

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

CLASSIFICAÇÃO DE INTOXICAÇÃO ALCOÓLICA POR
SINAIS DE VOZ E ALGORITMOS GENÉTICOS

GLEISON MARQUES LEMOS LEONI

ORIENTADOR: DÍBIO LEANDRO BORGES

DISSERTAÇÃO DE MESTRADO EM
ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPGEE.DM - 637/2016

BRASÍLIA/DF: DEZEMBRO/2016.

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**CLASSIFICAÇÃO DE INTOXICAÇÃO ALCOÓLICA POR SINAIS DE
VOZ E ALGORITMOS GENÉTICOS**

GLEISON MARQUES LEMOS LEONI

DISSERTAÇÃO DE MESTRADO PROFISSIONAL SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

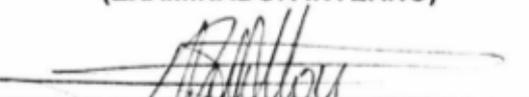
APROVADA POR:



DIBIO LEANDRO BORGES, Dr., CIC/UNB
(ORIENTADOR)



FLÁVIO ELIAS GOMES DE DEUS, Dr., ENE/UNB
(EXAMINADOR INTERNO)



NILTON CORREIA SILVA, Dr., FGA/UNB
(EXAMINADOR EXTERNO)

Brasília, 13 de Dezembro de 2016.

FICHA CATALOGRÁFICA

LEONI, GLEISON MARQUES LEMOS

Classificação de Intoxicação Alcoólica por Sinais de Voz e Algoritmos Genéticos.

[Distrito Federal] 2016.

xv, 65 p., 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2016).

Dissertação de Mestrado - Universidade de Brasília.

Faculdade de Tecnologia. Departamento de Engenharia Elétrica.

- | | |
|----------------------------|-------------------------|
| 1. Intoxicação Alcoólica | 2. Algoritmos Genéticos |
| 3. Aprendizagem de Máquina | 4. Análise Cepstral |
| I. ENE/FT/UnB | II. Título (série) |

REFERÊNCIA BIBLIOGRÁFICA

Leoni, G. M. L. (2016). Classificação de Intoxicação Alcoólica por Sinais de Voz e Algoritmos Genéticos. Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGENE.DM - 637/2016, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 65p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Gleison Marques Lemos Leoni.

TÍTULO DA DISSERTAÇÃO DE MESTRADO: Classificação de Intoxicação Alcoólica por Sinais de Voz e Algoritmos Genéticos.

GRAU / ANO: Mestre / 2016

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Do mesmo modo, a Universidade de Brasília tem permissão para divulgar este documento em biblioteca virtual, em formato que permita o acesso via redes de comunicação e a reprodução de cópias, desde que protegida a integridade do conteúdo dessas cópias e proibido o acesso a partes isoladas desse conteúdo. O autor reserva outros direitos de publicação e nenhuma parte deste documento pode ser reproduzida sem a autorização por escrito do autor.



Gleison Marques Lemos Leoni

Av. Walter Ananias nº 705, SETEC/SR/PF/AL, Jaraguá,
57025-080 Maceió - AL - Brasil.

DEDICATÓRIA

À minha família.

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus, por minha família, e a todas as pessoas que estão presentes em minha vida e que, de alguma forma, contribuíram com a realização deste trabalho. Agradeço ao professor Díbio Leandro Borges por aceitar o convite para participar dessa jornada. Agradeço a todos os colegas do Setor Técnico-Científico da Superintendência de Polícia Federal em Alagoas pelo incentivo e pela troca de experiências. Agradeço a todas as pessoas que contribuem com os impostos que possibilitam o ensino público de qualidade.

Este trabalho foi realizado com o apoio da Polícia Federal - PF, com recursos do Programa Nacional de Segurança Pública com Cidadania - PRONASCI, do Ministério da Justiça.

Gleison Marques Lemos Leoni

RESUMO

CLASSIFICAÇÃO DE INTOXICAÇÃO ALCOÓLICA POR SINAIS DE VOZ E ALGORITMOS GENÉTICOS

Autor: Gleison Marques Lemos Leoni

Orientador: Díbio Leandro Borges

Programa de Pós-graduação em Engenharia Elétrica

Brasília, Dezembro de 2016

Dados oficiais apontam que a intoxicação alcoólica se encontra entre as maiores causas de mortes no trânsito. A sociedade brasileira clama por uma solução, pois assiste, dia após dia, o aumento do número de acidentes envolvendo motoristas alcoolizados e/ou entorpecidos por outras drogas. O governo responde à sociedade endurecendo normativos legais e realizando campanhas de conscientização.

Atualmente, o nível de alcoolemia de um motorista é obtido por meio de exame sanguíneo ou por meio de etilômetros. A proposta deste trabalho é desenvolver um modelo, baseado em análise espectral, para possibilitar a indicação da alcoolização de uma pessoa embriagada por meio de sua fala.

Sabe-se que o álcool é uma droga depressora, que atua no sistema nervoso central provocando mudanças comportamentais do indivíduo alcoolizado e efeitos físicos pela ação direta em parte do trato vocal. O resultado das sensações física e psicomotora no indivíduo embriagado traz ao ouvinte de seu discurso a impressão de escutar uma voz mais ruidosa. Esta sensação é percebida por técnicas de análise espectral.

Neste trabalho foi realizada a análise espectral de tempo curto das falas de pessoas embriagadas e sóbrias por meio de GFCC (*Gamma-tone Frequency Cepstral Coefficients*). As características geradas foram selecionadas por um Algoritmo Genético (AG) com uma função de avaliação implementada com Regressão Logística. O objetivo do AG era alcançar a maior taxa de classificação com a menor quantidade de características espectrais possível.

A técnica obteve taxas de classificação equivalente a 81,18% quando avaliados por meio de validação cruzada LOO (*Leave One Out*) e conseguiu reduzir as características de entrada em mais de 90%. Foi também realizada a extração de características por meio de RFE (*Recursive Feature Elimination*). Por tal metodologia, alcançou-se uma taxa de classificação equivalente a 85,49%, tendo selecionado, contudo, 98 características.

Na comunidade científica não foram encontrados estudos de classificação de vozes intoxicadas pelo álcool utilizando algoritmos genéticos nem GFCC. Além disso, as taxas de classificação alcançadas neste trabalho superam pesquisas anteriores que utilizaram somente características cepstrais da voz para realizar a classificação de falas intoxicadas pelo álcool.

ABSTRACT

CLASSIFICATION OF ALCOHOL INTOXICATION BY SPEECH SIGNALS AND GENETIC ALGORITHMS

Author: Gleison Marques Lemos Leoni

Supervisor: Díbio Leandro Borges

Programa de Pós-graduação em Engenharia Elétrica

Brasília, December of 2016

Official data indicate that alcohol intoxication is currently among the leading causes of deaths in traffic. Brazilian society is clamoring for a solution, as it witnesses, day after day, the increase in the number of accidents involving drivers who are drunk and / or numb by other drugs. The government responds to society by hardening legal norms and conducting awareness campaigns.

