Universidade de São Paulo
Instituto de Física de São Carlos

Sylvio Barbon Júnior

Identificação de patologias na laringe com base na
Discriminative Paraconsistent Machine (DPM)

São Carlos - SP

2011

Sylvio Barbon Júnior

Identificação de patologias na laringe com base
na *Discriminative Paraconsistent Machine*
(DPM)

Tese apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Doutor em Ciências.

Área de concentração: Física Aplicada

Opção Física Computacional.

Orientador: Prof. Dr. Rodrigo Capobianco Guido

Versão corrigida

São Carlos - SP

2011

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

BARBON Jr, Sylvio
Identificação de patologias na laringe com base na
Discriminative Paraconsistent Machine (DPM) / Sylvio
BARBON Jr; orientador Rodrigo Capobianco Guido -
versão corrigida -- São Carlos, 2011.
94 p.

Tese (Doutorado - Programa de Pós-Graduação em
Física Aplicada Computacional) -- Instituto de Física
de São Carlos, Universidade de São Paulo, 2011.

1. Processamento de sinais. 2. Patologias na
laringe. 3. Discriminative Paraconsistent Machine.
4. Inteligência artificial. I. Capobianco Guido,
Rodrigo, orient. II. Título.

Dedico aos meus pais Sylvio e Sonia, a minha
esposa Aninha e familiares mais próximos.

Agradecimentos

A Deus, pela motivação e força em todos os momentos.

Ao meu orientador, pelo companheirismo, orientação e suporte neste trabalho, e ao Instituto de Física de São Carlos, por sediar e prover a infraestrutura necessária para o desenvolvimento do projeto de pesquisa.

À FAPESP, pelo apoio financeiro e institucional (2007/05070-6).

*“Para alcançar o conhecimento, acres-
cente coisas todos os dias. Para alcançar
a sabedoria, remova coisas todos os dias”*

Lao Tse

Resumo

BARBON JR, S. *Identificação de patologias na laringe com base na Discriminative Paraconsistent Machine*. 2011. 94p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

Este trabalho de doutorado apresenta duas inovações: a *Discriminative Paraconsistent Machine* (DPM), que consiste em um novo classificador elaborado com base na lógica paraconsistente anotada (LPA) e a aplicação da DPM para a identificação de patologias na laringe, por meio de exames nos sinais de voz de um locutor. Não há relatos na literatura sobre o uso da LPA para construção de um classificador e sobre suas aplicações para a finalidade proposta. Os resultados obtidos são motivadores, indicando um avanço na área.

Palavras-chave: Processamento de sinais. Patologias da laringe. Inteligência artificial. *Discriminative Paraconsistent Machine* (DPM).

Abstract

BARBON JR, S. *Identification of pathology in larynx based on Discriminative Paraconsistent Machine*. 2011. 94p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011.

This PhD thesis presents two novelties: the *Discriminative Paraconsistent Machine* (DPM), which is a new classifier built on the basis of the annotated paraconsistent logic (APL), and the applications of DPM to identify larynx pathologies, by inspecting a voice signal. There is neither a comment on literature about the use of APL to built a classifier nor about its applications for the proposed application. The results obtained create motivation, showing a clear progress in the field.

Key-words: Signal processing. Larynx pathologies. Artificial intelligence. Discriminative Paraconsistent Machine (DPM)

Lista de Figuras

2.1	Interpretação física simplificada do sistema bio-gerador de voz (adaptado de (17)). . .	30
2.2	Exemplo de nódulo nas pregas vocais (adaptado de (21))	32
2.3	Exemplo de pólipos nas cordas vocais (adaptado de (22))	33
2.4	Exemplo de cisto nas cordas vocais (adaptado de (22)).	33
2.5	Triângulo de Sierpinski (adaptado de (32)).	38
2.6	Janelas Triangulares que delimitam o MFCC segundo (3).	40
2.7	Conjuntos de treinamentos para diferentes tipos de hipóteses (13).	41
2.8	Distância d dos hiperplanos H_1 e H_2 (13).	43
2.9	Reticulado de Hasse representado pelo Quadrado Unitário adaptado de (44).	45
2.10	O espaço paraconsistente representado pelo diamante no plano Cartesiano. . . .	46
3.1	A estrutura proposta para a DPM. Os círculos brancos representam as PUs que formam a camada inicial. Os círculos negros representam as DBUs, que formam a camada discriminativa. Os quadrados cinza com um asterisco representam as SDUs e o círculo com um “O”, representa a POU, pertencente à última camada, que produz a saída da DPM.	52
3.2	Saída da DPM para o experimento com o vetor de entrada $\{1, 0, 1\}$. O círculo ($\circ$) representa a localização do ponto $[G_1, G_2]$, que é exatamente $[0, 1]$	62
3.3	Saída da DPM para o experimento com o vetor de entrada $\{3, 3, 3\}$. O círculo ($\circ$) dentro do losango representa a localização do ponto $[G_1, G_2]$, que corresponde a $[0, -0.589446]$	63
3.4	Saída da DPM para o experimento com o vetor de entrada $\{2, 2, 4\}$. O círculo ($\circ$) dentro do losango representa a localização do ponto $[G_1, G_2]$, que é exatamente $[0.960, 0.034]$	64

3.5	Saída da DPM para o experimento com o vetor de entrada $\{2, 2, 0\}$. O círculo ($\odot$) dentro do losango representa a localização do ponto $[G_1, G_2]$, que é exatamente $[-0.960, 0.034]$.	64
4.1	Janelamento de 1024 elementos.	70
5.1	Testes com 27 parâmetros.	76
5.2	Média de Reconhecimento	77
5.3	Média de Reconhecimento x Quantidade de Elementos.	78
5.4	Comparação entre Desvio Padrão e Reconhecimento.	79
5.5	Comparação de Parâmetros com um grupo de 27 características.	79

Lista de Tabelas

2.1	Sub-tipos de fonemas (adaptado de (17)).	30
2.2	Patologias e suas características (adaptado de (24)).	34
2.3	As 25 bandas críticas do sistema auditivo humano para obtenção dos MFCCs.	40
2.4	Estados possíveis e simbologia apresentados na Figura 2.9.	46
4.1	Banco de 87 vozes recebido do Hospital das Clínicas da FMRP-USP.	68
4.2	Divisão das vozes, totalizando 342 sinais.	69
4.3	Resumo dos parâmetros extraídos dos sinais de vozes.	71
5.1	Combinações dos parâmetros	74
5.2	Melhores resultados dos testes	75
5.3	Piores resultados dos testes	75

Lista de Abreviaturas

D	<i>Fractal Dimension</i>
DPM	<i>Discriminative Paraconsistent Machine</i>
DSP	<i>Digital Signal Processor</i>
FPGA	<i>Field Programmable Gate Array</i>
MFCCs	<i>Mel-Frequency Cepstral Coefficients</i>
RBF	<i>Radial Basis Function</i>
RNA	Redes Neurais Artificiais (<i>Artificial Neural Networks</i>)
SVM	<i>Support Vector Machine</i>
WAV	<i>Waveform Audio Format</i>

Sumário

1	Introdução	25
1.1	Motivação	25
1.2	Objetivos Específicos e Contribuições	27
1.3	Organização do Trabalho	27
2	Revisão Bibliográfica	29
2.1	Sistema de Produção da Fala Humana	29
2.2	Patologias na Laringe	31
2.3	Considerações Relativas à Análise Acústica de Sinais de Voz	34
2.4	Fractais	36
2.5	Coeficientes Cepstrais na Escala Mel (MFCCs)	39
2.6	Máquina de Vetor de Suporte (SVM) no Reconhecimento de Padrões	41
2.7	LPAv2 - Lógica Paraconsistente Anotada com Anotação de Dois Valores	43
3	O Classificador Proposto: <i>Discriminative Paraconsistent Machine</i> (DPM)	49
3.1	Conceituação	49
3.2	Arquitetura da DPM	51
3.2.1	Camadas	53
3.2.2	Treinamento da DPM	58
3.2.3	Usando a DPM no Reconhecimento de Padrões	59
3.3	Exemplo Numérico	59
3.3.1	Problema	59
3.3.2	Solução	60

3.3.3	Testando a DPM treinada	62
3.3.4	Analisando o Plano Paraconsistente	64
3.4	Problemas na Detecção de Normalidade e Patologia	65
4	Materiais e Métodos	67
4.1	Base de Dados	68
4.2	Metodologia	69
5	Testes, Resultados e Discussões	73
6	Conclusões e sugestões para trabalhos futuros	81
	Referências	84
	Apêndice I - Publicações durante o Doutorado	91

Capítulo 1

Introdução

Neste capítulo, a motivação, os objetivos específicos e a organização do trabalho são apresentados e explicados com detalhes.

1.1 Motivação

A busca de novas técnicas que visam a melhoria da qualidade de vida dos seres humanos é um dos principais estímulos à pesquisa, não só na criação de procedimentos, mas também no aprimoramento de abordagens. Com o intuito de contribuir para a distinção computacional entre vozes normais e vozes patológicas * (1) de forma não invasiva, com base em parâmetros acústicos (2-4) e não acústicos, este trabalho apresenta um novo classificador e um novo conjunto de parâmetros que, quando combinados, formam um sistema de classificação eficaz para a finalidade proposta.

Devido à peculiar estrutura do sistema fonador, algumas patologias do trato vocal podem ser confundidas com as definidas como roucas (5), por exemplo. Esse tipo de confusão ocorre principalmente em razão de o modelo tradicional de classificação de vozes ser binário,

*Embora possam parecer estranhos “ao pé da letra”, já que não são as vozes em si que são normais ou patológicas, mas sim as estruturas do aparelho fonador humano, os termos “voz normal” e “voz patológica” são aceitos e comumente utilizados na área.

isto é, emitir respostas do tipo “voz normal” ou “voz patológica” (6-8).

A Máquina Discriminativa Paraconsistente (*Discriminative Paraconsistent Machine* - DPM), que é o novo classificador aqui proposto, utiliza parâmetros combinados, de forma a distinguir as classes de vozes quanto ao grau de normalidade, patologia, além de contradições e indefinições entre tais estados. Uma vez que a classificação obtida no sistema proposto ocorra diferentemente da forma como a lógica clássica, também chamada de binária (9), a base para a elaboração da DPM foi a Lógica Paraconsistente (10). Este último modelo lógico visa oferecer alternativas aos princípios clássicos, que são insuficientes para tratar proposições cujo resultado seja diferente de “normal” ou “patológico”, por exemplo (11).

Assim, a DPM utiliza resultados formulados com base na Lógica Paraconsistente, mais especificamente a Lógica Paraconsistente Anotada com anotação de dois valores (LPA2v), consistindo em um caso particular da Lógica Paraconsistente que considera a existência de argumentos favoráveis e contrários a uma certa proposição, sendo expressos por meio de um grau de crença e um grau de descrença, respectivamente.

O sistema de classificação proposto utiliza um conjunto de parâmetros como entrada, extraídos de cada sinal de voz sob análise. Particularmente, sessenta e quatro parâmetros, acústicos e não acústicos, foram experimentados, incluindo: sexo do locutor, variância do *pitch*, variância das energias das sub-bandas Bark, variâncias dos coeficientes da escala Mel, variâncias das primeiras derivadas dos coeficientes de energia, entre outros. Os parâmetros citados são utilizados amplamente na área, de forma mais isolada (12-16), porém, a utilização combinada dos mesmos não havia sido muito explorada.

Este trabalho obteve taxa de reconhecimento médio de 87,20% para as classificações baseadas em lógica paraconsistente, utilizando uma base de dados de algumas centenas de vozes previamente laudadas por profissionais da área da saúde. As ideias e resultados iniciais foram publicados nos anais do 11th *IEEE International Symposium on Multimedia* (IEEE ISM2009)

com o título “*A Neural-network approach for speech features classification based on paraconsistent logic*”. Os resultados finais foram submetidos a publicação em periódico internacional.

1.2 Objetivos Específicos e Contribuições

O objetivo principal deste trabalho foi o estudo e o desenvolvimento de uma técnica para contribuir para a classificação de vozes, em sistemas de apoio à decisão médica, quanto à presença ou ausência de patologias na laringe, com graus intermediários de resposta. Particularmente, objetivou-se:

- com base nos conceitos básicos da LPA2v, estabelecer o classificador DPM, que permite aplicações práticas na área de classificação e reconhecimento de padrões (10);
- utilizar a DPM para fins de identificação de vozes com os padrões normal e patológico, assim como suas variantes;
- testar uma ampla gama de parâmetros e suas combinações, uma vez que influenciam na taxa de reconhecimento das patologias.

Como objetivo secundário, o estudo observou a viabilidade do uso de parâmetros não acústicos, tais como o sexo do locutor, combinados aos parâmetros acústicos, tradicionalmente utilizados. Nesse caso foram testadas, tanto a combinação ordenada, quanto a desordenada, com quantidades variadas e parâmetros aleatoriamente selecionados. Assim, também foi analisada a influência de um parâmetro sobre outros.

1.3 Organização do Trabalho

Tendo em vista os objetivos previamente detalhados, este trabalho está organizado da seguinte forma. No Capítulo 2 é realizada uma revisão da literatura, descrevendo os conceitos

teóricos sobre patologias na laringe, características acústicas da voz, classificadores tradicionais, LPA, entre outros tópicos. No Capítulo 3, é feita uma descrição detalhada da DPM e dos seus fundamentos teóricos. O Capítulo 4 apresenta a base de dados utilizada e os algoritmos associados para uso da DPM. Já o Capítulo 5 apresenta os testes, resultados e discussões referentes ao trabalho. No Capítulo 6, estão presentes os comentários finais, que levam às conclusões, além de possíveis sugestões para trabalhos futuros. Após as referências, um apêndice inclui as publicações do autor durante o curso de doutorado.

Capítulo 2

Revisão Bibliográfica

Neste capítulo os conceitos necessários ao desenvolvimento do presente projeto são brevemente revistos em um nível de profundidade suficiente para o fim à qual se destinam.

2.1 Sistema de Produção da Fala Humana

Segundo (17), a produção de voz, por mais complexa que seja, pode ser descrita de forma pormenorizada como sendo a propulsão de ar pelos pulmões, seguida de um processo de equalização garantido pelo trato vocal e elementos associados, conforme observado na Figura 2.1. Um detalhe importante corresponde às pregas vocais controlando o fluxo de ar fornecido pelos pulmões, tornando o sinal de excitação quase periódico, vibrando em uma determinada frequência, ou aperiódico, similar a um sinal ruidoso. No primeiro caso, o período do mesmo será chamado de período de *pitch* e a voz produzida será classificada como vozeadas, do contrário a mesma será classificada como não vozeadas, assemelhando-se a um ruído. Os principais exemplos de *vozeadas speech* são as vogais, e a maioria dos demais sons emitidos pelo ser humano podem ser classificados como *não vozeadas speech*. Uma característica importante é obtida com o véu palatino, pois o mesmo controla a passagem do fluxo de ar pelos tratos nasal e vocal, sendo que eles podem refinar a qualidade dos sinais para uma classificação conforme a Tabela 2.1.

Figura 2.1- Interpretação física simplificada do sistema bio-gerador de voz (adaptado de (17)).

