Independent Component Analysis - ICA) é um método que é visto como uma extensão da Análise de Componentes Principais (Principal Component Analysis - PCA). A ICA foi desenvolvida no contexto de separação cega de fontes (Blind Source Separation - BSS), em que o problema é definido na estimação da saída de uma fonte conhecida, quando esta fonte recebe vários sinais misturados e desconhecidos. A ICA tem sido aplicada em diversas áreas, como por exemplo: áudio, radar, instrumentação médica, comunicação móvel, engenharia biomédica e outras.

A ICA é utilizada em Separação Cega de Fontes (Blind Source Separation - BSS), porque consegue recuperar as fontes "não-observáveis" de uma mistura de diversas fontes. O termo *blind* refere-se ao fato de que existem fontes não observáveis no sinal e nenhuma ou pouca informação se tem sobre a mesma. Uma aplicação interessante de BSS é o problema *cocktail-party*, em que se separam as fontes originais de um sinal misturado, sem o conhecimento prévio dos coeficientes de mistura, nem a provável distribuição do sinal, usando apenas independência estatística como critério de separação de fontes (GIROLAMI; FYFE, 1997).

Jutten e Héroult (JUTTEN; HERAULT, 1991) desenvolveram o primeiro algoritmo de aprendizado para BSS. Bell e Sejnowski (BELL; SEJNOWSKI, 1995) desenvolveram uma rede neural capaz de aprender regras que minimizam a informação mútua dos nós de saída. Karhunen e Joutsensalo (KARHUNEN; JOUTSENSALO, 1997) propuseram algumas variações não-lineares de PCA, e demonstraram a utilidade destes algoritmos para estimação de frequência sinusoidais.

Separação Cega de Fontes representa um grande problema na engenharia, pois a técnica mais utilizada anteriormente era PCA, que utiliza apenas estatística de segunda ordem e é suficiente para descorrelacionar um conjunto de dados, mas não necessário para independência, que requer estatísticas de alta

ordem. Por esta razão, a ICA é vista como um método mais "robusto" que PCA, pois se PCA consegue descorrelacionar as fontes não observáveis, a ICA consegue deixá-los mútua e estatisticamente independentes entre si.

4.1 Definições

Considere que sejam observadas n misturas lineares x_1, \dots, x_n de n componentes independentes

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n \quad j = 1, \dots, n \quad (1)$$

e que cada mistura x_j , assim como cada componente independente s_k seja uma variável aleatória, e o a_j os coeficientes (pesos) da mistura linear.

Sem perda de generalidade, assume-se que tanto as variáveis da mistura quanto aquelas das componentes independentes têm média zero. Por conveniência, será usada a notação vetorial em vez de somas, como aquelas vistas na Equação (1), utiliza-se letras minúsculas e maiúsculas para representar, respectivamente, vetores e matrizes. Dessa maneira, pode-se reescrever a Equação (1) da seguinte forma:

$$x = As \quad (2)$$

em que $x \in \mathbf{R}^n$, $S \in \mathbf{R}^n$ e $A \in \mathbf{R}^n$.

O objetivo da técnica é recuperar as fontes s , através de x , sem nenhuma informação sobre as propriedades de A .

O modelo estatístico definido na Equação (2) é chamado de modelo de Análise de Componentes Independentes. Esse modelo descreve os dados observados pelo processo de mistura das componentes independentes s_i , que não

podem ser observadas diretamente. E preciso estimar tanto s quanto a matriz de mistura A , que também é desconhecida, pois tudo o que se observa é o vetor aleatório x .

O problema do modelo de dados de ICA é estimar a matriz A usando apenas a informação contida no vetor x . Para tanto, é preciso fazer suposições tão gerais quanto possível (HYVÄÄRINEN; KARHUNEN; OJA, 2001). Portanto, supõe-se que:

- a) As componentes s_i são estatisticamente independentes;
- b) As componentes têm distribuições não-gaussianas;
- c) Por motivos de simplicidade, a matriz A é quadrada.

4.2 Definição de independência

Sejam y_1 e y_2 duas variáveis aleatórias. Tais variáveis são ditas independentes se a ocorrência ou não ocorrência de y_1 não influenciar na ocorrência ou não ocorrência de y_2 , e vice-versa. Independência estatística é definida em termos de densidade de probabilidade. Seja $p(y_1, y_2)$ a função densidade de probabilidade (FDP) conjunta de y_1 e y_2 . Então, $p_1(y_1)$ denota a FDP de y_1 :

$$p_1(y_1) = \int p(y_1, y_2) dy_2 \quad (3)$$

e similarmente para y_2 . Duas variáveis aleatórias são estatisticamente independentes, se, e somente se, a FDP conjunta for dada por:

$$p_1(y_1, y_2) = p_1(y_1)p_2(y_2) \quad (4)$$

Pode-se definir decorrelação de duas variáveis aleatórias y_1 e y_2 , com

covariância zero, como:

$$E(y_1, y_2) = E(y_1)E(y_2) = 0 \quad (5)$$

Se duas variáveis são independentes, também são descorrelacionadas, mas o contrário não é verdadeiro.

4.3 Técnicas de estimação das componentes independentes

Sejam n misturas lineares x_1, \dots, x_n de n componentes independentes. Para estimar as componentes, deve-se encontrar a inversa da matriz A , que é chamada de W . A solução para a estimação das componentes independentes pode ser descrita da forma:

$$y = Wx = WAs \rightarrow DP_s \quad (6)$$

Sendo $D \in \mathbf{R}^{n \times n}$ uma matriz diagonal não-singular, e $P \in \mathbf{R}^{n \times n}$ uma matriz de permutação qualquer. Observa-se imediatamente que y_i é uma combinação linear de s_i . Baseado no teorema central do limite, y_i é mais gaussiano do que qualquer s_i e torna-se menos gaussiano, quando de fato é igual a uma componente de s .

Para finalmente estimar as componentes independentes, deve-se encontrar a matriz W que minimiza a não-gaussianidade de Wx .

A não-gaussianidade é um elemento chave para a estimação do modelo de ICA, pois a matrix A não é idêntica quando as componentes independentes têm distribuição gaussiana. Considere que o vetor x é distribuído de acordo com o modelo de ICA na equação (1), e que todas as componentes independentes têm distribuições iguais. Para estimar as componentes independentes, basta encontrar as combinações lineares corretas das variáveis da mistura x_i , de modo que:

$$s = A^{-1}x \quad (7)$$

Assim, pode-se expressar uma combinação linear de x_i por

$$y = b^T x \quad (8a)$$

$$= \sum_i b_i x_i \quad (8b)$$

$$= b^T A s \quad (8c)$$

em que b deve ser determinado. A partir da equação (8c) pode-se observar que y é uma combinação linear de s_i , com coeficientes dados por $q = b^T A$. Logo, obtém-se

$$y = q^T s \quad (9a)$$

$$= \sum_i q_i s_i \quad (9b)$$

Se b corresponder a uma das linhas da inversa de A , então y será uma das componentes independentes e, nesse caso, apenas um dos elementos de q será igual a 1, enquanto todos os outros serão iguais a zero. Não é possível determinar b exatamente, mas pode-se estimar seu valor com boa aproximação.

Uma forma de determinar b é variar os coeficientes em q e então verificar como a distribuição de $y = q^T s$ muda. Já que, conforme o Teorema Central do Limite (PAPOULIS, 2002), a soma de duas variáveis aleatórias independentes é mais gaussiana que as variáveis originais, $y = q^T s$ normalmente é mais gaussiana que qualquer uma das s_i e menos gaussiana quando se iguala a uma das s_i . Nesse caso, apenas um dos elementos q_i de q é diferente de zero (HYVÄÄRINEN; KARHUNEN; OJA, 2001).