At present, the level of a driver's blood alcohol content is obtained by blood test or by means of breathalyzer. The purpose of this work is to develop a model, based on spectral analysis, to enable the indication of the alcoholization of a drunk person through his speech.

It is known that alcohol is a depressant drug, which acts on the central nervous system causing behavioral changes of the alcoholic individual and physical effects by direct action in part of the vocal tract. The result of the physical and psychomotor sensation in the drunk individual brings to the listener of his speech the impression of hearing a noisy voice. This sensation is perceived by spectral analysis techniques.

In this work, the short-time spectral analysis of the speeches of intoxicated and sober people by GFCC (Gammatone Frequency Cepstral Coefficient) was performed. The generated features were selected by a Genetic Algorithm (AG) with an evaluation function implemented with Logistic Regression. The goal of GA was to achieve the highest classification rate with the least amount of spectral features possible.

The technique obtained classification rates equivalent to 81.18% when evaluated through LOO cross-validation (Leave One Out) and managed to reduce input features

by more than 90%. Feature extraction was also performed by RFE (Recursive Feature Elimination). By such methodology, a classification rate equivalent to 85.49% was reached, having, however, selected 98 characteristics.

In the scientific community, no classification studies of alcohol intoxicated speeches were found using genetic algorithms or GFCC. In addition, the classification rates achieved in this study surpass previous studies that only used cepstral features of the speech to perform the classification of speech intoxicated by alcohol.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	Justificativa e Motivação	2
1.2	Metodologia	3
1.3	Objetivo e Hipótese	6
1.4	Organização da dissertação	7
2	Revisão bibliográfica e o estado da arte	8
2.1	O básico da fonética acústica	8
2.2	Análise da intoxicação alcoólica por meio da voz: Aspectos históricos .	14
2.3	O estado da arte	15
2.3.1	Detecção de doenças por meio da voz	15
2.3.2	Detecção de sotaques	17
2.3.3	Detecção de intoxicação por meio do álcool	19
2.4	Considerações Finais	19
3	<i>Corpus</i> de falas alcoolizadas	21
3.1	Introdução	21
3.2	Desenvolvimento do <i>corpus</i>	22
3.3	Procedimento de gravação dos discursos	23
3.4	Considerações Finais	25
4	Algoritmos genéticos	26
4.1	Considerações iniciais	26
4.2	Inicialização	27
4.2.1	Técnicas estocásticas	29

4.2.2	Técnicas determinísticas	29
4.2.3	Técnicas de Não-composição	29
4.2.4	Técnicas de Composição	30
4.2.5	Técnicas genéricas	30
4.2.6	Técnicas para aplicações específicas	30
4.3	Avaliação	30
4.4	Seleção	30
4.4.1	Seleção por roleta	31
4.4.2	Seleção por amostragem estocástica universal	31
4.4.3	Seleção por ranking	32
4.4.4	Seleção por torneio	32
4.4.5	Seleção por NSGA-II	33
4.5	Cruzamento	34
4.5.1	Cruzamento de um ponto	34
4.5.2	Cruzamento de múltiplos pontos	35
4.5.3	Cruzamento uniforme	35
4.6	Mutação	35
4.6.1	Mutação de um ponto	36
4.6.2	Mutação de múltiplos pontos	36
4.6.3	Mutação uniforme	36
4.7	Parametrização	37
4.7.1	Tamanho da população	37
4.7.2	Taxas de cruzamento e mutação	37
4.7.3	Critério de parada	38
4.7.3.1	Critérios baseados na exaustão	38
4.7.3.2	Critérios baseados no progresso	38
4.7.3.3	Critérios baseados no movimento	38

4.7.3.4	Critérios baseados na distribuição	38
4.7.3.5	Critérios combinados	39
4.8	Considerações finais	39
5	Técnica acústica proposta para detecção de intoxicação alcoólica	40
5.1	Considerações iniciais	40
5.2	Extração do vetor de características	41
5.2.1	Janelamento	41
5.2.2	Características cepstrais	44
5.2.3	Características estáticas, dinâmicas e estatísticas	46
5.2.4	Definição dos vetores de características	47
5.3	Seleção das características por meio de Algoritmo Genético	48
5.3.1	Parametrização do algoritmo genético	49
5.3.1.1	Definição da população inicial	49
5.3.1.2	Definição do critério de seleção	49
5.3.1.3	Definição da taxa de cruzamento	50
5.3.1.4	Definição da taxa de mutação	50
5.3.1.5	Critério de parada	50
5.4	O processo de classificação	50
5.5	Comparação com trabalhos similares	50
5.6	Avaliação dos efeitos do gênero na classificação	52
5.7	Validação dos dados	53
5.8	Considerações Finais	54
6	CONCLUSÕES	55
6.1	Recomendações para pesquisas futuras	56

LISTA DE TABELAS

2.1	Variações encontradas pelos peritos nas falas do comandante do petroleiro Exxon Valdez, conforme apresentado em Johnson (JOHNSON; PISONI; BERNACKI, 1990). Adaptada de (HEIGL, Fall 2011).	15
2.2	Tipos de grupos e doenças em que foi possível realizar a classificação por redes neurais. Adaptada de (CROVATO, 2004) p.91.	16
3.1	Tipo de falas existentes no <i>corpus</i> ALC. Adaptada de (SCHIEL et al., 2010)	23
4.1	Valores de aptidão ajustados por meio da pressão de seleção.	33
4.2	Material genético dos pais selecionados para os exemplos de tipos de cruzamentos das Tabelas a seguir.	34
4.3	Exemplo de cruzamento de um ponto considerando.	35
4.4	Exemplo de cruzamento de múltiplos pontos.	35
4.5	Máscara gerada aleatoriamente	35
4.6	Exemplo de cruzamento uniforme.	36
4.7	Exemplo de mutação de um ponto.	36
4.8	Exemplo de mutação de múltiplos pontos.	36
4.9	Exemplo de mutação uniforme.	36
5.1	Comparação entre MFCC e GFCC.	46
5.2	Classificadores propostos em outros trabalhos similares	52
5.3	Avaliação do efeito do gênero na classificação.	53

LISTA DE FIGURAS

1.1	Esquemático diferenciando as características segmentais e suprasegmentais.	5
2.1	Sistema vocal humano. Adaptada a partir de (TRASCASA, 2013), p. 3.	9
2.2	Vista superior da laringe mostrando as posições da glote. (A) Glote na posição de repouso. (B) Glote em atividade. Adaptada de (CROVATO, 2004), p. 41.	9
2.3	Seção do eixo sagital do trato vocal, onde se demonstram os articuladores ativos (maiúsculos) e passivos (minúsculos). Adaptada de (ROSE, 2002), p. 153.	10
2.4	Representação esquemática do mecanismo de produção da voz. Em detalhe, no interior do retângulo tracejado, os órgãos que compõem o trato vocal. Adaptada de (NEY, 2003), p. 8.	11
2.5	Espaço acústico das vogais orais obtido para o português brasileiro. Adaptada de (SANTOS, 2013), p. 77.	12
2.6	Espaço acústico das vogais a, i e u para falas de pessoas nos estados sóbrio e alcoolizado, obtido para falantes do sexo feminino no idioma alemão. A linha contínua indica o centróide das vogais para as falas alcoolizadas, enquanto a linha tracejada o das falas sóbrias. Adaptada de (SCHIEL; HEINRICH; NEUMEYER, 2010).	13
2.7	Diferença entre as amostras de dois sinais de voz.(a) Voz de uma pessoa saudável. (b) Voz de uma pessoa diagnosticada com mal de Parkinson. Adaptada de (SHIRVAN; TAHAMI, 2011).	16
2.8	Diagrama de blocos proposto para detecção de sotaques. Adaptada de (HUANG; WANG, 2016). p.4.	18
3.1	Histograma apontando os níveis de alcoolização a que foram submetidos os indivíduos durante a formação do <i>corpus</i> de voz.	25
4.1	Diagrama de blocos genérico de um algoritmo genético. Adaptada de (HINÇAL, 2008), p.20.	27