Tabela 2.1- Sub-tipos de fonemas (adaptado de (17)).

tipo	descrição	exemplo
fricativas	não vozeadas com a presença de uma turbulência de ar entre a língua e os dentes superiores	<u>th</u> na palavra <i>thin</i>
oclusivas sussurradas	não vozeadas por impulso não vozeadas com uma barreira nas pregas vocais de forma que elas permaneçam parcialmente fechadas	<u>t</u> na palavra <i>top</i> <u>h</u> na palavra <i>he</i>
vozeadas fricativas	vozeadas misturados com ruído criado na constrição do trato vocal atrás dos dentes e contra o palato	<u>z</u> na palavra <i>zebra</i>
não vozeadas fricativas	idem anterior, porém as pregas vocais não vibram simultaneamente com a fricção	<u>h</u> na palavra <i>human</i>
vozeadas oclusivas	fonemas vozeadas misturados com ruído impulsivo criado no trato vocal	<u>g</u> na palavra <i>glass</i>
não vozeadas oclusivas	idem anterior, porém as pregas vocais não vibram simultaneamente com o impulso	<u>b</u> na palavra <i>boat</i>

Levando em consideração a estrutura e funcionamento descritos anteriormente, qualquer alteração estrutural causada por uma patologia no trato vocal, implica a obtenção de um sinal de voz diferente do normal. Essas alterações, sendo substanciais, têm efeito direto na voz do locutor, podendo ser facilmente perceptíveis por qualquer pessoa. Além disso, o profissional da área médica é capaz de realizar um pré-diagnóstico sem uma análise clínica mais detalhada. Porém, quando a alteração não é tão significativa, o seu impacto na voz do locutor pode não ser perceptível mesmo por um profissional da área médica. Nesses casos, a análise do sinal de voz por um programa de computador é aconselhável, podendo ser utilizada como uma técnica eficaz para alertar o locutor sobre a saúde de sua estrutura vocal, indicando a necessi-

dade de procurar um profissional especializado para realização de exames detalhados.

No tocante às pesquisas atuais referentes ao processamento digital de voz, envolvendo fonoaudiólogos, engenheiros, físicos, matemáticos, entre outros profissionais, têm-se estudado várias medidas de perturbação em tais sinais. Particularmente, diversos parâmetros são utilizados para medidas de perturbação em torno da frequência fundamental (*pitch*). Tal frequência é um importante parâmetro na avaliação anatômica e funcional da laringe (18), sendo determinada pela quantidade de vibrações das pregas vocais em um determinado período de tempo. As medidas de variação da frequência e amplitude ciclo-a-ciclo são respectivamente chamadas: *jitter* e *shimmer* (19-20).

2.2 Patologias na Laringe

Considerando critérios acústicos, as patologias relacionadas ao trato vocal são diferenciadas em relação a padrões perceptuais tais como rouquidão, aspereza e soprosidade (1) (6). A laringe, parte do trato vocal, não é diferente, tratando-se de um órgão fibromuscular suscetível a várias patologias. Dentre as mais comuns, e que afetam a voz, notam-se os nódulos vocais, exibidos na Figura 2.2 e popularmente conhecidos como “calos nas pregas vocais”, o pólip, ilustrados na Figura 2.3, o cisto, observados na Figura 2.4, além dos edemas de Reinke, também conhecidos como inchaços de Reinke.

A avaliação acústica da voz, baseada em termos perceptuais, conduz até a origem da patologia, uma vez que essa efetivamente causou modificações acústicas, embora possa ocorrer situações nas quais “ouve-se” um determinado padrão acústico e nenhuma patologia é detectada em exames tradicionais mais invasivos, porém mais precisos, tais como a videolaringoscopia e a videoestroboscopia (5).

A Tabela 2.2 exhibe as principais patologias, suas localizações e características acústicas, conforme (23).

Figura 2.2- Exemplo de nódulo nas pregas vocais (adaptado de (21))

Métodos não invasivos capazes de diferenciar patologias vocais, baseado em padrões e medidas acústicas, são motivo de muitas pesquisas científicas. Os métodos tradicionais, invasivos, além de gerar desconforto, apresentam um problema relacionado ao reflexo na cavidade supraglotal, que pode levar a avaliações equivocadas sobre as condições da laringe do paciente. Assim, além dos exames invasivos e da avaliação perceptual acústica, a utilização de outras tecnologias pode auxiliar cada vez mais no diagnóstico preciso e simplificado das doenças vocais. Este trabalho realizou sua detecção baseada em duas patologias: o Edema de Reinke e o Nódulo nas pregas vocais, que apresentam características acústicas similares em vários casos.

O nódulo é uma reação do tecido vocal ao constante estresse, induzido pelo movimento frequente oposicional brusco das pregas vocais (25). Esse é um dos problemas mais comuns na laringe, mas é uma lesão benigna, geralmente imóvel durante a fonação (26). O comportamento vocal inadequado e o uso excessivo da voz são indicados como os fatores principais do aparecimento de nódulos, sendo o álcool, o fumo, os distúrbios hormonais, as infecções e alergias, são os elementos predisponentes mais importantes. São caracterizados por uma protuberância esbranquiçada na borda glótica das pregas vocais, geralmente bilateral, localizados na porção

Figura 2.3- Exemplo de pólipos nas cordas vocais (adaptado de (22))

Figura 2.4- Exemplo de cisto nas cordas vocais (adaptado de (22)).

média da membrana vocal, local de máxima tensão muscular durante a fonação. Acusticamente, provocam uma vibração assimétrica e o incompleto fechamento das pregas vocais, acarretando soprosidade e variações no período de *pitch* e na amplitude da voz, além de acrescentar ainda um certo nível de ruído em alguns componentes espectrais da voz. Dessa forma, o estudo e avaliação de parâmetros acústicos que quantifiquem tais variações, podem indicar a presença da patologia.

A outra doença abordada pela proposta, segundo (23), o Edema de Reinke é um edema generalizado e bilateral que traduz a presença de uma laringite crônica. Também chamado de

Tabela 2.2- Patologias e suas características (adaptado de (24)).

Patologias	Localização	Características Vocais
Edema de Reinke	Ao longo de toda a prega vocal	Voz grave, fluida, rouca ou crepitante, no início agradável, virilizada
Nódulo	Metade da porção vibratória, junção de 1/3 anterior e 1/3 médio da prega vocal	Voz rouco-soprosa, grau discreto a extremo, ataques bruscos, ressonância baixa; alguns casos sem sinais vocais
Pólipo	Borda livre da prega vocal	Voz grave, rouca, rouco-soprosa, frequência dicrótica, fadiga vocal
Úlcera de contato	Porção cartilaginosa da prega vocal	Voz grave-trepitante, disfonia silente, odinofonia, odinofagia e otalgia
Leucoplasia	Na mucosa, concentrada na borda livre	Voz áspera, rouco-áspera, frequência aguda e fadiga vocal

degeneração polipoidal, usualmente é associada aos pacientes fumantes de longo período de tempo, bem como àqueles que fazem uso intensivo da voz, ou apresentam algum fonotrauma (28). A patologia recebeu esse nome devido a sua localização, próxima ao espaço anatômico com o nome do anatomista Reinke, que foi o primeiro a investigar a anatomia das pregas vocais. É também considerada uma lesão benigna na laringe, podendo ser caracterizada por um aumento fluídico que produz mudanças no movimento das pregas vocais. Essas mudanças têm o efeito acústico de rouquidão na voz do paciente.

2.3 Considerações Relativas à Análise Acústica de Sinais de Voz

Por meio da avaliação acústica, é possível a obtenção de informações sobre aspectos relevantes com relação à função dos componentes da laringe, incluindo particularidades relacionadas com a aerodinâmica e biomecânica, estabilidade fisiológica da fonação e coordenação

laríngea na fonoarticulação (29). Dessa forma, também é possível verificar as características do aspecto vocal, além de garantir a análise, extração e quantificação de padrões do sinal vocal.

De acordo com (27), muitos são os ganhos proporcionados pela análise acústica, dentre os quais se destacam: a utilidade de instrumentos para detecção precoce de alterações laríngeas; possibilidade de monitorar o andamento terapêutico por meio da comparações de resultados obtidos em épocas diferentes; compreensão acústica da saída vocal mais clara; melhor associação entre as análises preceptivo-auditiva e acústica.

Com relação à interpretação adequada dos sinais de voz, segundo a literatura, há duas maneiras de proceder com a análise:

- Sinal Direto: é a técnica que avalia os parâmetros acústicos diretamente do sinal de entrada (voz), no domínio do tempo. Esse mecanismo não foi objeto de estudo;
- Sinal Indireto: segundo esta técnica, os parâmetros são avaliados a partir de uma transformação de base, utilizando-se o domínio da frequência e o domínio cepstral (3). Por exemplo, pode-se avaliar os parâmetros no espectro do sinal com o cepstrum (3). Essa abordagem foi objeto de estudo neste trabalho.

O período de *pitch*, também chamado de frequência glótica, relacionado fortemente ao gênero e à idade do indivíduo, é considerado um dos parâmetros acústicos mais robustos para caracterização da voz. Segundo (23), em mulheres adultas, o *pitch* médio situa-se ao redor de 200Hz e nos casos dos homens adultos, cerca de 100Hz. O *pitch* é considerado a forma perceptiva da frequência fundamental, e a constatação de que um paciente tem uma regulação de frequência adequada envolve um julgamento, por especialista treinado, considerando altura, sexo, idade, entre outros fatores (30-31).

Indivíduos com algum tipo de patologia na laringe podem apresentar perturbações mais evidentes no *pitch* do que indivíduos com vozes normais, mesmo com um período médio fundamental semelhante. O fator de perturbação é sensível ao tamanho e à localização da alteração

tecidual patológica na laringe do sujeito (30).

A verificação da perturbação no domínio da frequência possibilita outro conjunto de informações, diferentes das obtidas no domínio do tempo. As informações utilizadas neste projeto são originárias desse tipo de análise. Com tal abordagem, o *pitch* e as frequências formantes, que são as frequências de ressonância do trato vocal das vogais sustentadas, são encontrados de maneira simplificada. Diversos estudos demonstram que existem significativas diferenças entre espectros de vozes normais e patológicas. Os componentes espectrais do ruído em vozes não saudáveis são originados pelo fluxo de ar gerado por características da glote ou vibração assíncrona das pregas vocais, sendo verificados de maneira clara com a análise espectral.

Uma das ferramentas tradicionais para a conversão do domínio do tempo para o da frequência é a Transformada de Fourier, neste caso a Transformada Discreta de Fourier (*Discrete Fourier Transform*- DFT), que está expressa na equação 2.1. Existem algoritmos eficientes para o cômputo da DFT, conhecidos como algoritmos de transformada rápida de Fourier (*Fast Fourier Transform* - FFT). Essas alternativas, segundo (20) reduzem sua ordem de complexidade computacional, de quadrática para logarítmica. O presente trabalho utiliza a implementação FFT para análise das informações no domínio da frequência.

$$X[\omega] = \sum_{n=0}^{N-1} x_n e^{-j2\pi n\omega/N} \quad , \quad \text{sendo } N \text{ o número de pontos do sinal temporal } x[\cdot]. \quad (2.1)$$

2.4 Fractais

Para (32), fractais são objetos gerados pela repetição de um mesmo processo recursivo, apresentando auto-semelhança, sendo auto-semelhança ou auto-similaridade a invariância da estrutura após uma transformação isotrópica, ou seja, que se dá com a mesma intensidade em todas as direções. Tomando um objeto S , constituído por um conjunto de pontos $R =$

$\{x_1, x_2, x_3, \dots\}$, a aplicação de uma transformação auto-similar com um fator de escala b , muda as coordenadas dos pontos $BR = \{b_{x1}, b_{x2}, b_{x3}, \dots\}$. Logo, o conjunto S formado pelos pontos de coordenadas R , é auto-similar se esse for invariante após a referida transformação (33).

Segundo (34), e considerando os conceitos que envolvem a geometria Euclidiana, chamada de dimensão topológica, um ponto possui dimensão com valor 0, 1 é o valor para uma reta, uma superfície tem dimensão 2 e o espaço possui dimensão 3. Dessa forma, é possível determinar a dimensão topológica de qualquer objeto, existindo uma correspondência proporcional entre este objeto e um espaço geométrico elementar. As formas geométricas dos fractais são incapazes de ser classificadas com a geometria Euclidiana, devido às três características fundamentais que as definem e distinguem de outras formas: a auto-semelhança em diferentes níveis de escala, sua complexidade infinita e a dimensão fractal (35). A dimensão fractal (D) tem relação com a dimensão espacial, ou seja, ao espaço ocupado por um sinal ou figura, quando o valor é correspondente a um número fracionário. Portanto, D consiste em uma medida para o nível de irregularidade ou auto-similaridade desse sinal. No presente trabalho, a variância de D será utilizada como medida de irregularidade dos sinais sob análise. Tal irregularidade pode ser associada com algumas patologias específicas.

De acordo com (34), pode-se definir D com a equação 2.2.

$$D = \frac{\log(N)}{\log(\frac{1}{T})} \quad , \quad (2.2)$$

o que significa que a regra para a formação do fractal consiste em substituir um segmento, parte ou objeto, por N segmentos de tamanho T , correspondentes a uma fração do tamanho anterior. Tradicionalmente, utiliza-se o triângulo de Sierpinski (32-24) (36), para demonstração do cálculo de D , conforme a equação 2.3.

$$D = \frac{\log(N)}{\log(\frac{1}{T})} = \frac{\log(3)}{\log(\frac{1}{\frac{1}{2}})} = \frac{\log(3)}{\log(2)} \approx 1.585 \quad (2.3)$$

A cada etapa da construção, um triângulo é substituído por outros três de área igual a $\frac{1}{2}$ da anterior. A Figura 2.5 possibilita visualizar o fato, sendo possível verificar a similaridade de formatos dos triângulos por meio da simples observação. Outros tipos de elementos, assim

Figura 2.5- Triângulo de Sierpinski (adaptado de (32)).

como o sinal de voz, podem ter um tipo de similaridade encontrado de outras maneiras, obtível por meio de ferramentas estatísticas. Independentemente do tipo de similaridade, os sinais unidimensionais da voz são tais que $1 \leq D \leq 2$, segundo (34) (37). Assim, durante o processo de construção de um fractal, com lei de formação bem definida, como, por exemplo, o triângulo de Sierpinski, é simples a obtenção de D . Os métodos tradicionais para a medição são: *walking-divider*, *box-counting*, *prism* e *power spectrum*. Segundo (34), a técnica do *power spectrum* é mais eficiente do que as demais, sendo baseada no espectro de potência do sinal fractal. O algoritmo consiste em obter $D = 2 - |H|$, sendo $H = \frac{\beta-1}{2}$, que é o expoente de Hurst, β é o coeficiente angular da reta média obtida por meio da regressão linear por Mínimos Quadrados, que, por sua vez, atende o conjunto de pontos da forma $\log_2(b_i)$ versus $\log_2(a_i)$, sendo b_i o valor normalizado do espectro de potência da banda i que possui comprimento a_i para N sub-bandas:

$$\beta = \frac{N \sum_{k=0}^{N-1} \log_2(a_k) \log_2(b_k) - \sum_{k=0}^{N-1} \log_2(b_k) \sum_{k=0}^{N-1} \log_2(a_k)}{N \sum_{k=0}^{N-1} \log_2(a_k)^2 - \left(\sum_{k=0}^{N-1} \log_2(a_k) \right)^2} \quad (2.4)$$

2.5 Coeficientes Cepstrais na Escala Mel (MFCCs)

A escala Mel, fundamentada em um modelo aprimorado para frequências de voz referente ao padrão perceptual humano, serve como base para avaliação do cepstrum. De acordo com (3), a análise cepstral tem sido largamente empregada em processamento digital de sinais de voz, e a implementação mais popular é aquela que combina a escala cepstrum com uma dilatação não-linear, de acordo com a escala Mel (38-39). Assim, os Coeficientes Cepstrais na Escala Mel (MFCCs - *Mel Frequency cepstral coefficients*), são uma representação dos componentes espectrais do sinal, de maneira mais conveniente e particularizada para um sinal de voz.