Como, na prática, os valores de q são desconhecidos e sabe-se que, através das Equações (8a) e (9a), temos:

$$b^T x = q^T s \quad (10)$$

Pode-se variar b e observar a distribuição de $b^T x$. Portanto, pode-se tomar, como b , um vetor que maximiza a não-gaussianidade de $b^T x$, sendo que esse vetor necessariamente corresponde a $q=A^T s$, vetor esse que possui apenas uma de suas componentes diferente de zero. Isso significa que y na Equação (8a) é igual a uma das componentes independentes. Logo, a maximização da não-gaussianidade de $b^T x$ permite encontrar uma das componentes.

4.4 Negentropia como medida de não-gaussianidade

Negentropia pode ser descrita como um conceito que se baseia na quantidade de diferença de entropia. A entropia de uma variável aleatória pode ser considerada como um grau de informação da observação de uma variável dada. Devido à complexidade ou aleatoriedade das variáveis, que podem envolver um processo, mais difícil será a sua predição e conseqüentemente maior será a sua entropia. Em uma visão mais ampla, a entropia é relatada como tamanho da codificação da variável aleatória, que para uma variável pode ser escrita na forma:

$$H(y) = -\sum P(y = a_i) \log P(y = a_i) \quad (11)$$

Em que a_i são os possíveis valores de y , P a função densidade de probabilidade e \log uma função logarítmica. Também se pode expressar a entropia de variáveis aleatórias para a Equação (12) para valores contínuos da seguinte forma:

$$H(y) = -\int f(y) \log f(y) dy \quad (12)$$

Como um dos resultados fundamentais da Teoria da Informação, sabe-se que uma variável gaussiana tem a maior entropia entre todas as variáveis aleatórias de igual variância (HYVÄÄRINEN; KARHUNEN; OJA, 2001); (PAPOULIS, 2002). Isso quer dizer que uma versão modificada da entropia diferencial pode ser usada como medida de não-gaussianidade.

Essa medida é chamada negentropia, sendo definida por:

$$J(y) = H(y_{gauss}) - H(y) \quad (13)$$

Em que y_{gauss} é a variável aleatória gaussiana de matriz de correlação (e covariância) de y . Devido às propriedades discutidas acima, a negentropia será sempre não-negativa e com média zero, se a distribuição de y for gaussiana.

Em contraste às suas qualidades como medida de não-gaussianidade, a negentropia é de difícil estimação. Por isso, é necessária a utilização de aproximações usando, por exemplo, momentos de alta ordem. Logo:

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} kurt(y)^2 \quad (14)$$

sendo $kurt(y)$, a curtose de y , é definida como o momento de quarta ordem da variável aleatória y , expresso por :

$$kurt(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (15)$$

No entanto, essa aproximação usa a curtose, que é uma medida de não-gaussianidade. Dessa forma, é mais conveniente utilizar outras abordagens, que substituem os momentos polinomiais y_3 e y_4 por outra função G . O método propõe a aproximação da negentropia, baseado em expectâncias (E). (HYVÄÄRINEN; KARHUNEN; OJA, 2001).

$$J(y) = k_1(E\{G_1(y)\}) + k_2(E\{G_2(y)\} - E\{G_2(v)\})^2 \quad (16)$$

sendo k_1 e k_2 constantes positivas, v uma variável gaussiana de média zero, e $G_1\{y\} = y^3$ e $G_2\{y\} = y^4$.

4.5 Seleção de características mais significantes

A técnica *Forward-Selection* é utilizada para encontrar a melhor combinação de características, iniciando com uma única característica, e acrescentando mais características, passo a passo. Dessa maneira, cada característica é adicionada no modelo de cada vez, e a característica que ainda não pertence ao modelo é testada para ser incluída. As características mais significativas, ou seja, que mostram um maior decréscimo na função de erro, que são adicionados ao modelo, até que se consiga um subconjunto f menor que o conjunto F selecionado.

Nesta técnica, cada etapa envolve o crescimento do subconjunto de funções bases, pela adição de uma função base (ou seja, uma característica). Adicionar uma função base nova é uma operação incremental. A equação fundamental da técnica é:

$$P_{m+1} = P_m - \frac{P_m \cdot f_j \cdot f_j^T \cdot P_m}{f_j^T \cdot P_m \cdot f_j} \quad (17)$$

que expressa a relação entre P_m , a matriz de projeção de m camadas escondidas do subconjunto corrente e P_{m+1} , a projeção sucedente do j -ésimo membro do último conjunto de características adicionado. Os vetores $\{f_j\}_{j=1}^M$ são colunas de um conjunto selecionado de funções-bases

$$F = [f_1, f_2, f_M] \quad (18)$$

sendo $M \gg m$.

A escolha das funções bases é baseada na minimização do *erro-médio-quadrático* (*MSE*). E das regras de atualização para a matriz de projeção e da equação para o erro-médio-quadrático, obtém-se:

$$T_m - T_{m+1} = \frac{(y^T \cdot P_m \cdot f_j)^2}{f_j^T \cdot P_m f_j} \quad (19)$$

sendo T_m o conjunto das funções anterior e T_{m+1} o conjunto das funções bases atual.

4.6 Conclusão

Este capítulo apresentou um histórico e a definição da Análise de Componentes Independentes, usada na presente pesquisa para o uso da extração das funções bases das vogais subvocais, as quais foram utilizadas na fase de aprendizagem dos dados. O próximo capítulo aborda sobre redes neurais artificiais que foram usadas na fase de classificação dos dados.

5 REDES NEURAIS ARTIFICIAIS

Redes Neurais Artificiais (RNA's) são modelos computacionais com propriedades particulares, tais como: habilidade de aprender, de generalizar, de agrupar e de organizar dados. RNA's podem ser utilizadas para classificação ou de reconhecimento de padrões. Nesta pesquisa, utiliza-se RNA's para classificação da fala subvocal (vogais subvocais), a qual é obtida através de um sistema de aquisição de dados e realizada sobre os sinais da fala subvocal digitalizadas.

Redes Neurais Artificiais (RNA's) foram desenvolvidas, originalmente, na década de 40, pelo neurofisiologista Warren McCulloch e pelo matemático Walter Pitts que, dentro do espírito cibernético, fizeram uma analogia entre células nervosas vivas e o processo eletrônico num trabalho publicado sobre "neurônios formais". O trabalho consistia num modelo de resistores variáveis e amplificadores, representando conexões sinápticas de um neurônio biológico (TAFNER, 2007)

Na década 80, diversos modelos de redes neurais artificiais surgiram com o propósito de solucionar problemas relacionados a diferentes contextos e foram aplicadas em reconhecimento de padrões, por exemplo, reconhecimento de voz e outras áreas, tais como, análise e processamento de sinais, robótica, o uso de robôs em suas operações críticas (e.g. desarmem de bombas, etc.).

A Figura 4 apresenta a arquitetura *Feedforward* de uma RNA de uma camada escondida e duas camadas de saídas.

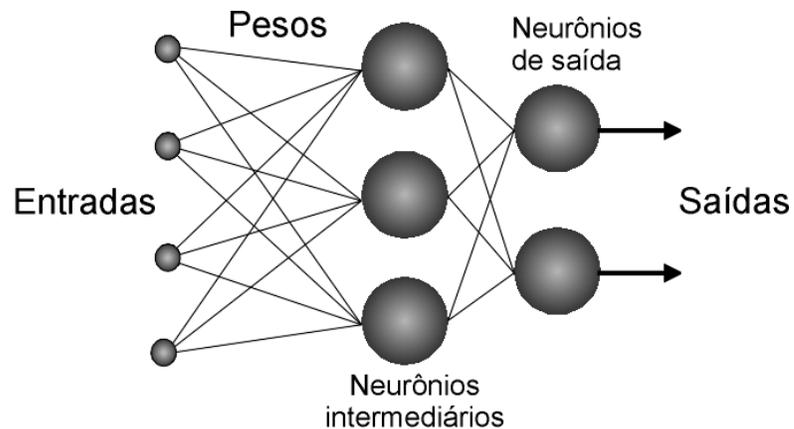


Figura 4: Arquitetura de RNA com duas saídas.
Fonte: TAFNER, 2007

Existem diversos modelos de redes neurais para solucionar instâncias específicas de problemas, onde a arquitetura destas redes pode sofrer modificações, dependendo de sua necessidade de aplicação. Basicamente, os itens que compõem uma rede neural são os seguintes:

- Conexões entre camadas (pesos sinápticos e limiares);
- Camadas intermediárias;
- Quantidade de neurônios em cada camada;
- Função de transferência de cada neurônio artificial;
- Algoritmo de aprendizado para modificar as conexões.