4.2	Exemplos de cromossomos e seus genes. Os genes do "Cromossomo A" estão codificados em binário, enquanto os do "Cromossomo B" estão codificados em real.	28
4.3	Classificação dos tipos de iniciação de população. Adaptada de (KAZIMIPOUR; LI; QIN, 2014).	28
4.4	Exemplo típico da seleção por roleta. Adaptada a partir de (NORAINI; GERAGHTY, 2011).	31
4.5	Exemplo típico da seleção por amostragem espectral universal. Adaptada a partir de (NORAINI; GERAGHTY, 2011).	32
4.6	Exemplo de modificação da roleta, baseado nos dados de ordenação e de aptidão ajustada calculada por meio da Equação 4.1 e demonstrada na Tabela 4.1.	33
4.7	Diagrama esquemático do funcionamento do algoritmo NSGA-II. Adaptado de (ZINI, 2009), p.77.	34
5.1	Diagrama de blocos básico do modelo proposto de sistemas de análise de voz.	41
5.2	Imagem demonstrativa dos tipos de janela existentes na literatura científica. Adaptada a partir de (DENG; O'SHAUGHNESSY, 2003).	41
5.3	Comparação das respostas em frequência dos janelamentos de Hamming, Hanning e Blackman. Adaptada a partir de (PODDER et al., 2014).	42
5.4	Espectrograma da fala da língua inglesa " <i>Hear aga now</i> " proferida por um indivíduo do sexo masculino. (A) Fala original ou sem perdas. (B) Fala processada com janela de 2 ms. (C) Fala processada com janela de 8 ms. (D) Fala processada com janela de 32 ms. (E) Fala processada com janela de 128 ms. (F) Fala processada com janela de 512 ms. Adaptada a partir de (PALIWAL; WÓJCICKI, 2008).	43
5.5	Exemplo de resposta em frequência de canais do filtro Gamma. Adaptada a partir de (ZOGHLAMI; LACHIRI, 2012).	44
5.6	Relação entre a largura de banda de um filtro genérico e a ERB. Adaptada a partir de (ABDULLA, 2012).	45
5.7	Ilustração da evolução do processo de redução de características do classificador implementado.	51

5.8	Ilustração da evolução da performance do classificador implementado. .	51
5.9	Ilustração dos resultados de busca das melhores características por RFE.	53

LISTA DE SÍMBOLOS, NOMENCLATURA E ABREVIACÕES

AG: Algoritmo Genético.

ALC: *Alcohol Language Corpus*.

BAS: *Bavarian Archive for Speech Signals*.

CBT: Código Brasileiro de Trânsito.

CLT: Consolidação das Lei do Trabalho.

CNH: Carteira Nacional de Habilitação.

ERB: *Equivalent Rectangular Bandwidth*.

GFCC: *Gammatone Frequency Cepstral Coefficients*.

GMM: *Gaussian Mixture Model*.

HMM: *Hidden Markov Model*

HNR: *Harmonics to Noise Ratio*.

KNN: *K Nearest Neighbor*.

LDA: *Linear Discriminant Analysis*.

LOOCV: *Leave One Out Cross Validation*.

LPC: *Linear Predictive Coding*.

MFCC: *Mel-Frequency Cepstral Coefficients*.

NSGA-II: *Non-dominated Sorting Genetic Algorithm*.

PCA: *Principal Component Analysis*.

PLC: Projeto de Lei da Câmara.

PNN: *Probabilistic Neural Network*.

RFE: *Recursive Feature Elimination*.

SAHS: *Self-adaptative Harmony Search*.

STF: Supremo Tribunal Federal.

SUS: Sistema Único de Saúde.

SVM: *Support Vector Machine*.

USA: *United States of America*.

GLOSSÁRIO

CEPSTRUM: O cepstrum é a transformada de Fourier do logaritmo da transformada de Fourier de um sinal. As vezes é chamado de espectro de um espectro.

DISFONIA: A disfonia é uma expressão médica que representa a dificuldade de fonação de qualquer origem.

FORMANTE: Um formante é um pico de energia em uma região do espectro sonoro.

JITTER: É a média absoluta das variações da frequência fundamental entre períodos consecutivos. (FARRÚS; HERNANDO; EJARQUE, 2007).

PITCH: É a média absoluta das diferenças entre períodos consecutivos da frequência fundamental, dividido pela média do período médio. (FARRÚS; HERNANDO; EJARQUE, 2007).

SHIRMER: É a média absoluta das diferenças entre períodos consecutivos da amplitude, dividido pela média do período médio. (FARRÚS; HERNANDO; EJARQUE, 2007).

TIMBRE: É uma característica acústica da fala, cujo efeito se correlaciona às dimensões das cavidades que vibram em ressonância com as pregas vocais.

WAVELET: É uma função capaz de decompor e representar outra função originalmente descrita no domínio do tempo, de forma a podermos analisá-la em diferentes escalas de frequência e de tempo.

1 INTRODUÇÃO

Diversas campanhas de conscientização tentam, há anos, demonstrar à população brasileira que a mistura álcool e direção, definitivamente, não combina. Apesar deste esforço, a sociedade tem observado, de forma atônita, o aumento do número de acidentes envolvendo motoristas alcoolizados e/ou entorpecidos por outras drogas. Afirma-se que, desde a entrada em vigor do Código Brasileiro de Trânsito (CBT) (BRASIL, 1997), a taxa de mortalidade em acidentes de tráfego no Brasil se mantém em torno 20 mortes por 100 mil habitantes e que, entre um quarto e a metade destes acidentes, houve consumo de álcool por algum dos envolvidos no sinistro (ECKSCHMIDT, 2015). Nos Estados Unidos, o abuso do álcool está ligado a 41% das mortes no trânsito, para indivíduos com menos de 33 anos (HINGSON; WINTER, 2003).

Dados do Distrito Federal, apresentados em (MODELLI; PRATESI; TAUIL, 2008), indicam que mais de 32% das vítimas fatais de atropelamento apresentaram taxas de alcoolização elevada. Além disso, mais de 57% das pessoas que vieram a óbito (87% destes estavam na direção do veículo) devido à capotagem do automóvel, tinham se excedido no consumo de álcool. Estatística similar foi obtida quando se estudou os índices de alcoolemia dos mortos devido a colisões dos veículos - mais de 44% tinham superado a taxa de 0,6 g/l de álcool no sangue.