O cálculo dos MFCCs pode ser realizado de diferentes maneiras. Segundo (40), o processo de extração dos MFCCs descarta informações para possibilitar que os vetores no domínio cepstral apresentem informações relevantes segundo a escala em questão. Por exemplo, informações relacionadas com a fase são perdidas durante a operação de magnitude. Isso acontece também para outras características do sinal, que durante a redução da resolução e a suavização provida pela aplicação da Transformada Cosseno Discreta (*Discrete Cosine Transform* - DCT) junto ao logarítmo.

Outra abordagem, utilizada neste trabalho, consiste no ajuste do espectro de potência S de cada *frame* sucessivo do sinal de voz, modelado tanto em frequência quanto em amplitude, com base em uma escala logarítmica. Os M coeficientes iniciais c_i da FFT são obtidos, sendo $9 \leq M \leq 14$. A equação 2.5 expressa o processo, sendo que, os X_k correspondem aos sinais filtrados de acordo com a escala Mel ou Bark. O conjunto de respostas em frequências dos vinte filtros utilizados tem um aspecto triangular, como ilustra a figura 2.6.

$$c_i = \sum_{k=1}^{20} X_k \cos\left(i\left(k - \frac{1}{2}\right)\frac{\pi}{20}\right) \quad , \quad \text{sendo} \quad i = 0, 1, \dots, M \quad . \quad (2.5)$$

A escala Bark mencionada anteriormente permite subdividir a faixa de frequências audíveis pelo ouvido humano em 25 intervalos chamados de bandas críticas. Quando dois ou mais sons, de frequências diferentes, mas pertencentes a uma mesma banda crítica, estão

Figura 2.6- Janelas Triangulares que delimitam o MFCC segundo (3).

sendo captados simultaneamente, o de maior amplitude pode mascarar os demais, e o de menor amplitude não é percebido. A Tabela 2.3 lista as bandas críticas nas escalas de frequências em Hertz (HZ) e na escala Bark (z), que é uma escala adotada para linearizar o intervalo entre as bandas. A conversão entre Hertz e Bark é dada pela equação $\frac{26,81}{1 + \frac{1960}{f}}$ 0.53, sendo que f representa a frequência em Hertz. Nos casos em que $z < 2$ aplica-se a correção $z \leftarrow z + 0.15 * (2 - z)$ e, quando $z > 20$, faz-se $z \leftarrow z + 0.22 * (z - 20.1)$.

Tabela 2.3- As 25 bandas críticas do sistema auditivo humano para obtenção dos MFCCs.

<i>Bark (z)</i>	<i>Hertz (Hz)</i>	<i>Bark (z)</i>	<i>Hertz (Hz)</i>
0	0 - 100	12	1720 - 2000
1	100 - 200	13	2000 - 2320
2	200 - 300	14	2320 - 2700
3	300 - 400	15	2700 - 3150
4	400 - 510	16	3150 - 3700
5	510 - 630	17	3700 - 4400
6	630 - 770	18	4400 - 5300
7	770 - 920	19	5300 - 6400
8	920 - 1080	20	6400 - 7700
9	1080 - 1270	21	7700 - 9500
10	1270 - 1480	22	9500 - 12000
11	1480 - 1720	23	12000 - 15500
		24	15500 - 22050

2.6 Máquina de Vetor de Suporte (SVM) no Reconhecimento de Padrões

Support Vector Machine ou Máquina de Vetor de Suporte (SVM) é um conjunto de métodos de Aprendizado de Máquina (AM) do tipo supervisionado para análise de informação e reconhecimento de padrões. As SVMs são baseadas em métodos de classificação estatística e análise regressiva (41).

As SVM são caracterizadas por um pré-processamento da informação para representar os padrões em uma dimensão mais elevada (12). A Teoria do Aprendizado Estatístico (TAE), considera f um classificador e F o conjunto de todos os classificadores que um determinado algoritmo de AM pode gerar. Esse algoritmo, durante o processo de aprendizado, utiliza um conjunto de treinamentos T , composto de N pares (x_i, y_i) , para gerar um classificador $f \in F$. A Figura 2.7 (13) exemplifica um conjunto de treinamentos em que é possível separar o conjunto das classes “círculo” e “triângulo”.

Figura 2.7- Conjuntos de treinamentos para diferentes tipos de hipóteses (13).

Com um uso de k transformações não-lineares $\phi_j, j = 1, 2, \dots, k$, assume-se que cada padrão x_k deve ser transformado em $y_k = \phi(x_k)$, considerando $z_k \pm 1$, e o padrão k pertencendo a ω_1 ou ω_2 . Segundo (13), as SVMs lineares com margens rígidas definem fronteiras lineares a

partir de dados linearmente separáveis. Assim, ao assumir um conjunto T para treinamento com N dados, $x_i \in X$, e seus respectivos rótulos $y_i \in Y$, X constitui o espaço de dados e $Y = -1, +1$, T é linearmente separável se for possível separar as classes $+1$ e -1 por meio de um hiperplano. Quando é possível separar dados por meio de um hiperplano, os classificadores utilizados são ditos lineares. O hiperplano é apresentado na Equação 2.6, sendo que $\vec{w} \cdot \vec{x}$ produto escalar entre os vetores $\vec{w}$ e $\vec{x}$. Considerando $w \in X$ como o vetor normal ao hiperplano descrito, e $\frac{b}{\|w\|}$ como a distância do hiperplano em relação a origem, com $b \in \mathbb{R}$.

$$f(x) = \vec{w} \cdot \vec{x} + b = 0 \quad (2.6)$$

A Equação 2.6 divide o espaço dos dados X em duas regiões: $\vec{w} \cdot \vec{x} + b > 0$ e $w \cdot x < 0$. Segundo (12) o vetor de suporte é a distância do hiperplano H_1 que delimita ω_1 , obtido por meio do treinamento baseado no padrão de uma classe, com o hiperplano H_2 , que delimita ω_2 . Dessa forma, sendo x_1 um ponto de $H_1 : \vec{w} \cdot \vec{x} + b = +1$ e x_2 um ponto no hiperplano $H_1 : \vec{w} \cdot \vec{x} + b = -1$, caso se projete $\vec{x}_1 - \vec{x}_2$ na direção de $\vec{w}$, perpendicularmente ao hiperplano médio, chamado por (12) como *optimal hiperplane* sendo $\vec{w} \cdot \vec{x} + b = 0$, é possível obter a distância entre os hiperplanos H_1 e H_2 . Essa projeção é obtida por meio da equação 2.7, descrita com mais detalhes em (13).

$$(\vec{x}_1 - \vec{x}_2) \left(\frac{\vec{w}}{\|\vec{w}\|} \right) \cdot \frac{(\vec{x}_1 - \vec{x}_2)}{\|\vec{x}_1 - \vec{x}_2\|} \quad (2.7)$$

Assim, a fase de treinamento da SVM consiste em encontrar o *optimal hiperplane*, o que torna possível verificar a máxima distância d entre os pontos mais próximos do padrão das classes avaliadas. As distâncias e identificações dos hiperplanos H_1 e H_2 podem ser verificados na Figura 2.8.

No presente trabalho, a concepção da DPM será embasada na SVM.

Figura 2.8- Distância d dos hiperplanos H_1 e H_2 (13).

2.7 LPAv2 - Lógica Paraconsistente Anotada com Anotação de Dois Valores

A Lógica Paraconsistente, diferentemente da lógica convencional ou lógica booleana, utilizada pelas máquinas computacionais, é caracterizada como uma lógica não-clássica criada pela necessidade de uma análise mais eficiente com capacidades distintas. Atualmente, a Lógica Paraconsistente apresenta duas abordagens: a Lógica Paraconsistente Anotada (LPA) e Lógica Paraconsistente Anotada com anotação de dois valores (LPAv2) (10). Esta segunda foi objeto dos estudos deste trabalho, para a implementação do algoritmo para-analisador DPM, que permite a implementação computacional da proposta.

As lógicas clássicas, que caracterizam qualquer situação como verdadeira ou falsa, perdem, muitas vezes, um enquadramento exato da realidade abordada. Com o uso de LPAv2, é possível avaliar situações reais, como ambiguidade e inconsistência. A LPAv2 é denotada por $p(\mu, \lambda)$, sendo que $p(\mu, \lambda) \in [0, 1]$ e $[0, 1]$ é um intervalo real unitário, sendo p a variável proposicional. Segundo (42), a seguinte relação pode ser estabelecida: $[0, 1] : (\mu_1, \lambda_1) \leftarrow (\mu_2, \lambda_2) \Leftrightarrow \mu_1 \leftarrow \mu_2 \wedge \lambda_1 \leftarrow \lambda_2$.

Considerando $p(\mu, \lambda)$, o mesmo pode ser lido intuitivamente como “considerando μ como grau de crença e λ sendo o grau de descrença da proposição...”. Com esse mesmo raciocínio, é possível considerar o primeiro parâmetro como evidência favorável, e o segundo como a evidência desfavorável. Sendo assim, um $p(1.0, 0.0)$ intuitivamente demonstra crença total, $p(0.0, 1.0)$ descrença total, $p(1.0, 1.0)$ indicando inconsistência e $p(0.0, 0.0)$, uma indefinição. No caso deste trabalho, pode-se contextualizar da seguinte forma:

Seja a proposição $p =$ “O paciente está com a sua voz saudável”. Tem-se, então:

- se anotada como $(1.0, 0.0)$, a leitura intuitiva será “O paciente está com a voz saudável com crença total”;
- se anotada como $(0.0, 1.0)$, a leitura intuitiva será “O paciente está com a voz saudável com descrença total”, ou seja, foi encontrada uma possível patologia;
- se anotada como $(1.0, 1.0)$, a leitura intuitiva será “O paciente está com a voz saudável como crença inconsistente”, ou seja, podem ser necessários mais exames ou procurar a opinião de outro especialista, ou exame, para validar a possibilidade de patologia;
- se anotada como $(0.0, 0.0)$, a leitura intuitiva será “O paciente está com a voz saudável com ausência total de crença”, ou seja, não é possível detectar qualquer presença de patologia, sendo esse chamado “estado paracompleto”.

Segundo (43), a LPAv2 provê um ágil mecanismo de distinção entre pacientes acometidos por Mal de Alzheimer, com uma boa classificação e desempenho satisfatório, sendo uma boa ferramenta de apoio, a decisão médica. Para se tornar um mecanismo robusto e analisar

informações como o grau de certeza e contradição, muitos trabalhos (10)(42)(44)(43)(45) utilizam o reticulado de *HASSE* com anotação para a LPAv2, sendo $\tau = (\mu_1, \mu_2) | (\mu_1, \mu_2) \in [0, 1] \subset R$ agora tem τ substituindo p, μ_1 no lugar de μ e μ_2 usado para simbolizar o segundo parâmetro λ . O reticulado citado é composto por quatro vértices, podendo ser representado por um Quadro Unitário no Plano Cartesiano (QUPC), como ilustrado na Figura 2.9.

Figura 2.9- Reticulado de Hasse representado pelo Quadrado Unitário adaptado de (44).

Em um sistema de análise paraconsistente, baseando-se no QUPC, é possível calcular o grau de contradição G_{ct} pela equação 2.8.

$$G_{ct} = \mu_1 + \mu_2 - 1 \quad . \quad (2.8)$$

O grau de contradição varia de -1 a $+1$, e é facilmente identificado no QUPC, cujo valor é semelhante à distância do ponto de interpolação entre os graus de crença e descrença e à reta que liga o ponto $D = (1, 0)$ - Verdadeiro ao ponto $B = (0, 1)$ - Falso. O valor $G_{ct} = -1$, que ocorre no ponto $A = (0, 0)$, representa uma contradição máxima negativa e o valor

Figura 2.10- O espaço paraconsistente representado pelo diamante no plano Cartesiano.

Tabela 2.4- Estados possíveis e simbologia apresentados na Figura 2.9.

Estado	Símbolo
<i>Verdade</i>	V
<i>Falso</i>	F
<i>Inconsistência</i>	$\top$
<i>Para completo</i>	$\perp$
<i>Quase-Verdade tendendo a Inconsistência</i>	$Qv \rightarrow \top$
<i>Quase-Verdade tendendo a Para completo</i>	$Qv \rightarrow \perp$
<i>Quase-Falso tendendo a Inconsistência</i>	$Qf \rightarrow \top$
<i>Quase-Falso tendendo a Para completo</i>	$Qf \rightarrow \perp$
<i>Quase-Inconsistente tendendo a Verdade</i>	$Q\top \rightarrow V$
<i>Quase-Inconsistente tendendo a Falso</i>	$Q\top \rightarrow F$
<i>Quase-Paraconsistente tendendo a Verdade</i>	$Q\perp \rightarrow V$
<i>Quase-Paraconsistente tendendo a Falso</i>	$Q\perp \rightarrow \top$

para $G_{ct} = +1$, que acontece no ponto $C = (1, 1)$, representa a contradição máxima positiva. Assim, em um sistema de análise paraconsistente, quanto mais interpolação entre os graus de crença e de descrença for próximo do segmento BD , visível pelo QUPC, mais o resultado da soma dos graus de crença e descrença (μ_1 e μ_2) se aproxima de 1, minimizando o valor de G_{ct} . Essa diminuição de G_{ct} representa uma menor contradição entre as informações parametrizadas. Finalmente, quando a soma dos graus de crença e descrença (μ_1 e μ_2) for igual a 1, o grau de contradição é zero e o ponto de interpolação estará sobre a reta BD . Nesse caso, $G_{ct} = 0$ e não há contradição entre os sinais. Uma outra análise facilmente realizada com o uso do QUPC é o

grau de certeza G_c , dado pela equação 2.9.

$$G_c = \mu_1 + \mu_2 \quad (2.9)$$

A quantificação do grau de certeza G_c pode variar de -1 a $+1$. Verificando no QUPC o valor corresponde à distância do ponto de interpolação entre os graus de crença e de descrença com relação à reta que liga o ponto $A = (0, 0)$ ao ponto $C = (1, 1)$. O valor de $G_c = -1$, que corresponde ao ponto $B = (0, 1)$, significa o valor de certeza máxima da negação da proposição. O valor de $G_c = +1$, que corresponde ao ponto $D = (1, 0)$, significa intuitivamente que existe uma certeza máxima na afirmação da proposição. Na Figura 2.9, é possível observar que quanto mais a interpolação entre os graus de crença e de descrença juntar-se ao segmento de reta AC , mais o resultado da subtração dos graus de crença pela descrença se aproxima de 0, diminuindo o valor de G_c , representando menor certeza entre as informações de entrada. Caso os graus de crença e descrença forem de valores iguais ($\mu_1 = \mu_2$), o grau de certeza será zero e o ponto de interpolação estará sobre a reta AC , significando indefinição entre os sinais. Sendo assim, os graus de crença, inconsistência, contradição (G_{ct}) e certeza (G_c) tornam a LPAv2 mais próxima das reações naturais, nas quais, muitas análises não têm a rigidez da lógica booleana, com afirmações somente verdadeiras ou falsas.

Capítulo 3

O Classificador Proposto: *Discriminative Paraconsistent Machine* (DPM)

Este capítulo apresenta a arquitetura da DPM, seu funcionamento, um exemplo numérico de uso, entre outros detalhes.