Todas as RNA's possuem alguma regra de treinamento, em que os pesos de suas conexões são ajustados de acordo com os padrões apresentados, de tal forma que se pode dizer que sua principal propriedade é a capacidade de aprendizado. Este processo é realizado através de interações de ajustes aplicados aos pesos das conexões ao qual se denomina treinamento. O aprendizado ocorre

quando a RNA atinge uma solução generalizada para uma classe específica de problemas.

Um conjunto de regras bem definidas para que a RNA possa aprender a solução de um problema é chamado de algoritmo de aprendizado. Existem muitos algoritmos de aprendizado aplicados em determinados modelos de redes neurais, estes algoritmos diferem entre si principalmente pelo modo como os pesos são modificados. Outra propriedade importante é a categorização das situações de aprendizado das RNA's. Neste contexto, podem-se citar as seguintes classes de aprendizado (BRAGA, 2007).

- **Aprendizado Supervisionado:** A rede neural recebe um conjunto de entradas padronizadas e seus correspondentes padrões de saída, onde ocorrem ajustes nos pesos sinápticos até que os erros entre os padrões de saída gerados pela rede atinjam um valor próximo ao desejado;
- **Aprendizado Não-Supervisionado:** A rede neural trabalha os dados de forma a extrair algumas propriedades estatísticas dos conjuntos de dados. A partir destas propriedades é que o aprendizado é constituído.

Nesta pesquisa, optou-se pela rede neural perceptron multicamadas (MLP), por sua capacidade de trabalhar com problemas não-linearmente separáveis, com utilização de, pelo menos uma camada intermediária entre a camada de entrada e de saída. Estas camadas intermediárias implementam uma extração de características, que ficam armazenadas nos pesos sinápticos. Nesta rede neural o algoritmo de treinamento mais utilizado é o *Backpropagation*, que é um tipo de algoritmo de Aprendizado Supervisionado por correção de erro.

5.1 Redes neurais perceptron multicamadas (MLP)

Redes Perceptron Multicamadas (MLP's) são as redes neurais mais freqüentemente usadas em reconhecimento de padrões (DUDA; HART; STORK, 2000); (BISHOP, 2007). Uma RNA do tipo MLP é constituída por um conjunto de nós fontes, os quais formam a camada da entrada da rede, uma ou mais camadas escondidas, e uma camada de saída, que extraem informações durante o aprendizado, distribuindo entre os pesos das várias camadas de entrada.

O número de nós fontes na entrada da rede é determinado pela dimensionalidade do espaço de observação, que é responsável pela geração dos sinais de entrada. O número de neurônios na camada de saída é determinado pela dimensionalidade requerida na resposta desejada. Assim, o projeto de uma rede MLP deve levar em consideração três aspectos importantes, que são:

- I) A determinação do número de camadas escondidas;
- II) A determinação do número de neurônios em cada uma das camadas escondidas;
- III) A especificação dos pesos sinápticos que interconectam os neurônios nas diferentes camadas da rede.

Os aspectos I e II determinam a complexidade do modelo da RNA escolhida. A função das camadas escondidas em uma RNA é a de influir na relação entrada-saída da rede, de forma a permitir a contribuição de superfícies de decisão não-lineares. Uma RNA com uma ou mais camadas escondidas é apta a extrair as estatísticas de alta ordem de algum desconhecido processo computacional subjacente, responsável pelo comportamento dos dados de entrada-saída, processo sobre o qual a rede está tentando adquirir conhecimento. A RNA adquire uma perspectiva global do processo, apesar de sua conectividade local, em virtude do

conjunto adicional de pesos sinápticos e da dimensão adicional de interações neurais proporcionada pelas camadas escondidas.

O aspecto III envolve a utilização de algoritmos de treinamento supervisionado. Redes MLP's têm sido aplicadas na solução de diversos e difíceis problemas através da utilização de tais algoritmos. O algoritmo de treinamento geralmente utilizado é o algoritmo de *retropropagação*, popularmente conhecido como *error backpropagation*. O algoritmo *retropropagação* baseia-se na metáfora do aprendizado por correção de erro.

Este algoritmo pode ser visto com o uma generalização do algoritmo LMS (Least Mean Square), desenvolvido por Bernard Widrow (WIDROW; STEARNS 1985). Basicamente, o algoritmo *backpropagation* consiste de dois passos através das diferentes camadas do MLP: um passo direto e um passo reverso.

- No passo direto, um padrão de atividade do processo a ser aprendido (vetor de entrada) é aplicado aos nós de entrada do MLP e o seu efeito se propaga através da rede, camada por camada, produzindo na camada de saída a resposta do MLP à excitação aplicada. Durante o passo direto os pesos sinápticos são mantidos todos fixos.
- No passo reverso, os pesos sinápticos são todos ajustados de acordo com a regra de aprendizado por correção de erro. Especificamente, a resposta do MLP à excitação é subtraída de um padrão de resposta desejado para aquela excitação aplicada, de forma a produzir um sinal de erro, de forma semelhante ao algoritmo LMS.

A Figura 5 mostra Ilustração das direções de dois fluxos básicos de um sinal em uma rede neural MLP: propagação direta dos sinais e a retropropagação dos sinais de erro.

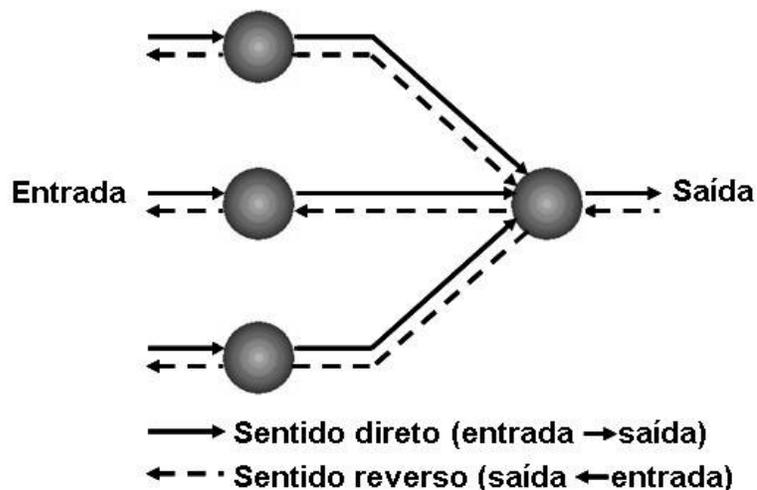


Figura 5: Ilustração das direções de dois fluxos básicos de um sinal em uma rede neural MLP:
 Sentido direto e Sentido reverso
 Fonte: Próprio autor, São Luís, 2007

Este sinal de erro é, então, propagado de volta aos mesmos neurônios utilizados no passo direto, mas no caminho contrário do fluxo de sinal nas conexões sinápticas, daí o nome *Backpropagation*. Os pesos sinápticos são ajustados de forma que a resposta obtida da MLP aproxime-se mais do padrão de resposta desejado, de tal forma que

$$e_j(n) = d_j(n) - y_j(n) \quad (20)$$

em que e_j é o sinal de erro, d_j o sinal desejado e y_j o sinal direto.