Tais números assombram a sociedade de tal maneira que se encontra em tramitação no Congresso Nacional o Projeto de Lei da Câmara 144/2015 (PLC 144/2015) (BRASIL. Câmara dos Deputados, 2015), que endurecerá as penas privativas de liberdade dos crimes de homicídio culposo e de lesão corporal quando o motorista causador do acidente se apresentar com a capacidade psicomotora alterada em razão do álcool ou de outras drogas que causem dependência. Entre as justificativas do projeto de lei, afirma-se que a "*violência no trânsito é responsável pela terceira maior causa de mortes no Brasil, ficando atrás apenas das mortes decorrentes de doença do coração e do câncer*".

Este capítulo é subdividido da seguinte forma: a seção 1.1 apresenta o histórico das atualizações normativas relativas à problemática envolvida nesta pesquisa. A seção 1.2 aborda a metodologia empregada. A seção 1.3 traça os objetivos deste trabalho. Por último, a seção 1.4 descreve a estrutura desta dissertação.

1.1 Justificativa e Motivação

O texto do atual CTB tipifica como crime, no Art. 306, o ato de *“conduzir veículo automotor, na via pública, sob a influência de álcool ou substância de efeitos análogos, expondo a dano potencial a incolumidade de outrem”*. Contudo, a Lei 11.705/2008 (BRASIL, 2008b) alterou a redação desse artigo, que passou a não mais exigir a exposição potencial a algum dano, mas, em contrapartida, passou a exigir a determinação da concentração alcoólica do motorista, penalizando-o se ele *“conduzir veículo automotor, na via pública, estando com concentração de álcool por litro de sangue igual ou superior a 6 (seis) decigramas, ou sob a influência de qualquer outra substância psicoativa que determine dependência”*. Além disso, em seguida, entrou em vigor o Decreto 6.488/2008 (BRASIL, 2008a) que estabelece que *“Qualquer concentração de álcool por litro de sangue sujeita o condutor às penalidades administrativas do art. 165 da Lei no 9.503, de 23 de setembro de 1997 - Código de Trânsito Brasileiro, por dirigir sob a influência de álcool.”*

Contudo, desde a sanção da Lei 11.705/2008, os mais renomados juristas discutem a respeito da constitucionalidade da utilização do etilômetro (bafômetro) como meio de prova do estado de embriaguez, sob a argumentação de que tal dispositivo violaria o princípio da não auto incriminação - *“nemo tenetur se detegere”* (ARAÚJO, 2015), repetido popularmente como *“ninguém é obrigado a produzir prova contra si mesmo”*. Além disso, o Superior Tribunal de Justiça Federal (STJ), em 28/03/2012, confirmou a necessidade de aferição da concentração alcoólica, por meio de etilômetro ou exames de sangue, para a comprovação da prática do crime de dirigir sob efeito de álcool (BRASIL. Superior Tribunal de Justiça, 2012), embora não tenha se manifestado sobre as penas administrativas, como a suspensão da Carteira Nacional de Habilitação (CNH), impostas quando o motorista se recusa a realizar os exames.

Em 20/12/2012, a Presidente da República sancionou a Lei 12.760/2012 (BRASIL, 2012), que alterou alguns artigos da Lei 9.503/1997, entre os quais estabeleceu que a prova do estado de embriaguez *“... poderá ser obtida mediante teste de alcoolemia, exame clínico, perícia, vídeo, prova testemunhal ou outros meios de prova em direito admitidos, observado o direito à contraprova”*. Contudo, já em 04/05/2016, houve nova intervenção do poder executivo com a sanção da Lei 13.281/2016 (BRASIL, 2016), que ampliou a punição para quem *“Recusar-se a ser submetido a teste, exame clínico, perícia ou outro procedimento que permita certificar influência de álcool ou outra substância psicoativa, (...)”*.

Sabe-se que os exames de alcoolemia são realizados por quaisquer policiais, utilizando-se de etilômetros que o cidadão desconhece quando, ou por quem, foram aferidos. Todo o processo da estimativa do estado de alcoolemia de um motorista é questionável juridicamente, razão pela qual os normativos se alteram tão seguidamente. Além disso, um caso de recusa da realização de exames de alcoolemia receberá do Estado sanção desproporcional à sua atitude por meio de multas de alta monta ou até prisão, o que provavelmente trará uma nova *pièce de résistance* ao vale tudo probatório do Estado. Além disso, a definição de alcoolemia por meio de etilômetros exige a utilização de piteiras descartáveis, o que aumenta os custos do Estado com a aquisição de produtos consumíveis.

Todos os normativos legais citados alhures, somados aos problemas jurídicos que constantemente ocorrem e aos custos de manutenção das grandes campanhas de lei seca pelo país, abrem um grande nicho à Criminalística: A determinação da alcoolização por meio da análise computacional das vozes de indivíduos embriagados.

Em complemento, juridicamente, a necessidade da determinação da condição de intoxicação alcoólica de um indivíduo não está presente somente no CTB. Diversas outras leis apresentam artigos que tratam de pessoas embriagadas, como, por exemplo, a Lei 13.146/2015 (BRASIL, 2015) que estabelece que ébrios habituais ou viciados em tóxicos são absolutamente incapazes de exercer os atos da vida civil. Já conforme a Consolidação das Leis do Trabalho (CLT) (BRASIL, 1943), a embriaguez habitual pode ser punida com demissão por justa causa. Além disso, existem diversos casos concretos, na esfera criminal, em que é relevante a indicação do estado de sobriedade de um suspeito. Outro exemplo é a necessidade da realização da identificação de um pessoa por meio da voz (Exames de Comparação de Locutores) e na única prova existente, tal indivíduo sob suspeição se encontra alcoolizado.

1.2 Metodologia

O experto deve depreender o que esperar dos sinais acústicos das falas das pessoas alcoolizadas. Para tanto deve conhecer os princípios de Processamento Digital de Sinais, de Fonética Forense e de Reconhecimento de Padrões. Com esse objetivo, foi realizada uma extensa revisão destes conceitos.

Verificou-se que as abordagens disponibilizadas pela comunidade científica para estudo da voz são vastas. No entanto, não foi encontrado amplo material de estudo que

visasse a classificação de falas de pessoas alcoolizadas. Os estudos mais relevantes disponíveis foram produzidos por pesquisadores da Universidade de Munich. Assim, além da revisão do estado da arte existente neste campo do processamento digital dos sinais, foram exploradas áreas de estudo correlatas que tinham como objetivo identificar doenças por meio da voz, sotaques e emoções por meio da fala.