3.1 Conceituação

Como mencionado nos capítulos anteriores, modelos baseados em treinamento para classificação de padrões têm sido amplamente discutidos e indicados pela literatura. Neste capítulo, está descrita a técnica proposta, *Discriminative Paraconsistent Machine* (DPM), que é uma modelagem discriminativa baseada em lógica paraconsistente combinada com classificadores inteligentes. A DPM consiste em um algoritmo de aprendizado semi-supervisionado, com algumas similaridades em relação às SVMs e outras técnicas semelhantes, são consideradas duas importantes questões: a formulação de uma função objetiva que restringe parâmetros que induzem a classificações incorretas, assim como a habilidade de lidar com as contradições e inconsistências. A DPM foi especificamente modelada para resolver problemas no tocante ao reconhecimento de padrões que envolvem contradições e inconsistências, não sendo resolvidas por meio da lógica tradicional. Para entender melhor a aplicabilidade da DPM, considera-se

uma situação na qual um paciente procura um médico com suspeita de que há algum problema com a sua garganta e/ou cordas vocais. Para a obtenção de um pré-diagnóstico, o médico grava o som de uma vogal sustentada durante alguns segundos. Após a gravação, o médico pode extrair algumas características da voz, sendo que o *software* utilizado gera um relatório final com as características analisadas.

SVMs, *Radial Base Functions* (RBF) e árvores de decisão (41), além de outras técnicas, certamente poderiam implementar uma solução para o contexto proposto no parágrafo anterior, porém o uso da DPM não forneceria os mesmos resultados. Isso ocorre devido ao fato de as técnicas citadas considerarem apenas duas possibilidades: “normal” ou “patológica”.

Considerando o problema destacado, a DPM promove as seguintes classificações:

- normal;
- patológica;
- normal tendendo à patológica;
- patológica com tendência à normalidade;

A DPM é uma ferramenta que recebe um sinal de dimensão X como entrada e produz sete saídas numéricas, correspondendo à solução esperada para o problema que está sendo tratado. Particularmente, o primeiro valor numérico corresponde à classificação principal, que pode ser: a entrada pertence a classe C_1 (normal, por exemplo), a classe C_2 (patológica, por exemplo), ambas ou nenhuma. O segundo e o terceiro valores de saída correspondem às coordenadas no espaço paraconsistente, comentado no capítulo que descreve a lógica paraconsistente como reticulado de Hasse. O quarto e quinto parâmetros correspondem ao grau de crença e descrença, assumindo que determinada entrada pertence a classe C_1 ou C_2 . O sexto parâmetro expressa o grau de crença nas classes C_1 e C_2 . O último parâmetro indica o grau de não pertencente a C_1 nem C_2 .

3.2 Arquitetura da DPM

A arquitetura da DPM está ilustrada na Figura 3.1. Seus elementos foram inspirados em alguns modelos de Redes Neurais Artificiais (RNAs), seguindo as seguintes características:

- a camada inicial contém X unidades passivas (*Passive Units* - PUs), cada uma com sua respectiva entrada alimentada por valores externos, além de T saídas que estão conectadas com as unidades de balanço discriminativo (*Discriminative Balance Units* - DBUs) da camada discriminativa;
- A camada discriminativa contém T DBUs com X entradas, cada qual associada a um respectivo peso $h_{i,j}$, sendo j o j -ésimo peso da unidade i , ($0 \leq i \leq X - 1$) e ($0 \leq j \leq T - 1$). Cada DBU tem uma saída conectada com a entrada de ambas as unidades de semi-decisão (*Semi-Decison Units* - SDUs) da camada de ajuste;
- A camada de ajuste contém dois SDUs (SDU_1 e SDU_2) com T entradas e uma saída para cada uma, sendo g_{i,C_1} e g_{i,C_2} ($0 \leq i \leq T - 1$), respectivamente, seus pesos. As saídas da SDU_1 e SDU_2 conectam apenas uma saída paraconsistente (*Paraconsistent Output Unit* - POU) na camada final;
- A camada final contém um POU com duas entradas e uma saída. Os pesos associados para ambas as entradas são sempre 1. O componente desta camada é responsável pela produção do resultado final.

Finalizando a estrutura da DPM, duas condições devem ser seguidas:

- T é igual ao número de exemplos de treinamento que serão utilizados para treinar a máquina, sendo X a dimensão de cada exemplo;
- $T > X$ é uma exigência, sendo $T \gg X$ a tendência para os casos reais de reconhecimento de padrões.

As unidades de processamento de cada uma das camadas já mencionadas foram inspiradas nas RNAs, embora cada unidade não deve ser considerada como sendo um neurônio.

Pelo contrário, elas devem ser consideradas como elementos de processamento ativos ou passivos, sendo que o conjunto de todos os elementos formarão uma estrutura inteligente, ou seja, a DPM. Essa distinção se deve ao fato de que um único neurônio artificial tem a capacidade de aprender, de uma maneira isolada e com características limitadas, mas apenas uma unidade da DPM não tem essa capacidade. Assim sendo, a nova máquina proposta deve considerar toda a estrutura e unidades que dela participam. Nas próximas sessões será discutido como cada uma das unidades funciona, assim como exibido na Figura 3.1.

Figura 3.1- A estrutura proposta para a DPM. Os círculos brancos representam as PUs que formam a camada inicial. Os círculos negros representam as DBUs, que formam a camada discriminativa. Os quadrados cinza com um asterisco representam as SDUs e o círculo com um "O", representa a POU, pertencente à última camada, que produz a saída da DPM.

3.2.1 Camadas

1-Camada Inicial

Cada PU da camada inicial encaminha suas entradas para todos os DBUs da camada discriminativa para posterior processamento, sendo considerados elementos simbólicos. A j -ésima amostra do t -ésimo exemplo de treinamento, que será enviado para a j -ésima PU, é representado por $x_{j,t}$, com $(0 \leq j \leq X - 1)$ e $(0 \leq t \leq T - 1)$.

2-Camada Discriminativa

As DBUs da camada discriminativa são ativas, isto é, elas processam os valores presentes nas entradas para a produção das respectivas saídas, como exibido na equação 3.1, sendo $c_{y,t}$ a saída do y -ésima unidade do t -ésimo exemplo de treinamento e $(0 \leq y, t \leq T - 1)$.

$$c_{y,t} = \sum_{k=0}^{X-1} x_{k,t} h_{y,k} \quad . \quad (3.1)$$

Inspirada em algumas implementações das SVMs, a camada discriminativa da DPM tem como primeiro objetivo aumentar a dimensão das informações de entrada, de X para T , o que produzirá, de forma simples, a separação entre as classes.

O segundo objetivo é que a t -ésima DBU tenha como saída a máxima resposta para o t -ésimo exemplo de treinamento e mínima resposta aos demais, sendo $(0 \leq t \leq T - 1)$. Deixando a mínima e máxima resposta assumir, respectivamente, os valores -1 e 1 , tem-se que, para a t -ésima unidade, o sistema linear com T equações e X incógnitas será:

$$c_{y,t} = \sum_{k=0}^{X-1} x_{k,t} h_{y,k} = \begin{cases} 1 & \text{for } y = t. \\ -1 & \text{for all } y \in [0, T - 1], y \neq t. \end{cases} \quad (3.2)$$

Cada um dos T sistemas gerados pela variação de y da Equação 3.2, consiste em uma formulação discriminativa adotada pela função objetivo de cada unidade correspondente. Essa formulação, contudo, apresenta o seguinte inconveniente: pode não existir uma solução exata

para o sistema, em razão de o número de equações (T) ser maior que o número de variáveis (X).

As seguintes matrizes demonstram um sistema exemplo:

$$\underbrace{\begin{pmatrix} x_{0,0} & x_{1,0} & x_{2,0} & \dots & \dots & x_{X-2,0} & x_{X-1,0} \\ x_{0,1} & x_{1,1} & x_{2,1} & \dots & \dots & x_{X-2,1} & x_{X-1,1} \\ x_{0,2} & x_{1,2} & x_{2,2} & \dots & \dots & x_{X-2,2} & x_{X-1,2} \\ \vdots & \vdots & \vdots & & & \vdots & \vdots \\ x_{0,T-2} & x_{1,T-2} & x_{2,T-2} & \dots & \dots & x_{X-2,T-2} & x_{X-1,T-2} \\ x_{0,T-1} & x_{1,T-1} & x_{2,T-1} & \dots & \dots & x_{X-2,T-1} & x_{X-1,T-1} \end{pmatrix}}_{\text{matriz } A_{[T][X]}} \cdot \underbrace{\begin{pmatrix} h_{y,0} \\ h_{y,1} \\ h_{y,2} \\ \vdots \\ h_{y,X-2} \\ h_{y,X-1} \end{pmatrix}}_{\text{vetor } B_{[T]}} = \underbrace{\begin{pmatrix} c_{y,t} \\ c_{y,t} \\ c_{y,t} \\ \vdots \\ c_{y,t} \\ c_{y,t} \end{pmatrix}}_{\text{vetor } C_{[T]}} . \quad (3.3)$$

Definindo $A_{[X][T]}^*$ como sendo a transposta de $A_{[T][X]}$, tem-se o seguinte sistema linear:

$$A_{[X][T]}^* \cdot A_{[T][X]} \cdot B_{[T]} = A_{[X][T]}^* \cdot C_{[T]} \quad (3.4)$$

tendo um total de X equações com X incógnitas cada uma, e apresentando solução exata que se aproxima da equação do sistema original da equação 3.2. Esse ajuste estatístico, que corresponde ao método dos Mínimos Quadrados para resolver equações de sistemas lineares incompatíveis (46), torna-se viável por três razões. Primeiramente, a DPM sendo treinada não estará sujeita ao *over-fitting*, que corresponde à situação na qual o classificador começa a se tornar muito especializado para os exemplos de treinamento, produzindo um modelo para futuros testes. Segundo, a solução obtida adiciona na DPM a capacidade de generalização. Por último, o sistema sempre convergirá para uma solução que é independente de procedimentos iterativos.

3-Camada de Ajuste

A camada de ajuste, a qual necessita de dois procedimentos para treinamento supervisionado, é responsável pela produção de dois valores, I_{C_1} e I_{C_2} , cada um gerado por uma unidade SDU. O primeiro valor representa uma decisão provisória, assumindo que o classificador foi treinado baseado em rótulos binários, os quais descrevem quando determinado exemplo de treinamento pertence ou não à classe C_1 . O último valor representa outra decisão provisória, mas com rótulo descrevendo quando este elemento pertence à classe C_2 . Os cálculos de I_{C_1} e I_{C_2} são os seguintes:

$$I_{C_1} = \sum_{k=0}^{P-1} \Phi(c_{k,t}) g_{k,C1} \quad \text{and} \quad I_{C_2} = \sum_{k=0}^{P-1} \Phi(c_{k,t}) g_{k,C2} \quad , \quad (3.5)$$

sendo $\Phi(a) = \frac{1}{1+e^{-a}}$ a função de treinamento adotada para manter $c_{k,t}$ delimitado, isto é, $(0 \leq c_{k,t} \leq 1)$. Para encontrar a equação 3.7, os pesos $g_{k,C1}$ e $g_{k,C2}$, $(0 \leq k \leq P-1)$ são determinados durante a fase de treinamento. Eles correspondem, respectivamente, à solução do seguinte sistema linear com T equações e T incógnitas:

$$\begin{pmatrix} \Phi(c_{0,0}) & \Phi(c_{1,0}) & \Phi(c_{2,0}) & \dots & \dots & \dots & \Phi(c_{P-2,0}) & \Phi(c_{P-1,0}) \\ \Phi(c_{0,1}) & \Phi(c_{1,1}) & \Phi(c_{2,1}) & \dots & \dots & \dots & \Phi(c_{P-2,1}) & \Phi(c_{P-1,1}) \\ \Phi(c_{0,2}) & \Phi(c_{1,2}) & \Phi(c_{2,2}) & \dots & \dots & \dots & \Phi(c_{P-2,2}) & \Phi(c_{P-1,2}) \\ \vdots & \vdots & \vdots & & & & \vdots & \vdots \\ \Phi(c_{0,T-2}) & \Phi(c_{1,T-2}) & \Phi(c_{2,T-2}) & \dots & \dots & \dots & \Phi(c_{P-2,T-2}) & \Phi(c_{P-1,T-2}) \\ \Phi(c_{0,T-1}) & \Phi(c_{1,T-1}) & \Phi(c_{2,T-1}) & \dots & \dots & \dots & \Phi(c_{P-2,T-1}) & \Phi(c_{P-1,T-1}) \end{pmatrix} \cdot \begin{pmatrix} g_{0,C1} \\ g_{1,C1} \\ g_{2,C1} \\ \vdots \\ g_{P-2,C1} \\ g_{P-1,C1} \end{pmatrix} = \begin{pmatrix} s_{0,C1} \\ s_{1,C1} \\ s_{2,C1} \\ \vdots \\ s_{P-2,C1} \\ s_{P-1,C1} \end{pmatrix}, \quad (3.6)$$

e

$$\begin{pmatrix} \Phi(c_{0,0}) & \Phi(c_{1,0}) & \Phi(c_{2,0}) & \dots & \dots & \dots & \Phi(c_{P-2,0}) & \Phi(c_{P-1,0}) \\ \Phi(c_{0,1}) & \Phi(c_{1,1}) & \Phi(c_{2,1}) & \dots & \dots & \dots & \Phi(c_{P-2,1}) & \Phi(c_{P-1,1}) \\ \Phi(c_{0,2}) & \Phi(c_{1,2}) & \Phi(c_{2,2}) & \dots & \dots & \dots & \Phi(c_{P-2,2}) & \Phi(c_{P-1,2}) \\ \vdots & \vdots & \vdots & & & & \vdots & \vdots \\ \Phi(c_{0,T-2}) & \Phi(c_{1,T-2}) & \Phi(c_{2,T-2}) & \dots & \dots & \dots & \Phi(c_{P-2,T-2}) & \Phi(c_{P-1,T-2}) \\ \Phi(c_{0,T-1}) & \Phi(c_{1,T-1}) & \Phi(c_{2,T-1}) & \dots & \dots & \dots & \Phi(c_{P-2,T-1}) & \Phi(c_{P-1,T-1}) \end{pmatrix} \cdot \begin{pmatrix} g_{0,C2} \\ g_{1,C2} \\ g_{2,C2} \\ \vdots \\ g_{P-2,C2} \\ g_{P-1,C2} \end{pmatrix} = \begin{pmatrix} s_{0,C2} \\ s_{1,C2} \\ s_{2,C2} \\ \vdots \\ s_{P-2,C2} \\ s_{P-1,C2} \end{pmatrix}, \quad (3.7)$$

sendo $s_{i,C1}$ e $s_{i,C2}$, $(0 \leq i \leq P-1)$, dois grupos de rótulos que informarão as SDUs quando o t -ésimo exemplo de treinamento pertence ou não a cada classe, C_1 e C_2 , respectivamente. Para ambos $s_{i,C1}$ e $s_{i,C2}$, os valores de rótulo 1 e -1 , respectivamente, são usados para uma indicação positiva e negativa. Embora ambos os grupos de rótulos sejam complementares, e dessa forma um possa ser descartado, existem algumas circunstâncias nas quais um ou mais exemplos de treinamento pertencem a mais de uma classe.

Também existe a possibilidade de utilizar qualquer valor entre -1 e 1 para os rótulos acima mencionados, em vez de apenas -1 e 1 . Nesse caso, tais valores descrevem o grau de certeza ou não, da entrada, sendo -1 e 1 os valores mínimo e máximo. Isso é fundamental para os casos em que a DPM é usada para tomar uma decisão considerando inconsistência e contradição, que é um dos seus objetivos básicos.