O modelo de cada neurônio MLP inclui uma função de ativação não-linear. É importante salientar que esta não linearidade é suave, ao contrário da função utilizada no modelo Perceptron de Rosenblatt (função degrau). Uma forma comumente utilizada de não-linearidade que satisfaz este requisito é a não-linearidade do tipo sigmóide definida pela função logística:

$$y_j = \frac{1}{1 + \exp(-v_j)} \quad (21)$$

sendo v_j o potencial de ativação (isto é, a soma ponderada de todas as entradas sinápticas acrescida do *bias* (polarização), do neurônio j , e y_j é a saída do neurônio. A presença de não-linearidade é importante porque, do contrário, a relação de entrada-saída da rede poderia ser reduzida àquela de um perceptron de camada única. Além disso, a utilização da função logística tem motivação biológica, pois procura levar em conta a fase refratária de neurônios reais (HAYKIN, 2001).

Durante o treinamento, as MLP's constroem um espaço multidimensional definido pela ativação dos nós das camadas escondidas, de modo que as classes sejam mais separáveis possíveis. A superfície de decisão se adapta aos dados.

5.2 Conclusão

Este capítulo apresentou definição sobre Redes Neurais Artificiais, seus tipos de aprendizados e também abordou sobre Rede Neural Perceptron Multicamadas (MLP), usada na presente pesquisa para classificação dos dados. O próximo capítulo apresenta o modelo proposto para resolver a instância da problematização desta pesquisa.

6 MODELO PROPOSTO

A Figura 6 mostra o sistema de reconhecimento da fala subvocal proposto, que é utilizado para solucionar a instância da problematização desta pesquisa, como sendo um sistema robusto e testado com as pessoas que participaram da composição dos dados, independente do sexo.

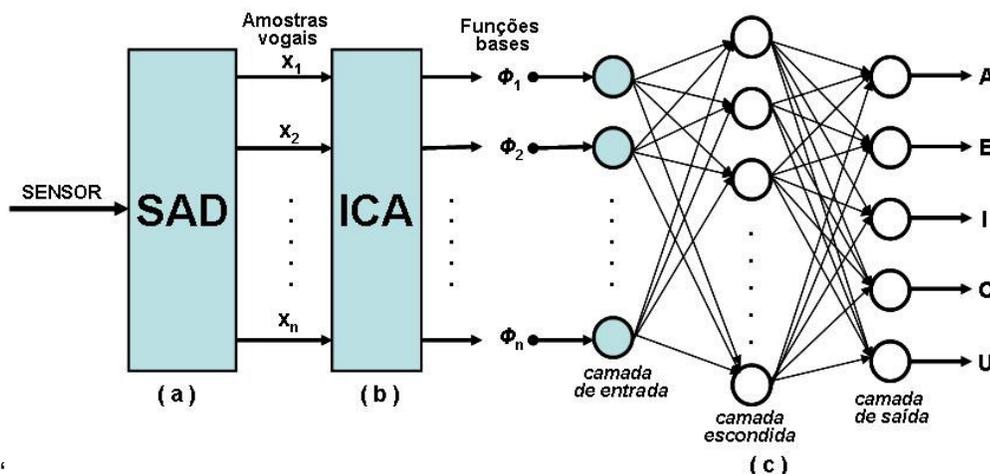


Figura 6: Sistema de reconhecimento da fala subvocal proposto: (a) - Fase de Aquisição de Dados. (b) - Fase de Aprendizagem dos Dados (c) - Fase de Classificação dos Dados
Fonte: Próprio autor, São Luís, 2007

O sistema consiste de 3 (três) fases: aquisição, aprendizagem e classificação dos dados. Na fase de aquisição dos dados, usou-se um sistema de aquisição de dados (SAD) e scripts Matlab[®]. Na fase de aprendizagem dos dados, usou-se a ICA para extração das características básicas (funções bases), e na fase de classificação dos dados, usou-se uma Rede MLP.

6.1 Fase de aquisição de dados

Nesta fase, usou-se um sistema de aquisição de dados (SAD), composto de uma placa de circuito eletrônico (interface EMG), sensores (eletrodos de superfície) e scripts Matlab[®] para compor o banco de dados das vogais subvocais.

6.1.1 Placa de aquisição de dados

A placa da aquisição de dados é composta de um circuito de amplificação e de filtragem conectados a um microcontrolador PIC. A comunicação com o computador é estabelecida através da sua porta serial COM1.

A Figura 7 mostra o circuito amplificação e filtragem e também mostra os 2 (dois) canais (CH1 e CH2) usados na aquisição dos sinais subvocais.

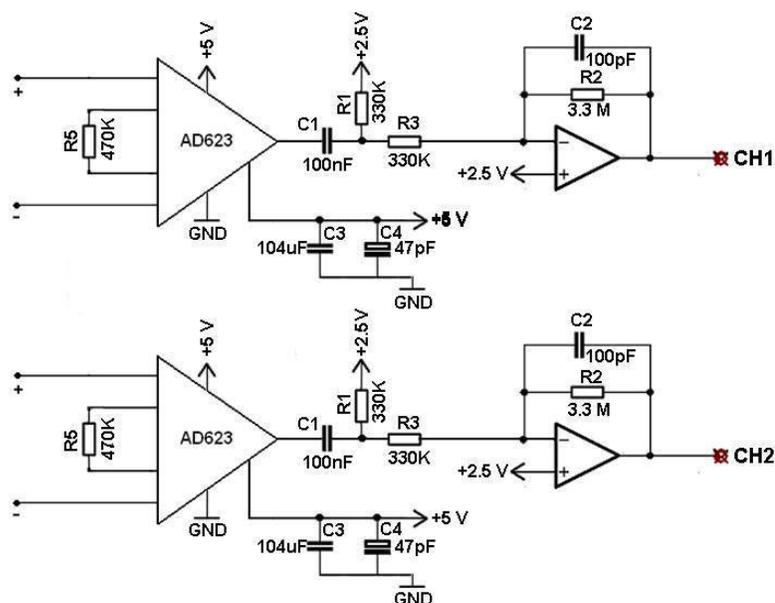


Figura 7: Circuito de Amplificação e Filtragem (Canal 1 (CH1), Canal 2(CH2)).
Fonte: EAGLE Layout Editor, 2006

O circuito foi inspirado em trabalhos similares como (SEGURA-JUÁREZ, 2004); (MARTIN; GLEESON, 2000) e é baseado no amplificador de instrumentação AD623 dos dispositivos análogos. Este amplificador de instrumentação tem baixo custo e baixo consumo como características chaves. O modelo AD623⁴ tem CMRR⁵

⁴ Datasheet deste amplificador de instrumentação AD623 está disponível no seguinte site:
http://www.analog.com/UploadedFiles/Data_Sheets/AD623.pdf, domínio público.

⁵ Common Mode Rejection Ratio

maior que 100 dB dentro do range de 1 até 60 hertz. O ganho da amplificação G_1 é dado pela equação (22) e foi ajustado a 10, para evitar o efeito da saturação causado pelo nível DC da entrada que é introduzida pelos eletrodos. O nível DC tem um valor baixo de 300mV.

$$R_1 + R_2 = 100K(G_1 - 1) \quad (22)$$

A Filtragem é executada por um filtro passa-alta de segunda ordem. A freqüência de corte do filtro é ao redor de 60 hertz, que é o padrão da freqüência da linha de energia. A filtragem passa-alta visa reduzir a variação no nível DC da entrada e é dado pela equação (23), ou seja,

$$f_1 = \frac{1}{2\pi R_5 C_3} \quad (23)$$

Logo após, um filtro passa-baixa de primeira ordem é utilizado para restringir o conteúdo espectral à freqüência de nyquist (equação 24).

$$f_2 = \frac{1}{2\pi R_3 C_1} \quad (24)$$

O ganho do filtro G_2 é dado pela equação (25) e é ajustado para 2K, totalizando um ganho de amplificação de 20K para os 2 (dois) canais 1 e 2 (CH1, CH2).

$$G_2 = \frac{R_2}{R_3} \quad (25)$$

O amplificador operacional (AMPOP) utilizado nestes filtros é o OPA4344⁶ da Burr-Brow⁷, que tem uma saída rail-to-rail, além do baixo custo, baixo ruído e baixo consumo.