Em seguida, para subsidiar os estudos, foi buscado na literatura científica, e em contatos realizados com pesquisadores brasileiros em fonética forense, um *corpus* de vozes de falantes do português brasileiro que apresentasse indicações dos estados de embriaguez bem determinados, ou seja, que informasse o grau de alcoolemia mensurado por exame de sangue ou por etilômetro. A resposta comum apontou para a não existência de tal banco de falas no Brasil. Desta forma, o material de suporte em que se baseia toda esta pesquisa é o *Alcohol Language Corpus* (ALC), que é uma fração do *Bavarian Archive for Speech Signals* (BAS), fornecido gratuitamente para efeitos de pesquisa pela Universidade de Múnich - Alemanha, no idioma alemão.

Considerando a estrutura do *corpus* de falantes de pessoas alcoolizadas, que será tratado em seção específica, esta dissertação propõe a realização do processamento dos discursos por meio de técnicas de fonética acústica, através de análise cepstral. Os vetores de características (*features*) foram extraídos de forma segmental e selecionados por meio por meio de um Algoritmo Genético (AG).

A fonética é a ciência que classifica os sons das línguas naturais, bem como descreve fisiológica e fisicamente os processos envolvidos na produção da fala (KIKUCHI, 2001). Mais especificamente, a fonética acústica "realiza a investigação dos caracteres acústicos envolvidos na produção dos sons da voz humana por meio de programas e equipamentos que permitem mensurar o tempo, a amplitude e a frequência das ondas sonoras complexas, próprias desses sons"(SANTOS, 2013).

O elemento principal da análise cepstral é o *cepstrum*. Seu nome vem da reorganização da escrita de *spectrum*, do inglês, espectro. É comumente utilizado em aplicações de processamento digital de sinais de imagem e de voz. Em análise de sinais acústicos da fala é utilizado para separar e estimar o conteúdo espectral dos discursos a partir das frequências de sua tonalidade (PROAKIS; MANOLAKIS, 1996). Geralmente o *cepstrum* pode ser gerado pela Transformada de Fourier da função logaritmica da Transformada de Fourier do sinal de voz original (MCLOUGHLIN, 2009), por exemplo.

A criação do vetor de características (*features*) pode se dar de forma segmental (tempo curto) ou de forma suprasegmental (tempo longo), de acordo com sua estrutura temporal (ANAGNOSTOPOULOS; ILIOU; GIANNOUKOS, 2012). Na primeira delas os parâmetros são calculados a cada intervalo pequeno de tempo, enquanto na outra estes parâmetros são obtidos considerando-se todo o enunciado. Normalmente a técnica de segmentação utilizada é aquela em que as janelas são fixas e deslocadas em uma ordem invariável no tempo. Comumente, os valores utilizados para cada janela são equivalentes a 25 ms e o deslocamento padrão é de 10 ms (LEE; CHO, 2016).

A Figura 1.1 ilustra as diferenças entre os parâmetros segmentais e suprasegmentais. São exemplos de características em nível suprasegmental o *Pitch*, *Shirmer*, *Jitter*, taxa de fala, quociente normalizado de amplitude, entre outros. São exemplos de características em nível segmental o MFCC (*Mel-Frequency Cepstral Coefficients*) e o GFCC (*Gammatone Frequency Cepstral Coefficients*).

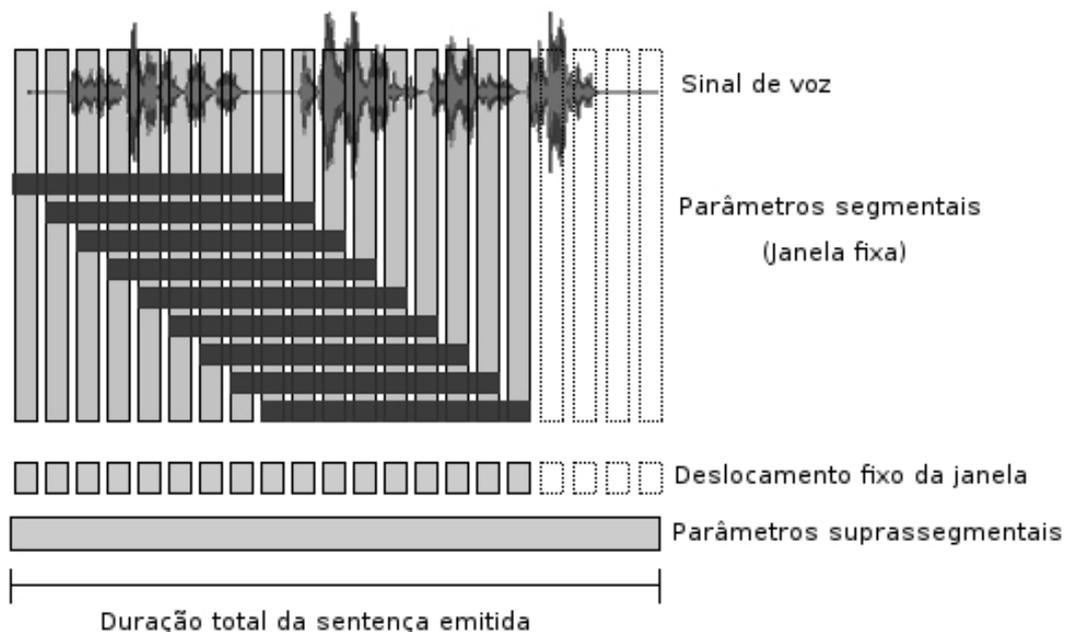


Figura 1.1: Esquemático diferenciando as características segmentais e suprasegmentais.

Geralmente os vetores de características apresentam diversos parâmetros que são redundantes ou irrelevantes. Contudo, a presença destes parâmetros levam a uma baixa precisão dos classificadores (ZHANG et al., 2009). Há, então, a necessidade da realização da seleção destes parâmetros. Os benefícios desta seleção para o processo de classificação são, além da melhoria das taxas de assertividade, a utilização de menor quantidade de recursos computacionais durante a aprendizagem, refletindo num tempo menor de aprendizado, e a construção de modelos mais simples e precisos (TAN, 2007).

A seleção dos parâmetros foi realizada por meio de algoritmo genético porque esta técnica oferece a possibilidade de soluções elegantes com rápida convergência em espaços não lineares e mal compreendidos (SHAHAMAT; POUYAN, 2015). Além disso, pode-se escolher, ou não, características por meio de recursos booleanos considerando sempre aquelas que beneficiam a classificação e evitam os máximos locais devido a uma certa aleatoriedade intrínseca (ZAMALLOA et al., 2006).

Apesar de extensa pesquisa bibliográfica, não foram encontradas quaisquer abordagens paralelas a esta para a resolução do problema apresentado, qual seja, a classificação de falas de pessoas embriagadas.

1.3 Objetivo e Hipótese

Como é possível inferir de subseção anterior, a sociedade sofre com o problema da alcoolemia de motoristas irresponsáveis ao dirigir. O poder público responde com o endurecimento dos normativos legais, numa tentativa de suprir os anseios da população que clama por uma solução.

Contudo, isso acontece de forma desordenada e sem a observação de normativos superiores, pois se ao contrário fosse, não haveria a necessidade de uma frequente modificação das leis e decretos que regulam o tema, conforme listado no breve histórico já comentado. O resultado é o constante aumento no número de acidentes e do número de vítimas em tratamento pelo Sistema Único de Saúde (SUS), sem falar no custo incalculável que é a perda da vida.