4-Camada Final

O único POU não necessita qualquer procedimento de treinamento e é a unidade que possibilita à DPM executar as classificações propostas. Essa unidade produz a saída final da DPM, sendo acionada a partir de dois valores recebidos da camada de ajuste, baseando seu

funcionamento no mecanismo com o qual a LPAv2 lida. Particularmente, a POU considera os valores recebidos de ambos os SDUs para decidir se a entrada pertence à C_1 , a C_2 , a C_1 , mas tendendo para C_2 , ou a C_2 mas tendendo para C_1 . Essa decisão é baseada no grau de crença (γ_1) e grau de descrença (γ_2) para cada classificação, sendo esses valores medições independentes não complementares. Os valores γ_1 e γ_2 , ($0 \leq \gamma_1, \gamma_2 \leq 1$), são definidos com base nas saídas de ambos os SDUs, como segue:

$$\gamma_1 = e^{-|I_{C1}-1|} \quad (3.8)$$

$$\gamma_2 = e^{-|I_{C2}-1|} \quad (3.9)$$

Particularmente, γ_1 descreve o grau de crença em que a entrada da DPM pertence à classe C_1 . Por outro lado, γ_2 descreve o grau de descrença em que a mesma entrada pertence a C_1 . Seis coeficientes são encontrados, utilizando os seguintes cálculos:

$$G_1 = \gamma_1 - \gamma_2 \quad , \quad (3.10)$$

($-1 \leq G_1 \leq 1$), sendo chamado de coeficiente de certeza, que descreve o nível de certeza de que a entrada da DPM pertence a C_1 ,

$$G_2 = \gamma_1 + \gamma_2 - 1 \quad , \quad (3.11)$$

($-1 \leq G_2 \leq 1$), o qual é chamado de coeficiente de contradição, que descreve o nível de contradição da entrada da DPM pertencer a C_1 ,

$$R_1 = 1 - \frac{D_1}{D_1 + D_2 + D_3 + D_4} \quad , \quad (3.12)$$

($0 \leq R_1 \leq 1$), sendo chamada de crença em C_1 ,

$$R_2 = 1 - \frac{D_2}{D_1 + D_2 + D_3 + D_4} \quad , \quad (3.13)$$

($0 \leq R_2 \leq 1$), sendo chamada de crença em C_2 ,

$$R_3 = 1 - \frac{D_3}{D_1 + D_2 + D_3 + D_4}, \quad (3.14)$$

($0 \leq R_3 \leq 1$), sendo chamada de crença em C_1 e C_2 , e

$$R_4 = 1 - \frac{D_4}{D_1 + D_2 + D_3 + D_4}, \quad (3.15)$$

($0 \leq R_4 \leq 1$), sendo chamada de crença nem em C_1 e nem em C_2 . Nas equações 3.12, 3.13, 3.14, e 3.15, $D_1 = \sqrt{(G_1 - 1)^2 + G_2^2}$, $D_2 = \sqrt{(G_1 + 1)^2 + G_2^2}$, $D_3 = \sqrt{G_1^2 + (G_2 - 1)^2}$ e $D_4 = \sqrt{G_1^2 + (G_2 + 1)^2}$ correspondendo às distâncias entre os pontos $(1, 0)$, $(-1, 0)$, $(0, 1)$, respectivamente, ao ponto (G_1, G_2) no plano paraconsistente (reticulado de Hasse). Após executar os cálculos anteriores, o resultado final da classificação é gerado de acordo com o algoritmo A-1.

Algoritmo A-1: Procedimentos utilizados pela POU para classificação de um vetor de entrada

INÍCIO

MARQUE o ponto (G_1, G_2) no plano paraconsistente e calcule as distâncias $D_1, \dots, D_4$ e as crenças $R_1, \dots, R_4$.

SE $(R_1 > R_2)$ e $(R_1 > R_3)$ e $(R_1 > R_4)$ então

$saída \leftarrow \{-1, G_1, G_2, R_1, R_2, R_3, R_4\}$, o que significa que o vetor de entrada tende a classe C_1 ;

SENÃO SE $(R_2 > R_1)$ e $(R_2 > R_3)$ e $(R_2 > R_4)$ então

$saída \leftarrow \{1, G_1, G_2, R_1, R_2, R_3, R_4\}$, o que significa que o vetor de entrada tende a classe C_2 ;

SENÃO SE $(R_3 > R_1)$ e $(R_3 > R_2)$ e $(R_3 > R_4)$ então

$saída \leftarrow \{2, G_1, G_2, R_1, R_2, R_3, R_4\}$, o que significa que o vetor de entrada tende a classe C_1 e C_2 ;

SENÃO

$saída \leftarrow \{0, G_1, G_2, R_1, R_2, R_3, R_4\}$, o que significa que o vetor de entrada não tende nem a classe C_1 nem a C_2 ;

FIM DO SENÃO;

FIM.

3.2.2 Treinamento da DPM

A-2 é o algoritmo semi-supervisionado para treinamento da DPM, sendo o passo T_{r1} não supervisionado e o passo T_{r2} a única etapa supervisionada. A camada inicial apenas encaminha os valores presentes nas unidades passivas X , sem mais atividades, sendo que o treinamento se inicia na camada discriminativa. Os requisitos são: T exemplos de treinamentos de dimensão X cada um, T rótulos binários para as saídas (1, -1, ou valores dentro desse intervalo) descrevendo a qual classe pertence cada exemplo.

Algoritmo A-2: Processo de Treinamento da DPM

INÍCIO

Passo T_{r1} - Treinamento da camada discriminativa: Para cada um dos T exemplos de treinamento, utilizar as equações 3.3 e 3.4 para ajustar os pesos h_{ji} . Depois, utilizar T vezes a equação 3.1 para encontrar o grupo c_y ;

Passo T_{r2} - Treinamento na camada de ajuste: Resolver as equações 3.6 e 3.7 para reconhecer os grupos $g_{i,C1}$ e $g_{i,C2}$;

FINAL

3.2.3 Usando a DPM no Reconhecimento de Padrões

O algoritmo A-3 foi adotado para utilizar uma DPM treinada para classificar um vetor de dimensão X .

Algoritmo A-3: Processo de utilização de uma DPM treinada para reconhecimento de padrões

INÍCIO

Passo U_{s1} - Utilizar a equação 3.1 T vezes até obter o grupo c_y ;

Passo U_{s2} Utilizar a equação 3.5 para obter I_{C1} e I_{C2} ;

Passo U_{s3} Utilizar as equações 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, 3.14 e 3.15. Finalmente, a saída deverá ser encontrada aplicando o algoritmo A-1;

FINAL

3.3 Exemplo Numérico

Para ilustrar os procedimentos de teste e treinamento, utilizou-se um exemplo básico com valores numéricos. Uma classificação real necessitaria de muitos exemplos de treinamentos, sendo este apenas uma demonstração da técnica.

3.3.1 Problema

O Problema consiste em diferenciar as classes C_1 e C_2 com o uso da DPM. A primeira classe contém um vetor de três dimensões no qual seu terceiro elemento corresponde à soma do primeiro e do segundo. Por outro lado, a outra classe contém um vetor no qual o terceiro elemento corresponde à diferença entre o primeiro e o segundo. Os seguintes exemplos

foram utilizados para treinar a DPM: $E_1 = 0, 1, 1$, $E_2 = 1, 0, 1$, $E_3 = 1, 1, 2$, $E_4 = 0, 1, -1$ e $E_5 = 1, 1, 0$. Um ponto interessante é que E_2 pertence a ambas as classes, combinando com as duas classificações.

3.3.2 Solução

Uma DPM com $X = 3$ entradas e $T = 5$ DBUs é necessária, pois existem cinco exemplos de treinamento de dimensão três. De acordo com o algoritmo A-2, que exhibe o processo de treinamento, serão obtidos cinco sistemas lineares com o uso da equação 3.3 para $0 \leq y \leq 4$.

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} h_{0,0} \\ h_{0,1} \\ h_{0,2} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}; \quad \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} h_{1,0} \\ h_{1,1} \\ h_{1,2} \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -1 \\ -1 \\ -1 \end{pmatrix}; \quad \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} h_{2,0} \\ h_{2,1} \\ h_{2,2} \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 1 \\ -1 \\ -1 \end{pmatrix};$$

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} h_{3,0} \\ h_{3,1} \\ h_{3,2} \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} h_{4,0} \\ h_{4,1} \\ h_{4,2} \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \\ 1 \end{pmatrix}.$$

Então, com o uso da equação 3.4 os sistemas serão convertidos da seguinte forma:

$$\begin{pmatrix} 3 & 2 & 3 \\ 2 & 4 & 2 \\ 3 & 2 & 7 \end{pmatrix} \cdot \begin{pmatrix} h_{0,0} \\ h_{0,1} \\ h_{0,2} \end{pmatrix} = \begin{pmatrix} -3 \\ -2 \\ -1 \end{pmatrix}; \quad \begin{pmatrix} 3 & 2 & 3 \\ 2 & 4 & 2 \\ 3 & 2 & 7 \end{pmatrix} \cdot \begin{pmatrix} h_{1,0} \\ h_{1,1} \\ h_{1,2} \end{pmatrix} = \begin{pmatrix} -1 \\ -4 \\ -1 \end{pmatrix}; \quad \begin{pmatrix} 3 & 2 & 3 \\ 2 & 4 & 2 \\ 3 & 2 & 7 \end{pmatrix} \cdot \begin{pmatrix} h_{2,0} \\ h_{2,1} \\ h_{2,2} \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix};$$

$$\begin{pmatrix} 3 & 2 & 3 \\ 2 & 4 & 2 \\ 3 & 2 & 7 \end{pmatrix} \cdot \begin{pmatrix} h_{3,0} \\ h_{3,1} \\ h_{3,2} \end{pmatrix} = \begin{pmatrix} -3 \\ -2 \\ -5 \end{pmatrix}; \quad \text{e} \quad \begin{pmatrix} 3 & 2 & 3 \\ 2 & 4 & 2 \\ 3 & 2 & 7 \end{pmatrix} \cdot \begin{pmatrix} h_{4,0} \\ h_{4,1} \\ h_{4,2} \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ -3 \end{pmatrix},$$

cujas soluções são as seguintes:

$$\begin{aligned} h_{0,0} = -1.500, \quad h_{0,1} = 0, \quad h_{0,2} = 0.500 & ; \quad h_{1,0} = 0.500, \quad h_{1,1} = -1.250, \quad h_{1,2} = 0; \\ h_{2,0} = -0.500, \quad h_{2,1} = -0.500, \quad h_{2,2} = 0.500 & ; \quad h_{3,0} = -0.500, \quad h_{3,1} = 0, \quad h_{3,2} = -0.500; \\ e & ; \quad h_{4,0} = 0.500, \quad h_{4,1} = -0.500, \quad h_{4,2} = -0.500. \end{aligned}$$

Aplicando a equação 3.1 sobre os valores $h_{j,i}$, para $0 \leq t, y \leq 4$, tem-se:

$$\begin{aligned} c_{0,0} = 0.500, \quad c_{1,0} = -1.250, \quad c_{2,0} = 0, \quad c_{3,0} = -0.500, \quad c_{4,0} = -1.000; \\ c_{0,1} = -1.000, \quad c_{1,1} = 0.500, \quad c_{2,1} = 0, \quad c_{3,1} = -1.000, \quad c_{4,1} = 0; \\ c_{0,2} = -0.500, \quad c_{1,2} = -0.750, \quad c_{2,2} = 0, \quad c_{3,2} = -1.500, \quad c_{4,2} = -1.000; \\ c_{0,3} = -0.500, \quad c_{1,3} = -1.250, \quad c_{2,3} = -1.000, \quad c_{3,3} = 0.500, \quad c_{4,3} = 0; \\ c_{0,4} = -1.500, \quad c_{1,4} = -0.750, \quad c_{2,4} = -1.000, \quad c_{3,4} = -0.500, \quad c_{4,4} = 0. \end{aligned}$$

O próximo passo é utilizar as equações 3.6 e 3.7 para encontrar respectivamente:

$$\begin{pmatrix} 0.62 & 0.22 & 0.50 & 0.38 & 0.27 \\ 0.27 & 0.62 & 0.50 & 0.27 & 0.50 \\ 0.38 & 0.32 & 0.50 & 0.18 & 0.27 \\ 0.38 & 0.22 & 0.27 & 0.62 & 0.50 \\ 0.18 & 0.32 & 0.27 & 0.38 & 0.50 \end{pmatrix} \cdot \begin{pmatrix} g_{0,C_1} \\ g_{1,C_1} \\ g_{2,C_1} \\ g_{3,C_1} \\ g_{4,C_1} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{pmatrix} e \begin{pmatrix} 0.62 & 0.22 & 0.50 & 0.38 & 0.27 \\ 0.27 & 0.62 & 0.50 & 0.27 & 0.50 \\ 0.38 & 0.32 & 0.50 & 0.18 & 0.27 \\ 0.38 & 0.22 & 0.27 & 0.62 & 0.50 \\ 0.18 & 0.32 & 0.27 & 0.38 & 0.50 \end{pmatrix} \cdot \begin{pmatrix} g_{0,C_2} \\ g_{1,C_2} \\ g_{2,C_2} \\ g_{3,C_2} \\ g_{4,C_2} \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \\ 1 \end{pmatrix}.$$

O lado direito do primeiro sistema implica que o primeiro, segundo e terceiro exemplos de treinamento pertencem a C_1 e o quarto e quinto, não. Por outro lado, o lado direito do segundo sistema implica que o segundo, quarto e quinto exemplos de treinamento pertencem a classe C_2 e o primeiro e terceiro, não. Os sistemas têm, respectivamente, as seguintes soluções:

$$\begin{aligned} g_{0,C_1} = 1.37, \quad g_{1,C_1} = 6.14, \quad g_{2,C_1} = 0.77, \quad g_{3,C_1} = 1.37, \quad e \quad g_{4,C_1} = -7.89; \\ g_{0,C_2} = 1.37, \quad g_{1,C_2} = 6.14, \quad g_{2,C_2} = -7.89, \quad g_{3,C_2} = 1.37, \quad e \quad g_{4,C_2} = 0.77. \end{aligned}$$

Considerando os cálculos anteriores, o processo de treinamento da DPM está finalizado.

3.3.3 Testando a DPM treinada

Quatro vetores de entrada aplicados ao algoritmo A-3 são utilizados para testar a DPM. O primeiro é $\{1, 0, 1\}$ pertencente ao grupo de treinamento; o segundo, terceiro e quarto, não pertencem a este grupo, sendo respectivamente $\{3, 3, 3\}$, $\{2, 2, 4\}$ e $\{2, 2, 0\}$. Testando o vetor $\{1, 0, 1\}$, produzem-se os seguintes resultados: $c_{0,\text{vetor}\{1,0,1\}} = -1.00$; $c_{1,\text{vetor}\{1,0,1\}} = 0.50$, $c_{2,\text{vetor}\{1,0,1\}} = 0$, $c_{3,\text{vetor}\{1,0,1\}} = -1.00$ e $c_{4,\text{vetor}\{1,0,1\}} = 0$. Na sequência, a equação 3.5 produz $I_{C_1} = 1$ e $I_{C_2} = 1$. Ao final, as equações 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, 3.14 e 3.15 produzem $\gamma_1 = 1$, $\gamma_2 = 1$, $G_1 = 0$, $G_2 = 1$, $R_1 = 0.707$, $R_2 = 0.707$, $R_3 = 1$, e $R_4 = 0.5856$. De acordo com o algoritmo A-1, as saídas serão $\{2, 0, 1, 0.707, 0.707, 1, 0.586\}$, isto é, o vetor de entrada combina com ambas as classes. O ponto (G_1, G_2) é representado na Figura 3.2.

Figura 3.2- Saída da DPM para o experimento com o vetor de entrada $\{1, 0, 1\}$. O círculo (o) representa a localização do ponto $[G_1, G_2]$, que é exatamente $[0, 1]$.