O microcontrolador é o bloco subsequente, imediatamente depois do amplificador operacional. Ele compreende todas as funções restantes da placa de aquisição com conversão A/D e transmissão do sinal através do IrDA (STUART, 1995).

A Figura 8 mostra o esquema do bloco subsequente, composto de um microcontrolador da Microchip Technology Inc⁸. O modelo do microcontrolador utilizado é PIC16873P⁹. Este microcontrolador está funcionando em sua configuração padrão em associação com um dispositivo conectado em uma porta serial de comunicação.

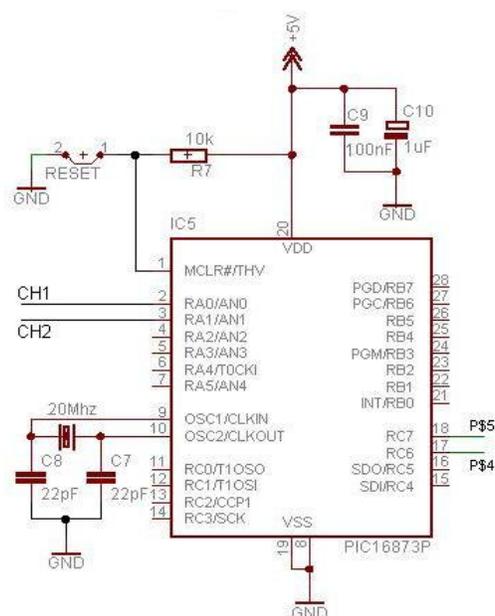


Figura 8: Microcontrolador (PIC 16873P)
Fonte: EAGLE Layout Editor, 2006

⁶ Datasheet deste amplificador operacional OPA4344 está disponível no site:
<http://focus.ti.com/lit/ds/symlink/opa4344.pdf>, domínio público.

⁷ www.burr-brown.com

⁸ www.microchip.com

⁹ Datasheet deste microcontrolador PIC está disponível no seguinte site:
<http://ww1.microchip.com/downloads/en/DeviceDoc/30275a.pdf>, domínio público.

O microcontrolador tem como características principais um conversor A/D de 12 bits, baixo consumo e baixo custo.

O programa do microcontrolador foi escrito no padrão da linguagem C e trabalha usando somente duas interrupções de hardware. Estas interrupções são fornecidas pelo microcontrolador. A primeira é a recepção serial de algum valor na porta USART e a segunda é o overflow do contador Timer0.

O projeto da placa de circuito eletrônico (interface EMG) usada na aquisição dos dados (vogais subvocais) foi desenvolvido no Laboratório de Processamento de Informação Biológica (PIB), na Universidade Federal do Maranhão (UFMA). A confecção da placa foi de baixo custo e possui uma arquitetura simples. Ao contrário de outras pesquisas já citadas que usaram uma estrutura muito mais complexa e cara, utilizando eletromiógrafos e softwares proprietários.

6.1.2 Eletrodos (Sensores)

Nesta pesquisa, foram utilizados os eletrodos (fabricante SKINTACT PREMIER Ref.: F-521) de superfície aderidos à pele como sensores, constituindo uma superfície de detecção que capta a corrente elétrica (sinais EMG) na pele, através da interface pele-eletrodo. Eles geralmente são compostos por um sistema *Prata/Cloreto de Prata (Ag/AgCl)* associado a um gel condutor (eletrólito) uma combinação metal/gel que permita uma reação eletrolítica.

6.2 Fase de aprendizagem de dados

Nesta fase, aprendemos um subespaço através de ICA. Vamos ter $x = [x_1, x_2, \dots, x_n]^T$ sendo um conjunto de observações obtidas de uma mesma classe de dados (vogais subvocais) e escrita na Equação (26). Usando x como a entrada para treinamento (amostras de vogais subvocais), a ICA aprende as funções bases φ_i para a classe de dados, de modo que o conjunto de variáveis que compõe o vetor $a = [a_1, a_2, \dots, a_n]^T$ é mutuamente independente estatisticamente falando, ou seja,

$$x = a^T \varphi \quad (26)$$

Para conseguir a independência estatística, algoritmos de ICA trabalham com uma estatística de uma ordem mais alta que indicam direções, onde o dado é maximamente independente. Aqui, nós usamos o algoritmo *FastICA* (HYVÄÄRINEN; KARHUNEN; OJA, 2001).

6.3 Fase de classificação dos dados

A última fase é a classificação dos dados, onde uma rede MLP (HAYKIN, 2001), foi usada para classificação. A rede MLP, uma rede neural feedforward treinada pelo algoritmo *backpropagation* é a mais freqüente técnica de rede neural usada em reconhecimento de padrões (MARTIN, 2000).

MLP's são classificadores de aprendizagem supervisionados que consistem de uma camada de entrada, uma camada de saída e uma ou mais camadas escondidas, que extrai a informação útil durante a aprendizagem e atribuem coeficientes de peso modificáveis para componentes das camadas de entrada. Cada saída é comparada com o valor desejado. Um sinal do erro é então propagado para trás e os pesos de conexão são ajustados correspondentemente. Durante o treinamento, redes MLP constroem um espaço multidimensional, definido pela ativação dos nós escondidos, de modo que os dados sejam tão separáveis quanto possível neste espaço.

Cada saída é comparada com o valor desejado. Um sinal do erro é então propagado para trás e os pesos de conexão são ajustados correspondentemente. Durante o treinamento, as MLP's constroem um espaço multidimensional, definido pela ativação dos nós escondidos, de modo que os dados sejam tão separáveis quanto possível.

A Figura 9 mostra um esquema ilustrativo do processo de aprendizagem numa RNA.

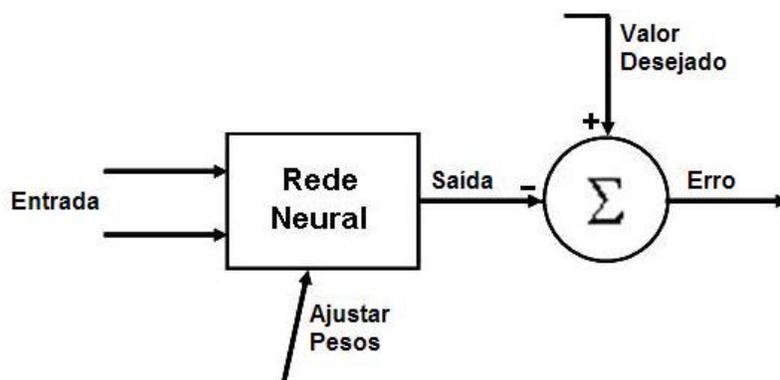


Figura 9: Esquema ilustrativo do processo de aprendizagem numa RNA
Fonte: Próprio autor, São Luís, 2007

Os dados são repetidamente apresentados à rede neural. A cada apresentação de um vetor de entrada, o erro entre a saída da rede e a saída desejada é computada e alimentado de volta para a rede neural. A rede neural usa este erro para ajustar seus pesos tal que o erro tenderá a diminuir. Esta seqüência de eventos é usualmente repetida até que um erro aceitável tenha sido alcançado ou até a rede não mais aprenda.