Assim sendo, o objetivo deste trabalho é contribuir com a sociedade, e em especial com a Criminalística, propondo um modelo de análise que permita a indicação do estado de alcoolização de pessoas por meio de suas falas. Tal modelo de análise se baseia na análise cepstral e de seleção de características por meio de algoritmo genético.

A hipótese é de que a seleção das características por meio de algoritmos genéticos irá reduzir de forma substancial a quantidade de parâmetros irrelevantes, melhorando consideravelmente as taxas de classificação. Além disso, tem-se a expectativa que o processo de seleção por esta metodologia deverá performar melhor que PCA (*Principal Component Analysis*) e LDA (*Linear Discriminant Analysis*) para a classificação de falas de indivíduos intoxicados pelo álcool, assim como obtido para reconhecimento de locutor (ZAMALLOA et al., 2008).

1.4 Organização da dissertação

Seguindo o exposto noutra parte, a dissertação está estruturada conforme a descrição a seguir:

O capítulo 2 relata a revisão bibliográfica realizada dos conceitos necessários para o andamento da pesquisa, assim como apresenta o estado da arte no tema deste trabalho, bem como para os estudos de reconhecimento de emoções e de detecção de doenças por meio da voz, por similaridade ao trabalho desenvolvido.

O capítulo 3 apresenta o banco de vozes utilizado e suas características. Uma exposição geral do *corpus* utilizado se torna necessária para expor suas qualidades e defeitos, pois a forma como foram estabelecidas as gravações deste banco de vozes interfere nos resultados finais da aplicação proposta.

O capítulo 4 apresenta uma introdução ao algoritmo genético. Ao longo do texto serão expostas as formas de parametrização do algoritmo, bem como as principais operações genéticas.

O capítulo 5 apresenta a técnica proposta para a detecção da intoxicação alcoólica. Além disso, avalia-se o modelo apresentado. Faz-se também uma comparação da técnica implementada com outras constantes na literatura científica que utilizaram o mesmo *corpus* de vozes com parâmetros acústicos da voz. Além disso, utiliza-se uma técnica recursiva de eliminação de parâmetros para avaliar o algoritmo genético.

Por fim, o capítulo 6 apresenta as conclusões obtidas com o desenvolvimento do modelo, assim como sugere trabalhos futuros.

2 Revisão bibliográfica e o estado da arte

Este capítulo é subdividido em quatro seções. Na seção 2.1 são apresentados os conceitos básicos de Fonética Forense. A seção 2.2 apresenta um breve histórico de como se deu o início das perícias de identificação de intoxicação alcoólica. Já a seção 2.3 apresenta os estudos estabelecidos como estado da arte nas perícias de áudio, entre os quais as pesquisas de identificação de doenças por meio da voz, reconhecimento de sotaques e de identificação da intoxicação alcoólica. Por fim, a seção 2.4 apresenta as considerações finais.

2.1 O básico da fonética acústica

Diz-se que a voz é o resultado da excitação do ar, que ao sair dos pulmões, atravessa os subsistemas laríngeo e/ou supralaríngeo. O primeiro é constituído de cartilagens que interligam os músculos que controlam as cordas vocais, por meio da abertura ou fechamento da glote. O segundo se estende desde as pregas vocais até a boca e as narinas, sendo também denominado de trato-vocal. A Figura 2.1 ilustra os órgãos que compõem o sistema vocal humano.

As pregas vocais são formadas por uma cobertura e um corpo. A primeira é formada pelo epitélio e pela camada superficial da lâmina própria. Por sua vez, o corpo é formado pelo músculo vocal, responsável pelos movimentos de adução e tensão das pregas vocais (VALE, 2012). A Figura 2.2 ilustra a visão superior da glote, onde é possível observar as cordas vogais, que podem vibrar de três modos distintos (BARBOSA; MADUREIRA, 2015):

- **Voz modal:** Aquela em que as pregas vocais vibram em toda a extensão com aritenóides aproximadas, sendo a fase de fechamento menor que a de abertura, mas com diferença pequena;
- **Voz soprosa:** Aquela em que as pregas vocais estão um pouco mais afastadas, mas ainda vibrando integralmente ou somente na parte anterior. As fases de aberturas são equivalentes e há muita energia na frequência fundamental.

- **Voz crepitante ou laringalizada (*Creaky voice*):** Aquela em que as aritenóides estão sob forte compressão e a vibração das pregas vocais é parcial. Há irregularidade do intervalo entre os pulsos glotais consecutivos, que é denominada *jitter*. Nesse modo, os pulsos glotais apresentam fechamento abrupto.

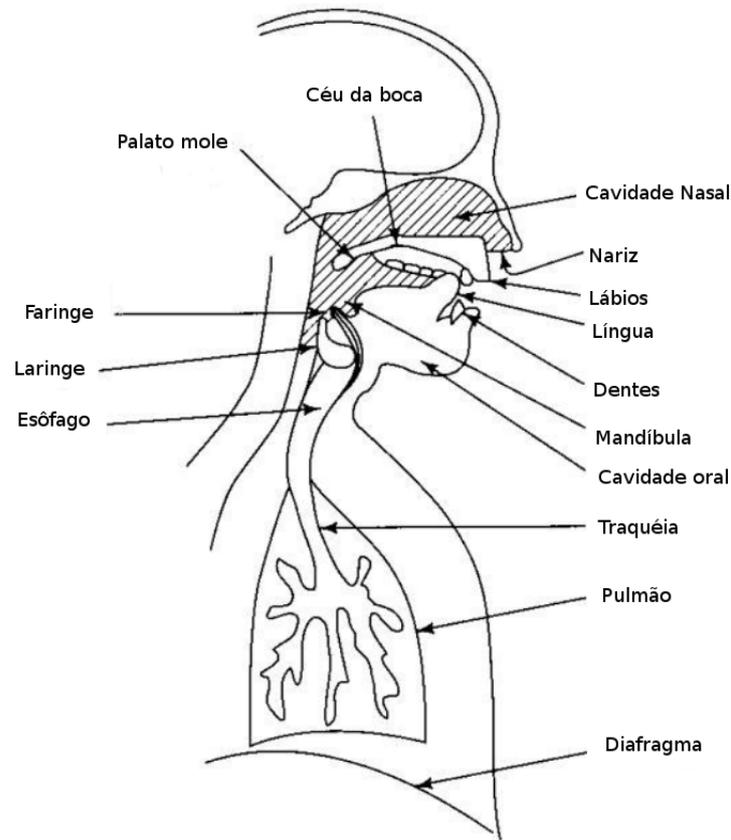
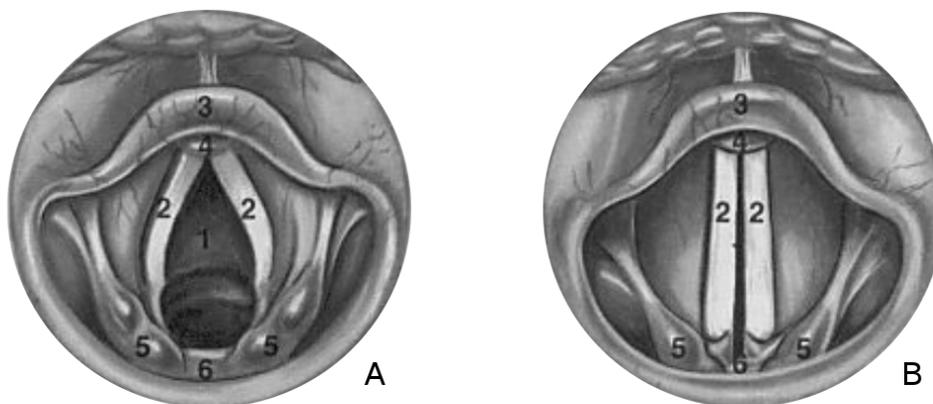


Figura 2.1: Sistema vocal humano. Adaptada a partir de (TRASCASA, 2013), p. 3.