Ao testar o vetor $\{3, 3, 3\}$, deve-se utilizar o procedimento citado anteriormente, o qual produzirá os seguintes resultados: $c_{0,\text{vetor}\{3,3,3\}} = -3.00$, $c_{1,\text{vetor}\{3,3,3\}} = -2.25$, $c_{2,\text{vetor}\{3,3,3\}} = -1.50$, $c_{3,\text{vetor}\{3,3,3\}} = -3.00$ e $c_{4,\text{vetor}\{3,3,3\}} = -1.50$. Na sequência, a equação 3.5 produz $I_{C_1} = -0.583396$ e $I_{C_2} = -0.583396$. Ao final, as equações 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, 3.14, e 3.15 produzem $\gamma_1 = 0.205277$, $\gamma_2 = 0.205277$, $G_1 = 0$, $G_2 = -0.589446$, $R_1 = 0.731396$, $R_2 = 0.731396$, $R_3 = 0.632208$ e $R_4 = 0.904999$. De acordo com o algoritmo A-1, as saídas serão $\{0, 0, -0.589, 0.731, 0.731, 0.632, 0.905\}$, isto é, o vetor de entrada combina com ambas

as classes, mas não exatamente. Este é um resultado interessante. O vetor de entrada está no meio das classes C_1 e C_2 porque $3 + 3 = 6$ e $3 - 3 = 0$. Esta é a razão pela qual a certeza, que pode variar entre -1 e 1 , resulta em 0 , isto é, a metade do intervalo. Nesse caso, o plano paraconsistente encontra-se na Figura 3.3.

Figura 3.3- Saída da DPM para o experimento com o vetor de entrada $\{3, 3, 3\}$. O círculo (o) dentro do losango representa a localização do ponto $[G_1, G_2]$, que corresponde a $[0, -0.589446]$.

Testando o vetor $\{2, 2, 4\}$, deve-se utilizar o procedimento citado anteriormente, o que produzirá os seguintes resultados: $c_{0, \text{vetor}\{2,2,4\}} = -1$, $c_{1, \text{vetor}\{2,2,4\}} = -1.5$, $c_{2, \text{vetor}\{2,2,4\}} = 0$, $c_{3, \text{vetor}\{2,2,4\}} = -3$, e $c_{4, \text{vetor}\{2,2,4\}} = -2$. Na sequência, a equação 3.5 produz $I_{C_1} = 0.99737$ e $I_{C_2} = -2.29874$. Ao final, as equações 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, 3.14 e 3.15 produzem $\gamma_1 = 0.997373$, $\gamma_2 = 0.0369297$, $G_1 = 0.960$, $G_2 = 0.034$, $R_1 = 0.989061$, $R_2 = 0.590365$, $R_3 = 0.715455$, and $R_4 = 0.705119$. De acordo com o algoritmo A-1, as saídas serão $\{-1, 0.960, 0.034, 0.989, 0.590, 0.715, 0.705\}$, isto é, o vetor de entrada combina com a classe C_1 . Nesse caso, o plano paraconsistente encontra-se na Figura 3.4.

Testando o vetor $\{2, 2, 0\}$, com a utilização o procedimento citado anteriormente, serão produzidos os seguintes resultados: $c_{0, \text{vetor}\{2,2,0\}} = -3$, $c_{1, \text{vetor}\{2,2,0\}} = -1.5$, $c_{2, \text{vetor}\{2,2,0\}} = -2$, $c_{3, \text{vetor}\{2,2,0\}} = -1$, e $c_{4, \text{vetor}\{2,2,0\}} = 0$. Na sequência, tem-se $I_{C_1} = -2.29874$ e $I_{C_2} = 0.99737$. Ao final, as equações 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, 3.14, e 3.15 produzem $\gamma_1 = 0.0369297$, $\gamma_2 = 0.997373$, $G_1 = -0.960$, $G_2 = 0.034$, $R_1 = 0.590365$, $R_2 = 0.989061$, $R_3 = 0.715455$, e $R_4 = 0.705119$. Com o algoritmo A-1, as saídas serão $\{1, -0.960, 0.034, 0.590, 0.989, 0.715, 0.705\}$, isto é, o vetor de entrada combina com a classe C_2 . Nesse caso, o plano paraconsistente

Figura 3.4- Saída da DPM para o experimento com o vetor de entrada $\{2, 2, 4\}$. O círculo (o) dentro do losango representa a localização do ponto $[G_1, G_2]$, que é exatamente $[0.960, 0.034]$.

encontra-se na Figura 3.5.

Figura 3.5- Saída da DPM para o experimento com o vetor de entrada $\{2, 2, 0\}$. O círculo (o) dentro do losango representa a localização do ponto $[G_1, G_2]$, que é exatamente $[-0.960, 0.034]$.

3.3.4 Analisando o Plano Paraconsistente

Os quatro exemplos apresentados na sessão anterior ilustram como um ponto (G_1, G_2) no plano paraconsistente pode ser interpretado. Além disso, exibem situações diferentes para a interpretação das saídas. Na Figura 3.2, cuja entrada foi o vetor $\{1, 0, 1\}$, evidencia-se que R_3 é o maior valor para crença entre as outras quatro possibilidades, isto é, de R_1 , R_2 e R_4 . Assim, o ponto (G_1, G_2) está próximo do ponto $(0, 1)$, vértice do losango, implicando que o vetor de entrada combina com ambas as classes. Quanto mais (G_1, G_2) se aproxima do ponto $(0, 1)$, mais

a combinação é exata, isto é, o vetor de entrada realmente faria parte de ambas as classes. Por outro lado, quanto mais (G_1, G_2) se aproxima do ponto $(0, -1)$, menos a combinação será exata, isto é, o vetor de entrada apenas se aproxima das duas classes, mas não combina com nenhuma perfeitamente, estando na fronteira de decisão entre ambas as classes. Esse é o caso do vetor $\{3, 3, 3\}$, que teve sua localização exibida na 3.3. Particularmente, se $G_1 = 0$ e $G_2 = -1$, o vetor de entrada não é compatível com nenhuma classe.

A figura 3.4, que apresenta o vetor $\{2, 2, 4\}$, tem seu ponto (G_1, G_2) próximo ao ponto $(1, 0)$, outro vértice do losango. Dessa forma, o vetor de entrada combina com a classe C_1 . Para qualquer situação, quanto mais o ponto (G_1, G_2) se aproxima do ponto $(1, 0)$, mais o vetor será compatível com a classe C_1 . Por outro lado, caso o ponto (G_1, G_2) estiver próximo ao ponto $(-1, 0)$, mais pertencerá ao padrão da classe C_2 . O último caso foi apresentado pelo vetor $\{2, 2, 0\}$, cujo plano paraconsistente foi apresentado na Figura 3.5.

3.4 Problemas na Detecção de Normalidade e Patologia

O melhor problema para demonstrar a aplicabilidade e utilidade da DPM é aquele que envolve a classificação de sinais biomédicos. Isso ocorre devido à presença de falso-negativos que possibilita considerações incorretas. Os testes descrevem a distinção entre vozes normais e patológicas. Isso não ocorre de uma forma tão simples quanto a classificação binária, mas sim com recursos extras, como mencionado anteriormente.

Capítulo 4

Materiais e Métodos

Neste capítulo são apresentados procedimentos, materiais e métodos, detalhando-se a base de dados utilizada e os procedimentos necessários para a obtenção dos valores utilizados como parâmetros de entrada para a DPM.

Este trabalho foi realizado no Laboratório *SpeechLab* do Instituto de Física de São Carlos (IFSC-USP), onde foram utilizados equipamentos para pesquisa, implementação e testes do projeto. O grupo de pesquisadores do laboratório possui diversos trabalhos publicados na área de saúde da voz, além de estrutura para experimentos voltados para processamento de áudio e voz.

Destaca-se também, o auxílio dos profissionais do Departamento de Otorrinolaringologia e Cirurgia de Cabeça e Pescoço do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP-USP), que disponibilizaram a base de vozes saudáveis e patológicas utilizada neste trabalho.

Tabela 4.1- Banco de 87 vozes recebido do Hospital das Clínicas da FMRP-USP.

Classificação	Sexo	Quantidade de indivíduos
Saudável (Normal)	Masculino	14
Saudável (Normal)	Feminino	38
Edema de Reinke	Masculino	1
Edema de Reinke	Feminino	8
Nódulo nas Pregas Vocais	Masculino	12
Nódulo nas Pregas Vocais	Feminino	14

4.1 Base de Dados

Uma das principais características deste trabalho foi o uso de um sistema de Aprendizado de Máquina, discutido nos capítulos anteriores. Assim, para uma boa taxa de reconhecimento é necessário o uso de uma variedade de exemplos de treinamento. Dessa forma, para criar os parâmetros de treinamento, utilizou-se um banco de dados com sinais de vozes com características apresentadas na Tabela 4.1.

O banco de dados contém fonemas do português brasileiro: /a/, /e/ e /i/. Para os experimentos foi utilizado o fonema /a/, pois, sustentado em potência confortável ao locutor, é considerado o melhor som vocálico para extração e medição dos parâmetros da voz (7). Os sinais foram armazenados seguindo o padrão wave PCM (47) amostrando as vozes a 22050Hz, 16 bits, mono-canal. Para a gravação, foi utilizado um microfone LeSon SM58, com características de banda larga unidirecional e controle dinâmico de ganho. Cada voz amostrada armazenou a sustentação do som vocálico por aproximadamente 5 segundos, em níveis confortáveis de potência, conforme um diálogo natural.

O processo de aquisição e classificação das vozes foi acompanhado por médicos da FMRP-USP, usando videolaringoscopia e videoestroboscopia para as confirmações. Um cuidado especial foi tomado para garantir que não haveria aumento da frequência fundamental ou mudanças que poderiam desestabilizar o sinal natural da fala, devido ao aumento anormal do

Tabela 4.2- Divisão das vozes, totalizando 342 sinais.

Classificação	Total de Sinais	Sinais para Treinamento	Sinais para Testes
Saudável (Normal)	142	50	92
Edema de Reinke	61	50	11
Nódulo nas pregas vocais	139	50	89

fluxo de ar pela glote. Cada uma das 87 vozes do banco foi dividida em três ou quatro segmentos, totalizando um aumento da quantidade de sinais utilizados em 342 trechos de vozes, conforme detalhado na Tabela 4.2.

4.2 Metodologia

Após o desenvolvimento e implementação da DPM, foi desenvolvido um algoritmo de extração de parâmetros, conforme descrito a seguir.

Cada um dos arquivos de voz foi pré-processado exatamente como recebido da FMRP-USP, com aproveitamento apenas da parte central, visando à redução da possibilidade de ruídos iniciais ou mesmo silêncio, garantindo a utilização eficiente das amostras da vogal armazenada.

As amostras acima foram divididas em segmentos, cada um com 16384 elementos. De cada um desses, foram extraídas as seguintes características:

- frequência fundamental (*pitch*);
- variância da dimensão fractal;
- variância das energias de sub-bandas;
- variância dos coeficientes cepstrais em escala Mel (MFCC);
- variância das primeiras derivadas dos valores do MFCC;
- variância das segundas derivadas dos valores do MFCC.

A maioria dos valores extraídos consiste em uma medida da variabilidade dos dados. A variância é baseada na diferença entre o valor de cada observação (x_i) e a média da população (μ). Neste trabalho, os dados são de uma população de elementos, sendo a média dos desvios elevados ao quadrado, chamada variância da população, denotada pelo símbolo grego σ^2 . Para uma população de N elementos e média μ , tem-se a equação 4.1.

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} \quad (4.1)$$

No caso de cada amostra, N continha o valor de 31, pois cada um dos segmentos de tamanho 16384 foi janelado com 1024 amostras (j_i), com uma sobreposição de 512 elementos (50%), de acordo com a Figura 4.1. Dessa forma, para a extração de cada característica, com exceção do *pitch*, foram utilizados grupos de 1024 elementos e a variância da população dos grupos foi tratada como a variância do segmento.

Figura 4.1- Janelamento de 1024 elementos.

O primeiro parâmetro fornecido para a DPM foi a variância da dimensão fractal de cada janela (j_i). Os valores obtidos na análise da geometria fractal de cada amostragem destacam justamente as irregularidades na forma do sinal verificado. Tais irregularidades contribuem para a avaliação da padronização do sinal sob análise.

O próximo parâmetro utilizado foi a frequência fundamental (F_0), ou seja o *pitch*. Para a extração do *pitch* foi utilizado o algoritmo baseado na análise cepstral (48), o qual foi aplicado nos elementos, porém sem a técnica do janelamento, cujo intuito era obter o valor para o sinal todo de 16384 amostras, e não a variância do mesmo. Na presente fase, estimou-se que o valor do *pitch* combinado com o sexo do locutor poderia indicar alguma patologia, uma vez

Tabela 4.3- Resumo dos parâmetros extraídos dos sinais de vozes.

Posição	Parâmetro
1	Variância da Dimensão Fractal
2	Valor do Pitch
3	Sexo (1 para Masculino e -1 para Feminino)
4 até 28	Variância da Energia de SubBanda
29 até 41	Variância dos Coeficientes MFCC
42 até 53	Variância da Primeira Derivada dos Coef. MFCC
54 até 64	Variância da Segunda Derivada dos Coef. MFCC

que cada classe, neste caso masculino e feminino, caracteriza-se em uma faixa de frequências reconhecidas (49). Assim, o terceiro parâmetro utilizado foi o sexo do locutor. Essa informação foi retirada da própria base de vozes, sendo utilizado o valor 1 para locutores do sexo masculino e -1 para locutores do sexo feminino.

Na sequência, os parâmetros de posição 4 a 28 são as energias das sub-bandas, sendo que uma variância elevada em cada sub-banda demonstraria dificuldades na sustentação de determinadas frequências, o que pode ser um indício de patologia.

O restante dos parâmetros foi baseado nos MFCCs, sendo que dos parâmetros 29 a 64, os 13 primeiros eram os valores da variância dos coeficientes MFCCs, os próximos 12 correspondem às derivadas primeiras e os 11 últimos correspondem às derivadas segundas.

Resumindo, um total de 64 parâmetros foi obtido para cada análise, listados na Tabela 4.3.

Mediante a obtenção dos 64 parâmetros para cada segmento de voz, eles foram utilizados para o treinamento e verificação do classificador desenvolvido.

Capítulo 5

Testes, Resultados e Discussões

Neste capítulo são descritos os resultados obtidos, bem como é apresentada uma discussão sobre os mesmos. Comparações com outros trabalhos na literatura também são realizadas, considerando, inclusive, obtenção e diferenças entre base de dados, equipamentos utilizados e condições de gravação.

Para a obtenção dos resultados da etapa de testes, com o uso da técnica proposta junto ao estudo para determinar qual a melhor combinação de parâmetros de entrada e quantidade destes, foi utilizado um algoritmo para avaliação da combinação dos parâmetros. Essa avaliação foi realizada para comparar tanto os parâmetros que, combinados, forneciam melhores resultados, como a quantidade de parâmetros. O algoritmo para verificação realizou a combinação automatizada de todos os parâmetros partindo do parâmetro inicial (variância da dimensão fractal) junto ao segundo parâmetro (valor do *pitch*) até a variância da segunda derivada do coeficiente 11 do MFCC, sempre utilizando somente dois parâmetros combinados. Em um próximo passo, o algoritmo executava a mesma bateria de testes anteriores, porém com o incremento de mais um parâmetro, ou seja, um a mais que o teste anterior, continuando a combinar os parâmetros até realizar a combinação dos sessenta e quatro parâmetros, conforme a tabela 5.1.

Considerando os resultados obtidos, a melhor média de reconhecimento encontrada foi a de 27 elementos, ou seja, independente de quais parâmetros, utilizando 27 foi possível atingir a média de 80,93% de reconhecimento entre as combinações de parâmetros, como exi-

Tabela 5.1- Combinações dos parâmetros

Número do Teste	Quantidade de Parâmetros	Quantidade de Testes
1	2 Elementos	2015
2	3 Elementos	1952
3	4 Elementos	1890
4	5 Elementos	1829
5	6 Elementos	1769
6	7 Elementos	1710
7	8 Elementos	1652
8	9 Elementos	1595
9	10 Elementos	1539
10	11 Elementos	1484
20	21 Elementos	989
30	31 Elementos	594
40	41 Elementos	299
50	51 Elementos	104
60	61 Elementos	9
61	62 Elementos	5
62	63 Elementos	2

bido na Tabela 5.2, sendo que com o uso da melhor combinação foi obtido 97%.