6.4 Medidas de avaliação de um sistema diagnóstico

A Sensibilidade (*Sens*), especificidade (*Spec*) e acurácia (*Accu*) são algumas das medidas mais usadas para descrever um sistema de diagnóstico. A Sensibilidade é a proporção de verdadeiro-positivos (VP) que é corretamente identificada pelo teste e é definida por $Sens = VP / (VP + FN)$. A Especificidade é a proporção de verdadeiro-negativos (VN) que corretamente identificada pelo teste e é definida por $Spec = VN / (VN + FP)$. A Acurácia (*Accu*) é uma das medidas para caracterizar a precisão de uma grandeza medida e é definida por $Accu = (TN + TP) / (TP + FP + FN + TN)$, onde FN é falso-negativo, FP é falso-positivo, FN é falso-negativo, VN é verdadeiro-negativo e VP é verdadeiro-positivo.

6.5 Conclusão

Este capítulo apresentou o modelo proposto, com uma metodologia de resolução da instância da problematização desta pesquisa, trazendo um sistema de reconhecimento da fala subvocal composto das seguintes fases: aquisição, aprendizagem fase de classificação dos dados e as medidas avaliação de um sistema diagnóstico: sensibilidade, especificidade e acurácia, usadas para avaliar os resultados desta pesquisa. O próximo capítulo apresenta os resultados e discussão desta pesquisa.

7 RESULTADOS E DISCUSSÃO

Neste capítulo estão descritos os resultados obtidos, usando o modelo proposto, o qual representa o reconhecimento da fala subvocal para os fonemas das vogais orais abertas.

7.1 Resultados

7.1.1 Dados capturados na fase de aquisição

O banco de dados usado nesta pesquisa é das vogais subvocais que foram armazenadas na fase de aquisição de dados através do uso de um sistema de aquisição de dados e de scripts Matlab[®]. Este banco de dados é composto de 150 vogais subvocais [a], 150 vogais subvocais [ɛ], 150 vogais subvocais [i], 150 vogais subvocais [ɔ] e 150 vogais subvocais [u], tal que o banco de dados das vogais subvocais possui 750 amostras para cada pessoa. Todas essas amostras foram armazenadas em um banco de vogais subvocais. Deste banco, foram selecionadas 50 vogais subvocais [a] 50 vogais subvocais [ɛ], 50 vogais subvocais [i], 50 vogais subvocais [ɔ] e vogais subvocais [u] de cada grupo de 750 vogais subvocais. Para cada grupo subvocal, uma amostra foi selecionada, contendo os sinais subvocais para extrair as funções bases, usando ICA. Estas novas amostras foram armazenadas em um banco de dados de características básicas (funções bases). Essas novas amostras foram usadas como entrada na rede neural MLP, para serem classificadas.

7.2 Dados subvocal

As amostras para realização desta pesquisa foram compostas por 10 (dez) pessoas (7 homens e 3 mulheres) com idade entre 18 e 48 anos. Para a aquisição dos sinais EMG (vogais subvocais) utilizou-se um sistema de aquisição de dados, uma janela de aquisição de 2 segundos e um simples canal de aquisição de

dados EMG, coletadas via eletrodos de superfície com numa frequência de amostragem de 1221 Hz, com presença de ruído no ambiente de coleta. Cada pessoa pronunciou subvocalmente 150 vezes as 5 (cinco) vogais orais abertas da língua portuguesa brasileira, totalizando 750 sinais EMG (vogais subvocais) por pessoa e assim compondo um total geral de 7500 amostras armazenadas num banco de dados das vogais subvocais.

Os sinais EMG foram coletados de cada pessoa, através do uso de 3 (três) eletrodos auto-adesivo (eletrodos de superfície) Ag/Ag-CL. Um par de eletrodos foi colocado do lado de direito e esquerdo da laringe na região do pescoço, numa distância de 5 cm entre eles. O terceiro eletrodo adicional (eletrodo de referência), funcionando como sinal de terra, foi colocado atrás da orelha direita de cada participante do processo de amostragem dos dados, diferente dos outros trabalhos, onde o eletrodo de referência foi colocado no pulso.

A Figura 10 traz o diagrama do sistema subvocal, mostrando o posicionamento dos eletrodos e a gravação dos sinais EMG (vogais subvocais) de uma pessoa.

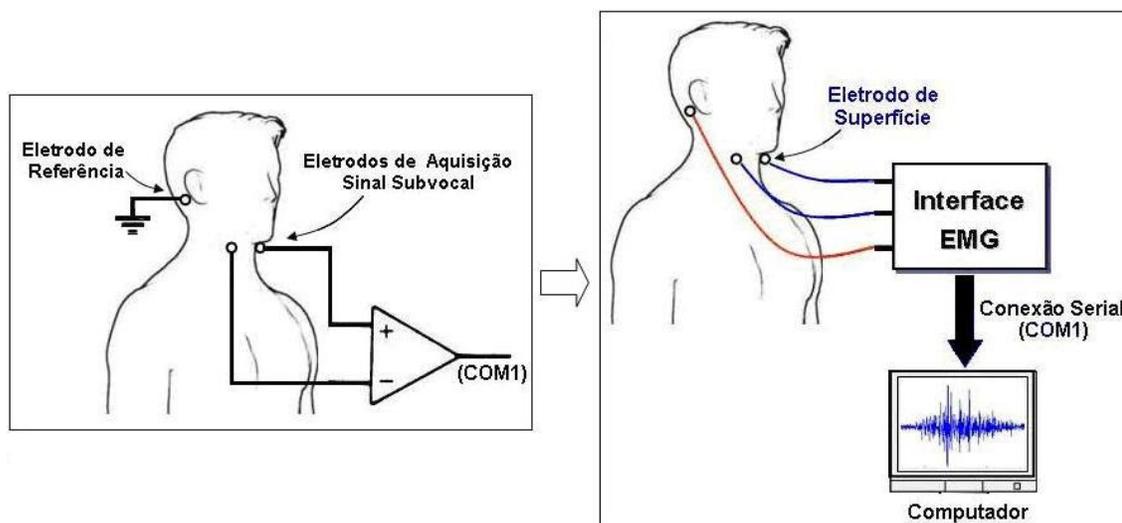


Figura 10: Diagrama do Sistema Subvocal (Posicionamento dos eletrodos e gravação)
 Fonte: Próprio autor, São Luís, 2007

Inicialmente, o resultado da aquisição dos sinais EMG indicou que a posição dos eletrodos de superfície alterou a resposta do sinal EMG, do reconhecimento discreto da vogal subvocal.

A aquisição realizada, utilizando 3 (três) eletrodos de superfícies, onde cada eletrodo foi conectado a uma placa de circuito (interface EMG), onde o gravador de sinal fez a gravação das respostas (sinais EMG), numa frequência de amostragem de 1221 Hz, que é validada para este tipo de aquisição de dados, usando eletrodos de superfície. E também foi utilizado um filtro digital notch de 60 Hz para remover a interferência do ambiente de coleta dos sinais EMG.

A Figura 11 mostra dois sinais EMG típicos das vogais subvocais [a] e [ɛ].

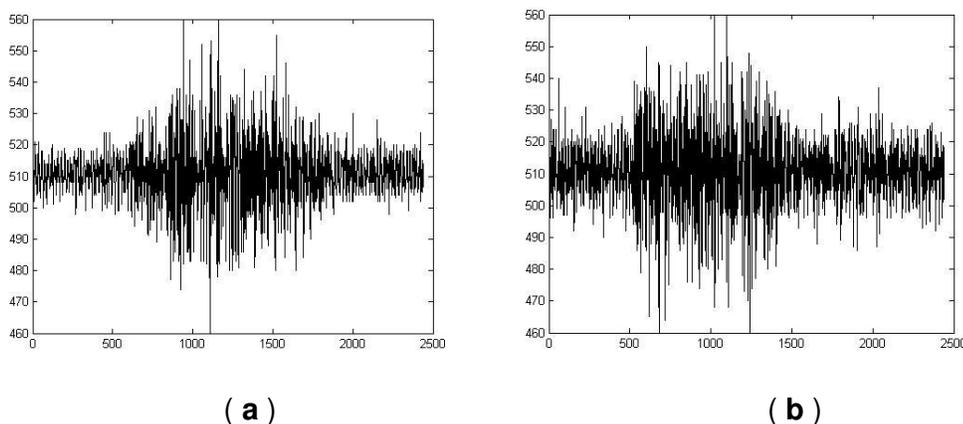


Figura 11: Sinais EMG das vogais subvocais: (a) – vogal subvocal [a], (b) – vogal subvocal [ɛ],
Fonte: MATLAB, 2004

Para aquisição da característica do sinal EMG, scripts Matlab© foram desenvolvidos para extrair e criar um sinal unificado (sinal EMG) para gravação, através do sistema de aquisição de dados. Artefatos EMG tais como tremores de fadiga de músculos de deglutição, ou tosses foram removidos, durante o pré-processamento dos arquivos das vogais subvocais.