1 - Glote 2 - Cordas vocais 3 - Epiglote 4 - Comissura anterior 5 - Cartilagens aritenóides 6 - Comissura posterior

Figura 2.2: Vista superior da laringe mostrando as posições da glote. (A) Glote na posição de repouso. (B) Glote em atividade. Adaptada de (CROVATO, 2004), p. 41.

Para que a voz possa ser compreendida, do ponto de vista segmental, é necessário a ocorrência de co-articulação por co-produção e modulação. Isso se dá de forma natural para pessoas saudáveis e é o resultado do que entendemos como sílabas. A partir daí se pode dizer que a voz transformou-se em fala; atividade complexa que envolve diversas musculaturas que são controladas de forma harmônica pelo cérebro. Os articuladores podem ser classificados em ativos e passivos (ROSE, 2002). A Figura 2.3 ilustra os articuladores ativos e passivos.

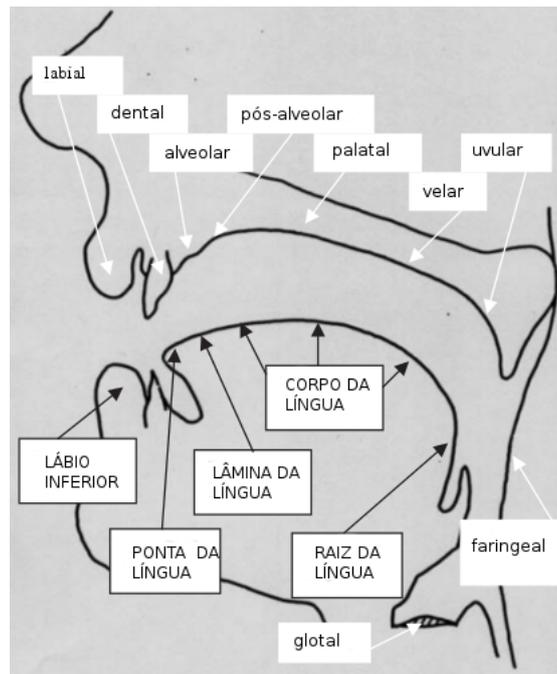


Figura 2.3: Seção do eixo sagital do trato vocal, onde se demonstram os articuladores ativos (maiúsculos) e passivos (minúsculos). Adaptada de (ROSE, 2002), p. 153.

A teoria moderna de produção da fala estabelece a nítida separação entre dois sistemas: a fonte e o filtro (Teoria Fonte-Filtro de Produção da Fala) (BARBOSA; MADUREIRA, 2015).

A fonte é o sítio da geração da energia da fala, que está localizado entre a glote e os lábios. Essa fonte pode ser periódica, aperiódica ou a combinação de ambas, dependendo do vocábulo a ser produzido.

As ondas periódicas podem ser decompostas em frequências fundamentais e suas frequências múltiplas (harmônicas), sendo de grande importância para a fonética acústica. Obviamente suas harmônicas tem amplitude menor que a sua frequência fundamental associada. A frequência fundamental apresenta variabilidade intra-falante devido a fatores linguísticos (entoação para ênfase, por exemplo), paralinguísticos (estado psicológico, por exemplo) e extralinguísticos (*creaky voice*, por exemplo). Além disso, tais

variações extra-falantes podem também ser explicadas devido às diferenças morfológicas de cada indivíduo (dimensões e massa), assim como nas divergências no controle da vibração das pregas vocais.

Por sua vez, o filtro é o trato vocal. Este conjunto de órgãos é composto por diferentes tipos de tecidos, em relação aos seus graus de elasticidade e rigidez, os quais influenciam sobremaneira na relação entre a absorção e a reflexão das ondas sonoras (BRANDÃO, 2011).

Na literatura científica são apresentadas diversas formas de modelá-lo. Contudo, o modelamento mais comum é aquele que compara o trato vocal a um tubo ressonante, que favorece algumas frequências de acordo com a configuração do posicionamento dos órgãos envolvidos. A Figura 2.4 ilustra um esquemático composto de um sistema de tubos ressonantes para representar o trato vocal.