Analisando separadamente a combinação dos 27 parâmetros, que se mostrou ótima para o reconhecimento, foi observado que alguns comprometeram negativamente os resultados. A Figura 5.1 exibe um gráfico com o percentual de reconhecimento de cada teste. Nessa imagem é possível verificar que próximo ao teste 75 não haviam sido atingidos resultados superiores a 95% de reconhecimento; e após esse teste foram obtidos resultados melhores. Isto ocorreu após o teste de número 75, quando houve uma troca de parâmetros, sendo que o parâmetro variância de *pitch* foi substituído pelo parâmetro de variância de energia da sub-banda 24. Após essa troca, observou-se que foi atingido um padrão de estabilidade. Com relação à melhor combinação de reconhecimento, verifica-se, na Figura 5.1, que, após o teste 547, foi detectada uma queda no reconhecimento, assim como variações mais bruscas nos resultados. Tal fato ocorreu justamente após a substituição do parâmetro variância de energia da sub-banda 14 para variância da segunda derivada da energia da sub-banda 6, tornando-se estável novamente quando os parâmetros acústicos dos MFCCs foram utilizados, isto é, quando o grupo de

Tabela 5.2- Melhores resultados dos testes

Teste	Média de Reconhecimento (%)	Desvio Padrão Reconhecimento (%)
27 Elementos	80,93	22,45
28 Elementos	79,31	24,05
40 Elementos	78,21	26,81
44 Elementos	77,22	26,44
39 Elementos	76,71	27,32
43 Elementos	76,64	28,46
48 Elementos	76,51	25,05
29 Elementos	76,01	26,87
26 Elementos	75,94	27,30
49 Elementos	75,75	27,31

Tabela 5.3- Piores resultados dos testes

Teste	Média de Reconhecimento (%)	Desvio Padrão Reconhecimento (%)
62 Elementos	40,47	6,81
63 Elementos	45,00	2,17
61 Elementos	47,72	8,19
59 Elementos	54,06	21,67
60 Elementos	57,26	17,55
58 Elementos	59,08	24,38
57 Elementos	60,88	25,86
2 Elementos	63,97	18,81
54 Elementos	64,38	30,17
56 Elementos	65,52	27,78

parâmetros não utilizava as variâncias das energias das sub-bandas, passando a contar apenas com as variâncias dos MFCCs, suas primeiras e segundas derivadas.

Considerando a totalidade da média de reconhecimento, os piores resultados foram aqueles com a maior e a menor quantidade de parâmetros. Por meio da Figura 5.2 é possível visualizar a curva que descreve o comportamento do sucesso no reconhecimento. O uso de uma quantidade média de parâmetros mostrou-se ideal segundo os critérios de reconhecimento. A Tabela 5.3 exhibe os resultados menos favoráveis, sendo o resultado com menor média final, o teste com 62 elementos, que atinge uma média de 40,47%.

É possível verificar na Figura 5.3 o acúmulo dos melhores resultados em um bloco

Figura 5.1- Testes com 27 parâmetros.

centralizado, partindo da quantidade de 26 elementos a 48 elementos como a faixa ideal para o reconhecimento. A figura citada também apresenta uma queda expressiva após os testes com 50 elementos.

Foram exibidos nas tabelas 5.2 e 5.3 os respectivos valores dos desvios padrões obtidos nos resultados de reconhecimento considerados bons, isto é, acima de 73,4%, e ruins, isto é, abaixo de 70,91%. Os valores mais baixos para o desvio padrão significam que o conjunto de parâmetros resulta em uma avaliação homogênea para os testes realizados, independentemente do nível de reconhecimento. O baixo desvio padrão não foi um sinal de bom reconhecimento, mas sim uma proximidade no valor de reconhecimento obtido. Assim, para os vários testes em uma mesma população, os valores foram semelhantes. Uma característica importante foi que as combinações de parâmetros com menor qualidade de reconhecimento tiveram valores mais próximos, com menores valores de desvio padrão, sendo possível afirmar que os melhores resultados obtidos no reconhecimento apresentaram uma variação muito grande entre os ele-

Figura 5.2- Média de Reconhecimento

mentos avaliados.

Os resultados apresentados na Figura 5.4 exibem a tendência dos conjuntos de parâmetros com menos elementos ou mais elementos, obtendo uma melhor homogeneidade para reconhecimento; os grupos com quantidades intermediárias têm os valores de desvio padrão mais elevados, porém os melhores índices de reconhecimento.

Outra avaliação, quanto aos parâmetros, foi a influência constatada de alguns deles junto ao restante do grupo. O parâmetro dimensão fractal não apresentou uma contribuição significativa para o reconhecimento, e, além disso, nos agrupamentos que continham tal parâmetro foi observada uma queda no nível de reconhecimento. A Figura 5.5 exibe uma comparação entre a utilização dos parâmetros dimensão fractal, variância do período de *pitch* e sexo do locutor. O que se pode constatar é que os resultados de reconhecimento utilizando o sexo do locutor foram superiores aos com agrupamentos de outros dois parâmetros, seguidos por variância do período de *pitch* e dimensão fractal. Este último chegou a obter reconhecimento inferior próximo a 17%.

A contribuição do parâmetro não acústico do sexo do locutor deve ser destacada, uma vez que a combinação deste parâmetro com outros, relacionados com o primeiro, resulta em um critério de reconhecimento. Um exemplo dessa combinação é com o parâmetro acústico *pitch*, pois para muitos autores é possível detectar o gênero do locutor devido ao estabelecimento da

Figura 5.3- Média de Reconhecimento x Quantidade de Elementos.

sua faixa de frequência. Assim, é possível combinar um locutor em uma faixa específica e especificar em outro parâmetro não acústico a especificação do seu gênero. Dessa forma, a DPM será treinada associando os dois parâmetros. Tal fato demonstrou ser positivo para o reconhecimento, como observado na Figura 5.5, pelo conjunto de parâmetros com o sexo do Locutor. Finalmente, os resultados de parâmetros acústicos e não acústicos mostram respostas consideráveis. No entanto, considerando apenas os não acústicos com a quantidade de 27 elementos, observou-se um resultado bom para o reconhecimento.

Cabe ressaltar que, para cada um dos resultados obtidos foi possível localizar a saída da DPM no plano paraconsistente, conforme exemplificado para um dos sinais hipotéticos do capítulo 3. Isso fez com que fosse possível obter, não somente uma resposta do tipo “normal” ou “patológica”, dependendo de que lado, esquerdo ou direito, estava localizada a saída da DPM, mas também, o nível de normalidade ou patologia. Na hipótese de uma certa saída da DPM estar afastada dos vértices esquerdo ou direito, atenção deve ser dedicada para que exames médicos sejam realizados, principalmente quando a referida localização indicar normalidade com tendência para patologia. Adicionalmente, é importante relatar que não houve

Figura 5.4- Comparação entre Desvio Padrão e Reconhecimento.

Figura 5.5- Comparação de Parâmetros com um grupo de 27 características.

nenhum caso em que a saída da DPM estivesse localizada próxima ao vértice indicador de qual o sinal de voz não pertencia a nenhuma das classes, isto é, a voz em questão não era nem normal nem patológica.

Capítulo 6

Conclusões e sugestões para trabalhos futuros

Neste capítulo seguem as conclusões obtidas a partir dos resultados mostrados no capítulo anterior, destacando a aplicabilidade da DPM e dos parâmetros selecionados como entrada para ela. Além disso, sugestões para trabalhos futuros também estão listadas.

Pode-se dizer que a principal contribuição deste trabalho foi o desenvolvimento de um algoritmo que associa uma técnica baseada em aprendizado de máquina com lógica para-consistente anotada, promovendo uma classificação adequada para distinção de vozes normais e patológicas, fundamentalmente as de pacientes com nódulos nas pregas vocais e Edema de Reinke. Utilizando diferentes combinações de parâmetros, foi possível obter um reconhecimento de mais de 80%. Este trabalho também descreveu uma boa sequência de parâmetros para obtenção do resultado esperado, demonstrando que os parâmetros acústicos selecionados contribuíram de uma forma ímpar para a melhoria no reconhecimento.

A utilização da dimensão fractal descreveu o comportamento do sinal de forma importante, porém, combinada aos demais parâmetros acústicos, não se mostrou muito propícia para o reconhecimento dos sinais estudados. O parâmetro não acústico sexo, não se mostrou relevante quando combinado aos outros parâmetros. Segundo a literatura, o indivíduo de um

sexo específico, cuja voz é saudável, segue uma faixa específica de frequências, porém, esta não foi uma característica destacável após os testes. O uso desse parâmetro foi considerado para o experimento, pois, combinado ao parâmetro *pitch*, seria possível verificar a faixa de voz saudável, como comentado anteriormente.

Os resultados mostraram que a utilização apenas dos parâmetros adquiridos do sinal completo não demonstraria toda a capacidade de reconhecimento do modelo proposto, reforçado pela literatura, que descreve que uma das principais características das não conformidades na voz é justamente a sustentação dos elementos que a compõem. Assim, os melhores resultados foram obtidos com o uso da variância dos valores dos parâmetros em trechos janelados das vozes. A utilização da variância destaca-se como importante elemento relacionado com as energias das sub-bandas e coeficientes da escala Mel, juntamente com suas derivadas primeiras e segundas.

As combinações de parâmetros acústicos apresentaram os melhores resultados de reconhecimento, em especial os conjuntos com 26 e 27 elementos. A quantidade comentada, com um resultado sequencial partindo da terceira energia de sub-banda, foi a escolha que atingiu mais de 80% de reconhecimento médio para a base de vozes testadas, sendo um reconhecimento de 97% em diversos casos, ajustados principalmente com o uso da Lógica Paraconsistente Anotada.

A Lógica utilizada foi eficiente justamente por possibilitar a detecção e efetiva avaliação da qualidade dos parâmetros, pois em algumas combinações o nível atingido comprometeria a certeza do resultado ou acusaria um nível alto de contradição com relação ao apontamento do diagnóstico, que era contornado pelo algoritmo na diminuição do nível de afirmação quanto ao resultado do teste. É importante ressaltar que a análise da localização da saída da DPM no plano paraconsistente possibilitou uma indicação adicional àquela obtida com outros classificadores, conforme descrito anteriormente.

Em função disso, algumas sugestões para trabalhos que pretendem seguir a mesma linha no reconhecimento, merecem destaque. Alguns estudos apresentam vários critérios relevantes para caracterização da voz, boa parte dos quais não são de origem acústica, por exemplo: a etnia, o índice de massa corporal, a profissão, os hábitos associados ao tabagismo e a idade dos pacientes. Esses parâmetros também poderiam ser combinados para melhorar ainda mais os resultados obtidos. Como mencionado na revisão bibliográfica, SVMs, RNAs, somadas a outros algoritmos, derivam diversas outras técnicas, inclusive a DPM. Porém, a técnica proposta apresenta característica para lidar de forma mais adequada com sinais biomédicos, devido à presença de contradições e inconsistências nos sinais naturais.

Alguns projetos de iniciação científica estão em desenvolvimento para a implementação de uma plataforma que possibilite a aquisição da voz e parâmetros não acústicos de forma remota (via Internet), para expandir a proposta deste projeto, formando um banco de vozes que associa características acústicas e não acústicas, com preliminares avaliações da saúde da voz do narrador, baseada em uma DPM já treinada. A principal contribuição desses trabalhos é a criação da plataforma que irá possibilitar a análise da voz de forma remota, contribuindo para um pré-diagnóstico e auxiliando como um Sistema de Apoio à Decisão Médica em locais de difícil acesso pelo especialista.

Assim, este trabalho alcançou a sua proposta de atingir um nível considerável de reconhecimento, contribuindo como uma ferramenta de Apoio à Decisão Médica, sendo um mecanismo menos invasivo quando comparado aos tradicionais aparelhos utilizados. Além disso, a implementação do algoritmo proposto em uma plataforma de computação distribuída, garantido pela velocidade de aplicação do algoritmo, irá proporcionar o pré-diagnóstico das possíveis patologias em um estágio antecipado, o que possibilita um tratamento com maiores chances de sucesso. Implementações para funcionamento em tempo-real com base em um *Digital Signal Processor* (DSP) ou *Field Programmable Gate Array* (FPGA) também estão sendo planejadas.

Por fim, é fundamental observar que nenhuma comparação entre os resultados obtidos

com a DPM e os obtidos com outros classificadores, tais como SVMs, foi inserida neste documento. Isto se deve ao fato de que a DPM foi criada para lidar com características, tais como inconsistências e indefinições, que outros classificadores não podem manipular com perfeição, pois não foram criados para isso. Como exemplo, basta notar que nenhum outro classificador existente na literatura tem condições de localizar uma determinada entrada (voz, no caso deste trabalho) no plano paraconsistente, como mostrado anteriormente. Assim, não seriam coerentes tais comparações.

Referências

- 1 BAKEN, R. J. *Clinical measurement of speech and voice*. San Diego: Singular, 2000.
- 2 NGOC, T. P.; BADIN, P. Vocal tract acoustic transfer function measurements: further developments and applications. *Journal de Physique IV*, v.4, n.C5, p.549-552, 1994.
- 3 DENG, L.; O'SHAUGHNESSY, O. *Speech processing: a dynamic and optimization-oriented approach*. New York: Marcel Dekker, 2003.
- 4 BEMMEL, J. H.; MUSEN, M. A. *Handbook of medical informatics*. 2a ed, Houten, Netherlands, Bohn Stafleu Van Loghum: Singular, 2000.
- 5 ISSHIKI, N.; OKAMURA, H.; TANAVE, M.; MORIMOTO, M. Differential diagnosis of hoarseness. *Folia Phoniatica*, v.21, n.1, p.9-19, 1969.
- 6 BLESS, D. M. Measurement of vocal function. *Otolaryngologic clinics of north america*, v.24, n.5, p.1023-1033, 1991.
- 7 ROSA, M.; PEREIRA, J. C.; GRELLET M. Adaptive estimation of residue signal for voice pathology diagnosis. *IEEE Transactions on Biomedical Engineering*, v.47, n.1, p. 96-104, 2000.
- 8 GRANQVIST, S.; HAMMARBERG, B. The correlogram: a visual display of periodicity. *The Journal of the Acoustical Society of America*, v.114, v.5, p.2934-2945, 2003.
- 9 COSTA, N. C. A.; KRAUSE, D.; BUENO, O. Paraconsistent logics and paraconsistency. In: GABBAY, D. M.; THAGARD, P.; WOOS, J. (Ed.). *Handbook of the philosophy of science*, Amsterdam:NorthHolland, 2005, v.5.
- 10 SILVA FILHO, J. I.; ABE, J. M. *Fundamentos de redes neurais paraconsistentes*. São Paulo: Arte e Ciência, 2000.