7.3 Classificação dos dados usando rede neural MLP

Após as 2 (duas) fases: aquisição de dados e aprendizagem dos dados, 37 funções bases para cada vogal subvocal, totalizando 185 funções bases, foram selecionadas como sendo as características mais significativas dos dados.

Foram selecionadas 80% das funções bases para treinamento e 20% para teste, com 1000 iterações e taxa de aprendizagem de 0,01, e as características básicas (funções bases) escolhidas (ϕ_i) são as entradas da rede neural MLP e RBF, utilizando um programa de simulação de rede neural, Trajan Neural Network Simulator, da Trajan Software Ltd¹⁰. Foram realizados testes com diferentes redes MLP e RBF. O melhor resultado foi obtido com a seguinte configuração 185-21-5, (185 neurônios na camada de entrada), (21 neurônios na camada escondida) e (5 neurônios na camada de saída).

A Figura 12 mostra a arquitetura da rede neural MLP usada nesta pesquisa, possuindo a seguinte configuração de 185:21: 5.

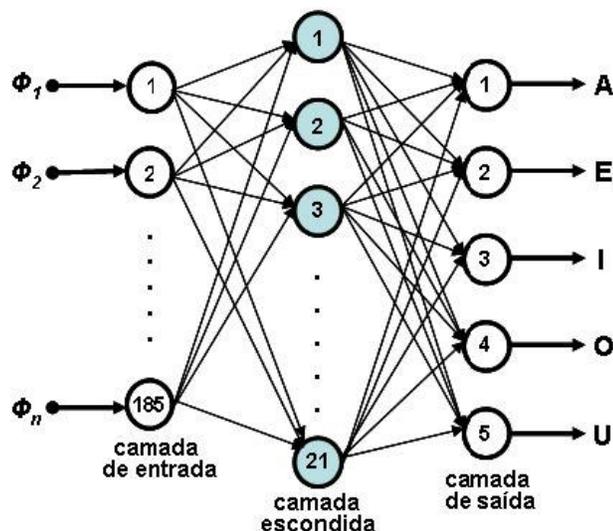


Figura 12: Arquitetura da rede neural MLP usada
Fonte: Próprio autor, São Luís, 2007

¹⁰ www.trajan-software.demon.co.uk

A tabela 1 mostra as 5 (cinco) redes neurais MLP com as suas respectivas camadas de entrada, camadas escondidas e seus desempenhos (validação cruzada).

MLP	Camada de Entrada	Camada Escondida	Desempenho
1	185	17	90,9%
2	185	18	92,3%
3	185	19	92,5%
4	185	20	94,0%
5	185	21	95,4%

Tabela 1: Desempenho das redes neurais MLP
Fonte: Próprio autor, São Luís, 2007

A tabela 2 mostra as 5 (cinco) redes neurais RBF com as suas respectivas camadas de entrada, camadas escondidas e seus desempenhos.

RBF	Camada de Entrada	Camada Escondida	Desempenho
1	185	17	70,9%
2	185	18	72,2%
3	185	19	72,6%
4	185	20	72,8%
5	185	21	73,2%

Tabela 2: Desempenho das redes neurais RBF
Fonte: Próprio autor, São Luís, 2007

O processo de treinamento analisou um conjunto de 20 redes neurais MLP e RBF, usando o programa de simulação de redes neurais, Trajan Neural Network Simulator, o qual auxiliou na obtenção dos resultados finais. Foram comparadas as 5 (cinco) melhores configurações das redes neurais MLP com desempenho entre 90,9% e 95,4% com as redes neurais RBF com desempenho entre 70,9% e 73,2%, utilizando a validação cruzada. Além de tudo isso, é necessário explicar sobre a validação cruzada, que avalia o desempenho de uma rede neural. Por causa disso, decidiu-se manter a rede neural MLP com valor de 95,4%, porque este valor contém o melhor desempenho na classificação com uma configuração de 185-21-5.

E observou-se que a classificação das vogais subvocais teve aproximadamente 95,4% de sucesso.

A tabela 3 mostra a matriz de confusão relativa à classificação das vogais subvocais do modelo proposto.

Vogais Subvocais	Vogal Subvocal A ([a])	Vogal Subvocal E ([ɛ])	Vogal Subvocal I ([i])	Vogal Subvocal O ([ɔ])	Vogal Subvocal U ([u])
A ([a])	35	3	0	0	0
E ([ɛ])	0	34	4	0	0
I ([i])	0	0	35	3	0
O ([ɔ])	2	0	0	34	1
U ([u])	0	0	0	0	36
Sucesso (%)	94,6%	91,8%	94,6%	91,8%	97,3%

Tabela 3: Classificação das vogais subvocais
Fonte: Próprio autor, São Luís, 2007

A análise dos dados (vogais subvocais) indicados na Tabela 2, mostra que o método obteve uma taxa de sucesso de 93,99% sobre o reconhecimento das vogais subvocais. A especificidade e a sensibilidade foram de 93,92% e 94,05% respectivamente. O método obteve 174 verdadeiro-positivos (VP), 139 verdadeiro-negativos (VN), 9 falso-positivos (FP) e 11 falso-positivos (FN).

A Classificação dos dados através da rede neural MLP foi realizada utilizando o programa Trajan Neural Network Simulator, que obteve um desempenho de 95,4% através da validação cruzada, com uma configuração de 185:21: 5.

As tabelas 4 e 5 mostram as propostas das pesquisas anteriores e a proposta desta pesquisa, com as técnicas utilizadas e a taxa de sucesso de cada uma respectivamente.

Pesquisas anteriores	Técnica utilizada	Taxa de sucesso
Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment	- HMM - Rede Neural	74%
Sub Auditory Speech Recognition Based on EMG/EPG Signals	- HMM - Rede Neural	92%
Sub-Auditory Speech Recognition	- Wavelets - Rede Neural	92%
Web Browser Control Using EMG Based Sub Vocal Speech Recognition	- Wavelets - Rede Neural	92%

Tabela 4: Proposta das pesquisas anteriores
Fonte: Próprio autor, São Luís, 2007

Esta Proposta	Técnica utilizada	Taxa de sucesso
Reconhecimento de Fala Subvocal baseado em Eletromiografia de superfície (EMG) utilizando Análise de Componentes Independentes (ICA) e Rede Neural MLP	- ICAI - Rede Neural	93,99%

Tabela 5: Proposta desta pesquisa
Fonte: Próprio autor, São Luís, 2007

Comparando as 2 (duas) tabelas 4 e 5, observou-se que a proposta desta pesquisa obteve uma taxa de sucesso de 93,99% com uma arquitetura simples de baixo custo, maior do que as das pesquisas anteriores, que utilizaram uma arquitetura complexa e cara com equipamentos e softwares proprietários.

7.4 Conclusão

Este capítulo apresentou os resultados desta pesquisa como uma taxa de sucesso maior do que os apresentados em outras pesquisas anteriores. O próximo capítulo apresenta a conclusão e as sugestões de trabalhos futuros.