- 11 LEME NETO, C. M.; VENSON, N. *Lógica paraconsistente*. Universidade de Santa Catarina, 2002.
- 12 DUDA, R.; HART, P.E.; STORK, D.G. *Pattern recognition*. 2a ed., Wiley John & Sons, 1999.
- 13 LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma introdução às Support Vector Machines. *RITA*, São Paulo, p.1-5, v.14, n.2, 2007.
- 14 GUIDO, R. C.; PEREIRA, J. C.; SLAETS, J. Introduction to the special issue on emergent applications of fractals and wavelets in biology and biomedicine. *Applied Mathematics and Computation*, v.207, p.3-4, 2009.
- 15 GUIDO, R. C.; DAJER, M.; PEREIRA, J. C. Relative entropy measures applied to healthy and pathological voice characterization. *Applied Mathematics and Computation*, v.207, p.95-108, 2009.
- 16 ADDISON, P.; WALKER, J.; GUIDO, R. C. Time-frequency analysis of biosignals. *IEEE Engineering in Medicine and Biology Magazine*, v. 28, p. 14-29, 2009.
- 17 BARBON JR, S. *Dynamic time warping baseado na transformada wavelet*. 2007. 113p. Dissertação (Mestrado em Ciência). Instituto de Física de São Carlos da Universidade de São Paulo, São Carlos, 2007.
- 18 HORII, Y. Vocal Shimmer in sustained phonation. *Journal of Speech Hear*, v.23, p.202-211, 1980.
- 19 SCHOENTGE, J.; GUCHTENEERE, R. Time series analysis of Jitter. *Journal of Phonetics*, v.23, p.189-201, 1995.
- 20 DELLER, J. R. JR.; PROAKIS, J. G.; HANSEM, J. H. L. *Discrete-Time processing of speech signals*. New York: MacMillan Publishing Co., 1993.
- 21 BHATTACHARYYA, A. K. *Voice disorders: an overview*. Disponível em: <http://www.totalhealth.co.uk/clinical-experts/mr-abir-bhattacharyya/voice-disorders-overview>, Acesso em: novembro/2010
- 22 HOFFMAN, H. Vocal nodules(vocal fold nodules, vocal cord nodules, singer's nodules). *Iowa Head and Neck Protocols*, Yowa, 2010. Disponível em: <http://wiki.uiowa.edu/display/protocols/Vocal+Nodules,+fold+nodules,+vocal+cord+nodules,+singer+nodes>. Acesso em: 15 Out. 2010.
- 23 BEHLAU, M. *Avaliação e tratamento das disfonias*. São Paulo: Lovise, 1995.

- 24 FONSECA, E.S. *Wavelets, predição linear e LS-SVM aplicados na análise e classificação de sinais de vozes patológicas*. 2008. 107p. Tese (Doutorado). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- 25 BRAGA, J. C.; OLIVEIRA, D. S. F.; ATHERINO, C. C. T.; SCHOTT, T. C. A.; SILVA, J. C. Vocal nodules: functional and anatomical analysis. *Revista CEFAC*, São Paulo, v.8, n.2, p.223-9, abr-jun, 2006.
- 26 McHUGH-MUNIER, C.M.; SHERER, K. R.; LEHMANN, W; SHERER, U. Coping strategies, personality and voice quality in patients with vocal folds nodules and polyps. *Journal of Voice*, v.11, n.4, p.452-461, 1997.
- 27 BEHLAU, M. S. *Voz o livro do especialista*. Rio de Janeiro: Revinter, 2001.
- 28 YATES, A.; DEDO, H. Carbon dioxide enucleation of polypoid vocal cords. *Laryngoscope*, v.94, p.731-735, 1984.
- 29 DELLACQUA, A. *Análise acústica da voz de obesos mórbidos pré e pós tratamento cirurgico para a obesidade*. Dissertação (Mestrado), Escola de Engenharia Elétrica de São Carlos, São Carlos, 2007.
- 30 ANDRADE, L. M. O. *Determinação dos limiares de normalidade dos parâmetros acústicos da voz*. Dissertação (Mestrado), Escola de Engenharia Elétrica de São Carlos, São Carlos, 2003.
- 31 NIETO, A., Del PALACIO, A. J., LORENZO, A., VEGAS, A., COBETA, I. Los ordenadores em el análisis de la voz: aplicaciones clínicas. *Acta Otorrinolaringologica Española*, v. 46, n. 4, 1995.
- 32 FALCONE, K. *Fractal geometry mathematical foundations and applications*. 2nd. ed. London, UK:Wiley, 2003.
- 33 MIRANDA, V. G. J, ASSIS, T. A., MOTA, F. B., ANDRADE, R. F. S., CASTILHO C. M. C. Geometria fraca: propriedades e características de fractais ideais, *Revista Brasileira de Ensino de Física*, v. 30, n. 2, 2008. doi.org/10.1590/S1806-11172008000200005
- 34 FANTINATO, P. C. *Segmentação de voz baseada na análise fractal e na transformada wavelet*. 2009. 123p. Dissertação (Mestrado). Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2009.
- 35 BACKES, A. R.; BRUNO O.M. Técnicas de estimativa da dimensão fractal: um estudo comparativo. *InfoComp - Journal Of Computer Science*, v. 4, p. 42-49, 2005.

- 36 SECCO, F.R., ROCHA, T. T. *Fractais*. Florianópolis:Universidade de Santa Catarina, 2004. Trabalho apresentado no curso de Ciência da Computação.
- 37 AL-AKAIDI, M. *Fractal speech processing*. Cambridge-UK: Cambridge University Press, 2004.
- 38 SANCHEZ, F. L. ; BARBON, S. J. ; GUIDO, R. C. ; FONSECA, E. ; SCALASSARA, P. R. ; MACIEL, C. D. ; PEREIRA, J. C. ; CHEN, S. H. Wavelet-based cepstrum calculation. *Journal of Computational and Applied Mathematics*, v. 227, p. 288-293, 2009.
- 39 VIEIRA, L. S. *Conversão de voz baseada na transformada wavelet*. 2008. 91p. Dissertação (Mestrado). Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008.
- 40 MILNER B.; SHAO X. Clean speech reconstruction from MFCC vectors and fundamental frequency using an integrated front-end. *Speech Communication*, v. 48, n. 6, p. 697-715, 2006.
- 41 HAYKIN, S. *Neural networks and learning machines*, 3rd. New Jersey:Prentice-Hall Internacional, 2008.
- 42 ABE, J. M. Paraconsistent artificial neural networks: an introduction. IN: *Lecture Notes in Artificial Intelligence*, 2004, v.3214, p.942-948.
- 43 LOPES, H. F. S.; ABE, J. M. Application of paraconsistent artificial neural networks as a method of aid in the diagnosis of Alzheimer disease. *Journal of Medical Systems*, v.34, n.6, p.1073-1081, 2009.
- 44 ABE, J. M.; PRADO, J. C. A.; NAKAMATSU, K. Paraconsistent artificial neural network: applicability in computer analysis of speech productions. IN: *Lecture Notes in Computer Science*, 2006, v.4252, p844-850.
- 45 BARBON, S.J., GUIDO, R.C.; VIEIRA, L.S. A neural-network approach for speech features classification based on paraconsistent logic. IN: IEEE INTERNATIONAL SYMPOSIUM ON MULTIMEDIA, 11, 2009, San Diego. *Proceedings...*San Diego:IEEE, 2009, p. 567-570.
- 46 MERRIMA, M. *A text-book on the method of least squares*. Michigan:University of Michigan Library, 2009.
- 47 BOSI, M; GOLDBERG, R. E. *Introduction to digital audio coding and standards*. Boston-USA: Kruwer Academic Press, 2003.
- 48 COLEMAN, J. *Introducing speech and language processing*. Cambridge: Cambridge University Press, 2005.

49 TRAUNMULLER, H.; ERIKSSON, A. *The frequency range of the voice fundamental in the speech of male and female adults*. Stockholm: University of Stockholm, 1994. Disponível em: < [citeseerx.ist.psu.edu/viewdoc/download?doi = 10.1.1.64.5133rep](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.5133rep) >. Acesso em: novembro/2010.

Apêndice I - Publicações durante o Doutorado

1. GUIDO, R.C.; BARBON, S. J. Introducing the Discriminative Paraconsistent Machine (DPM). *Cognition*, Elsevier, 2011. In press.
2. SANCHEZ, F. L.; BARBON, S. J.; GUIDO, R. C.; FONSECA, E.; SCALASSARA, P. R.; MACIEL, C. D.; PEREIRA, J. C.; CHEN, S. H. Wavelet-based cepstrum calculation. *Journal of Computational and Applied Mathematics*, v.227, p. 288-293, 2009.
3. BARBON, S. J.; GUIDO, R. C.; VIEIRA, L. S.; FONSECA, E. Wavelet-based dynamic time warping. *Journal of Computational and Applied Mathematics*, v. 227, p. 271-287, 2008.
4. GUIDO, R. C.; VIEIRA, L. S.; BARBON, S. J.; SANCHEZ, F. L.; MACIEL, C. D.; FONSECA, E.; PEREIRA, J. C. A neural-wavelet architecture for voice conversion. *Neurocomputing*, Amsterdam, v. 71, p. 174-180, 2007.
5. SCALASSARA, P. R.; MACIEL, C. D.; GUIDO, R. C.; PEREIRA, J. C.; FONSECA, E.; MONTAGNOLI, A.; BARBON, S. J.; VIEIRA, L. S.; SANCHEZ, F. L. Autoregressive decomposition and pole tracking applied to vocal folds nodule identification. *Pattern Recognition Letters*, v. 28, p. 1360-1367, 2007.
6. VIEIRA, L. S.; GUIDO, R. C.; BARBON, S. J.; SANCHEZ, F. L.; DUCATTI, J. A.; MOTA, T. B.; SANTOS, B. L. S.; HOU, T. H.; GOMES, M. H. Identificação de patologias por análise de sinais da fala humana usando a transformada discreta de Fourier. *Boletim Técnico da Faculdade de Tecnologia de São Paulo*, v. 22, p. 53-56, 2007.
7. GUIDO, R. C.; BARBON, S. J.; VIEIRA, L. S.; SANCHEZ, F. L.; MACIEL, C. D.; SCALASSARA, P. R.; PEREIRA, J. C.; PUIA, V. M. Spoken document summarization

- based on dynamic time warping and wavelets. *International Journal of Semantic Computing*, v. 1, p. 347-357, 2007.
8. BARBON, S. J.; VAZ, M. J. Frameworks RIA para JSF lado a lado entrada de dados. *Revista MundoJ*, Curitiba, p. 36 - 42, 01 ago. 2010.
 9. BARBON, S. J.; VAZ, M. J. Frameworks RIA para JSF lado a lado exibição de dados). *Revista MundoJ*, Curitiba, p. 34 - 41, 01 nov. 2010.
 10. BARBON, S. J.; GUIDO, R. C.; VIEIRA, L. S. A Neural-Network approach for speech features classification based on paraconsistent logic. In: IEEE INTERNATIONAL SYMPOSIUM ON MULTIMEDIA (ISM2009), 2009, San Diego. *Proceedings....* San Diego: IEEE, 2009. v. 1. p. 567-570.
 11. GUIDO, R. C.; BARBON, S. J.; VIEIRA, L. S.; SANCHEZ, F. L.; MACIEL, C. D.; PEREIRA, J. C.; SCALASSARA, P. R.; FONSECA, E. Introduction to the discrete shapelet transform and a new paradigm: joint time-frequency-shape analysis. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS (IEEE ISCAS 2008), 2008, Seattle. *Proceddings....*, Seattle:IEEE, 2008. v. 1. p. 2893-2896.
 12. VIEIRA, L. S.; SANCHEZ, F. L.; GUIDO, R. C.; BARBON, S. J. On the use of Wavelets for voice conversion. In: CONGRESSO DE ENGENHARIA DE ÁUDIO, 5, 2007, São Paulo. *Anais...* São Paulo, 2007. CD-ROM.
 13. SANCHEZ, F. L.; VIEIRA, L. S.; BARBON, S. J.; GUIDO, R. C. Wavelet-based audio watermarking. In: CONGRESSO DE ENGENHARIA DE ÁUDIO, 5, 2007, São Paulo. *Anais...* São Paulo, 2007. CD-ROM.
 14. BARBON, S. J.; VIEIRA, L. S.; GUIDO, R. C.; SANCHEZ, F. L.; Wavelets aplicadas a dynamic time warping. In: CONGRESSO DE ENGENHARIA DE ÁUDIO, 5, 2007, São Paulo. *Anais...* São Paulo, 2007. CD-ROM.
 15. VIEIRA, L. S.; GUIDO, R. C.; BARBON, S. J.; SANCHEZ, F. L.; SERGIO, K. I. C.; GUILHERM, M. B. A.; SOUZA, L. M.; SOLGON, R.; FONSECA, E. On the use of

- wavelets for voice morphing. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, CNMAC, 30, 2007, Florianópolis. *Anais...*, Florianópolis: 2007, CD ROM.
16. BARBON, S. J.; GUIDO, R. C.; CHEN, S. H.; VIEIRA, L. S.; SANCHEZ, F. L. Improved dynamic time warping based on the Discrete Wavelet Transform. In: IEEE INTERNATIONAL SYMPOSIUM ON MULTIMEDIA (IEEE ISM2007), 9, 2007, Taichung. *Proceedings...*, Taichung:IEEE, 2007. v. 1. p. 256-261.
 17. GUIDO, R. C.; BARBON, S. J.; VIEIRA, L. S.; RODRIGUES, L. C.; FANTINATO, P. C.; ESCOLA, J. P. Autenticação biométrica por comando de voz com implementação em hardware dedicado. In: WORKSHOP NO ÂMBITO DO PROJECTO, 2, 2008, São Carlos *Anais...* São Carlos: 2008. CD-ROM.
 18. GUIDO, R. C.; BARBON, S. J.; VIEIRA, L. S.; FANTINATO, P. C.; SANCHEZ, F. L.; ESCOLA, J. P. Speech compression based on the discrete shapelet transform. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, CNMAC, 31, 2008, Belém. *Anais...*, Belém: 2008, CD-ROM.
 19. BARBON, S. J.; VIEIRA, L. S.; RODRIGUES, L. C.; GUIDO, R. C. Uso de janelas triangulares no auxílio à percepção sonora por portadores de necessidades especiais auditivas. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, CNMAC, 33, 2010, Aguas de Lindoia. *Anais...*, Aguas de Lindoia: 2010, CD ROM.
 20. QUEIROZ, I. D. D.; NEVES, L. A.; BARBON, S. J. Simulação computacional tridimensional da bronquite crônica. In: SIMPÓSIO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA, 12, São Paulo. *Anais....* São Paulo:2010, v. 1. p. 51-51.
 21. SOUZA, D. R. G.; BARBON, S. J. Avaliação da necessidade de implantação de TI nas propriedades com citricultura. In: SIMPÓSIO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA, 12, São Paulo. *Anais....* São Paulo:2010, v. 1. p.9-9

22. DUARTE, A. O.; BARBON, S. J. Avaliação da redução da resolução em frequência de sinais de vozes considerando a conservação do pitch. In: SIMPÓSIO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA, 12, São Paulo. *Anais...* São Paulo:2010, v. 1. p.10-10.
23. SANCHEZ, F. L.; GUIDO, R. C. ; VIEIRA, L. S. ; BARBON, S. J. Inserção e extração de marca d'água em arquivos digitais. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, CNMAC, 30, 2007, Florianópolis. *Anais...*, Florianópolis: 2007, CD ROM.
24. PUIA, V. M.; VIEIRA, L. S. ; GUIDO, R. C. ; BARBON, S. J. Reconhecimento de locutores usando wavelets. In: JORNADA CIENTÍFICA DA UFSCAR, 7, 2007, São Carlos,*Anais...*, São Carlos:2007. CD-ROM.
25. DUARTE, A. O.; BARBON, S. J. Processamento das características da voz no domínio da frequência baseado em método com resolução reduzida. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E EXTENSÃO, 12, Frutal.*Anais...*Frutal:UEMG, 2010. CD-ROM.
26. RODRIGUES, T.; BARBON, S. J. Avaliação de características da voz baseado em redes neurais artificiais. In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E EXTENSÃO, 12, Frutal.*Anais...*Frutal:UEMG, 2010. CD-ROM.