8 CONCLUSÃO E TRABALHOS FUTUROS

O desenvolvimento desta pesquisa corroborou para compor os resultados apresentados, usando a Análise de Componente Independente (ICA) como uma ferramenta para discriminar os sinais subvocais na extração das funções bases. E também o uso da rede neural MLP, a qual teve um desempenho de 95,4% na classificação dos dados.

Baseado nestes resultados, observou-se que tais características fornecem a sustentação significativa sobre o sinal subvocal com muito mais detalhes. Os resultados incentivavam muito, quando as funções bases dos sinais foram extraídas usando ICA e classificadas que contempla o reconhecimento da fala subvocal, neste caso as 5 (cinco) vogais orais e abertas da língua portuguesa brasileira. A análise dos dados mostra que o modelo proposto obteve uma taxa de sucesso de 93,99% sobre o reconhecimento das vogais subvocais, com uma especificidade de 93,92% e uma sensibilidade de 94,05%.

A realização desta pesquisa mostra que no futuro poderão ser desenvolvidos sistemas que utilizam à fala subvocal em aplicações em diversas áreas (e.g. médicas e industriais). Nesta direção, estes tipos de sistemas poderão auxiliar pessoas a se comunicarem entre elas mesmas, sem uso da fala ou voz audível em qualquer ambiente.

Esta dissertação deu origem a um artigo cujo título “**Subvocal Speech Recognition Based on EMG signal Using Independent Component Analysis and Neural Network MLP**”, foi aceito e publicado no “**2008 International Congress on Image and Signal Processing (CISP2008)**”, sendo organizado pelo IEEE.

REFERÊNCIAS

- ARAUJO, Ezequiel.: **Apostila de Eletromiografia**. Disponível em:
<http://hp.br.inter.net/ezequiel.araujo/eb/arquivos/apostila_de_eletromiografia.doc>.
Acesso em: 15 de julho de 2007.
- BEHLAU, Mara.: **O Livro do Especialista**, Volume I, Editora REVINDER, 2004.
- BELL A. J., SEJNOWSKI T. J.: **An information maximization approach to blind separation and blind deconvolution**. Neural Computation, 1995, Volume 7, No.61, pp.1129-1159.
- BETTS, B. J., JORGENSEN, C.: **Small Vocabulary Recognition Using Surface Electromyography in an Acoustically Harsh Environment**. Interacting with Computers, Volume 18, Issue 6, December 2006, Pages 1242-1259.
- BEUTTENMÜLLER, Glorinha., BEUTTENMÜLLER, Vânia.: **Reequilíbrio da Musculatura Orofacial**, Editora Enelivros, 1995.
- BINSTED, Kim; JORGENSEN, Charles. **Sub-Auditory Speech Recognition**. Conference on Neural Networks, Portland Oregon, July 2003.
- BRAGA, Antonio de Pádua.: **Redes Neurais Artificiais - Teoria e Aplicações**, LTC. 2007.
- CADSOFT. EAGLE Version 4.16r1 for Windows, 2006
- COIMBRA, Cal.: **Análise acústica da voz**. Disponível em:
<http://www.acesa.com/viver/arquivo/vida_saudavel/2005/01/14-cal>. Acessado em 05 de setembro de 2007.
- D'AMBROSIO, Oscar.: **Os donos da voz**. Disponível em:
<<http://www.unesp.br/aci/jornal/162/fonoaudiologia5.htm> >. Acesso em: 13 de julho de 2007.
- DUDA, Richard O, HART, Peter E. STORK, David G.: **Pattern Classification**, 2nd Edition. Wiley Interscience, 2000
- FÉLIX, Sandra.: **Aparelho fonador**. Disponível em:
<http://www.musicaeadoracao.com.br/tecnicos/tecnica_vocal/aparelho_fonador.htm> . Acesso em: 10 de julho de 2007.
- GIROLAMI M., FYFE C.: **An extended exploratory pursuit network with linear and non-linear anti-hebbian lateral connections applied to the cocktail party problem**. Neural Networks, Volume 10, No.9, pp.1607-1618,1997
- GRAMÁTICA DESCRITIVA.: **Vogal**. Disponível em:
<<http://www.radames.manosso.nom.br/gramatica/vogal.htm>>. Acesso em: 26 de setembro de 2007.

_____.: **Vogais da língua portuguesa brasileira.**
Disponível em: <<http://www.radames.manosso.nom.br/gramatica/vogal.htm>>. Acesso em: 26 de setembro de 2007.

HYVÄÄRINEN A., KARHUNEN J., OJA E.: **Independent Component Analysis.**
Nova York: John Wiley & Sons, 2001.

HAYKIN, Simon.: **Neural Networks, A Comprehensive Foundation.** 2 eds.
Pearson Prentice Hall, Indian, 2001.

JORGENSEN, C., BINSTED, K.: **Web Browser Control Using EMG Based Sub Vocal Speech Recognition,** AI Magazine, Volume 21, No.1, pp.57-66, 2000.

JORGENSEN, C; LEE, D., AGABON, S.: **Sub Auditory Speech Recognition Based on EMG/EPG Signals,** In Proceedings of International Joint Conference on Neural Networks, Portland Oregon, July 2003.

JORGENSEN, Chuck.: **NASA Tech Briefs – Interview (May 2004).** Disponível em: <http://www.nasatech.com/NEWS/May04/who_0504.html>. Acesso em: 05 de março de 2007. Trad. MENDES, José da Assunção Gomes, 2007.

JUTTEN C., HERAULT J.: **Blind separation of sources, signal processing,** 24:1-10.1991.

KARHUNEN J., JOUTSENSALO, J.: **Representation and separation of signals using nonlinear PCA type learning.** Neural Networks, 1997.

MARCHETTI, Paulo Henrique; DUARTE, Marcos.: **Instrumentação em Eletromiografia.** Disponível em: <<http://lob.incubadora.fapesp.br/portal/p/EMG.pdf>>. Acesso em: 05 de agosto de 2007.

MARTIN, J.; GLEESON, B.: **A Micropower Dry-Electrode ECG Preamplicifier.** IEEE Transactions on Biomedical engineering, Volume 47, No. 2, pp. 155-162, 2000.

MATHWORKS Inc. MATLAB Version 7.0.1.240704, 2004

PAULA, Maurício Braga de.; **Reconhecimento de palavras faladas utilizando Redes Neurais Artificiais.** Monografia. Pelotas-RS, 2000.

PAPOULIS, Athanasios; PILLAI, S. Unnikrishna.: **Probability, Random Variables and Stochastic Processes.** 4 ed. Nova York: McGraw-Hill. 2002.

SEGURA-JUÁREZ, J.J., Cuesta-Frau, D., Samblas-Pena, L., Aboy, M.: **A Microcontroller-Based Portable Electrocardiograph Recorder.** Volume 51, No.9, pp.1686—1690, 2004.

STUART, W.: **The IrDA Platform. Hewlett-Packard Laboratories,** Technical Report HPL-95-29, Bristol, 1995.

SOUZA, Lourdes Bernadete Rocha de.: **Fonoaudiologia Fundamental.** São Paulo: Editora Revinder, 2000.

TAFNER, Malcon Anderson.: **O Que São as Redes Neurais Artificiais.** Disponível em: < http://www.cerebromente.org.br/n05/tecnologia/rna_i.htm >. Acesso em: 10 de agosto de 2007.

VIANA, Lisley.: **Aparelho fonador.** Disponível em: <http://www.musicaeadoracao.com.br/tecnicos/tecnica_vocal/curso_fisiologia/curso_fisiologia_01.htm>. Acesso em: 10 de julho de 2007.

WIDROW, Bernard., STEARNS Samuel D.: **Adaptive Signal Processing.** Prentice-Hall signal processing series, 1985.