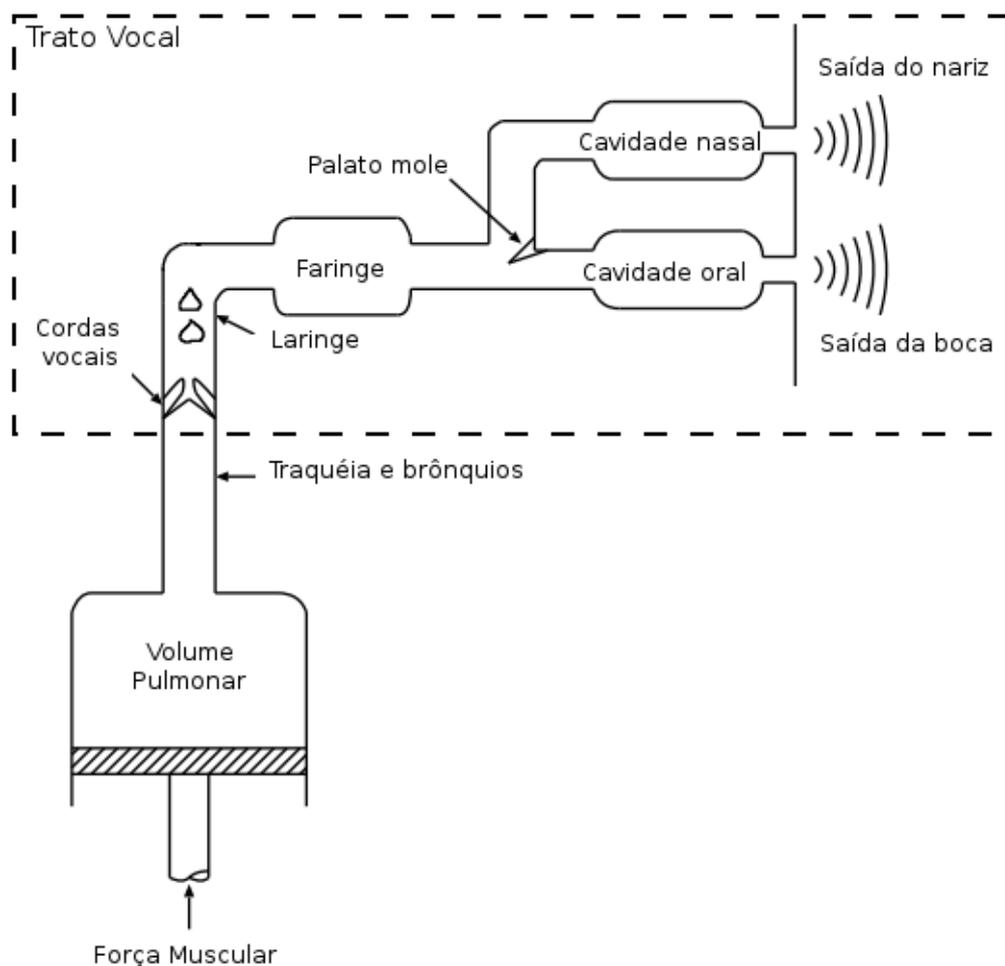


Figura 2.4: Representação esquemática do mecanismo de produção da voz. Em detalhe, no interior do retângulo tracejado, os órgãos que compõem o trato vocal. Adaptada de (NEY, 2003), p. 8.

Como dito anteriormente, os órgãos do trato vocal se configuram para permitir a emissão de determinados vocálicos. Alguns estudos demonstram relações entre F1 e F2, que são as duas primeiras harmônicas do sinal sonoro emitido, com a altura do sistema línguo-mandibular e a posição de constrição no eixo sagital. Quanto mais alto o sistema línguo-mandibular, menor é o valor de F1 e quanto mais anterior for a constrição maior será o valor de F2 para vogais (BARBOSA; MADUREIRA, 2015).

Em (SANTOS, 2013) foi demonstrada uma análise fonético-acústica das vogais orais e nasais nas línguas portuguesa brasileira e europeia, considerando as variações intergênero. A autora constatou que os triângulos articulatórios contidos no espaço acústico gerado pelas vozes masculinas geralmente se encontram no interior daqueles produzidas pelas vozes femininas. Tais resultados são, obviamente, relacionados às diferenças na anatomia das pregas vocais envolvendo ambos os gêneros. A Figura 2.5 ilustra o espaço acústico das vogais orais para o português brasileiro.

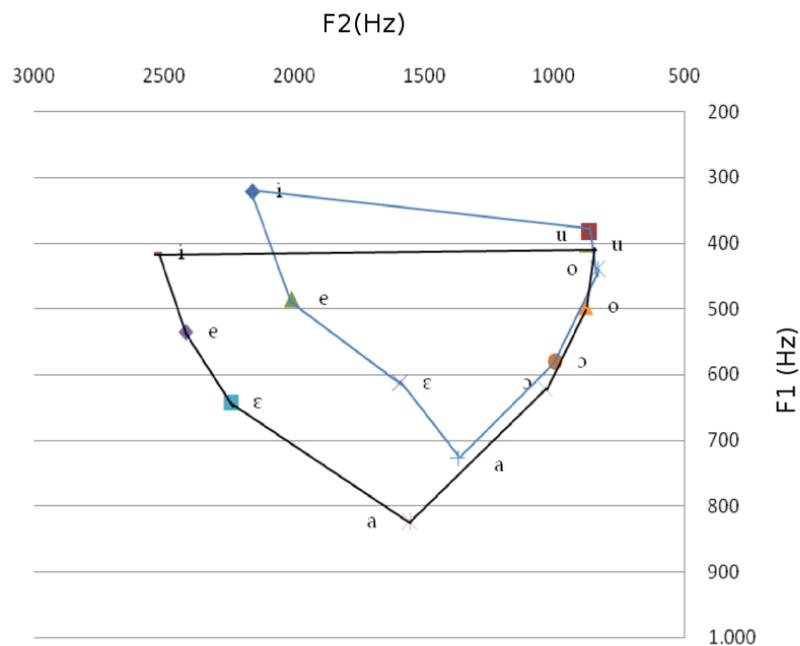


Figura 2.5: Espaço acústico das vogais orais obtido para o português brasileiro. Adaptada de (SANTOS, 2013), p. 77.

A modificação do espaço acústico das vogais, motivada pela intoxicação alcoólica, foi demonstrada em (SCHIEL; HEINRICH; NEUMEYER, 2010) para locuções de mulheres falantes da língua alemã. A Figura 2.6 ilustra a abertura do triângulo acústico com a intoxicação alcoólica. O autor justificou este fenômeno afirmando que há alteração morfológica do trato vocal, devido ao aumento de massa das pregas e ao ressecamento do canal após a administração de bebida alcoólica.

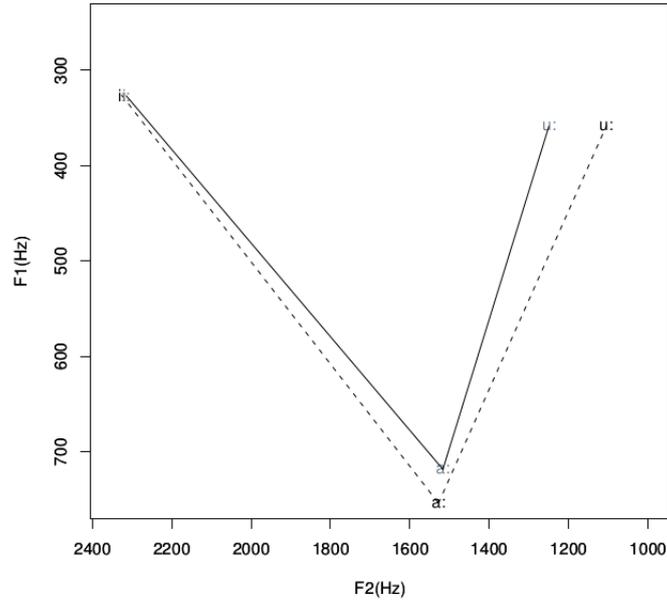


Figura 2.6: Espaço acústico das vogais a, i e u para falas de pessoas nos estados sóbrio e alcoolizado, obtido para falantes do sexo feminino no idioma alemão. A linha contínua indica o centróide das vogais para as falas alcoolizadas, enquanto a linha tracejada o das falas sóbrias. Adaptada de (SCHIEL; HEINRICH; NEUMEYER, 2010).

Da mesma forma que para vogais, as consoantes também podem ser modeladas como um tubo acústico. Em (SILVA, 2007) é apresentada a classificação das consoantes quanto aos seus modos de articulação:

- **Oclusivas:** São geradas em duas fases: Uma de oclusão e outra de explosão. A primeira fase é caracterizada nos espectrogramas pela ausência de sinal acústico, que corresponde ao momento em que os articuladores obstruem a passagem de ar no trato vocal. A segunda fase é a explosão, que nada mais é que a desobstrução instantânea da passagem do ar no trato. É vista no espectrograma como barras de alta energia. São exemplos: [ˈpɪrə], em "pira", e [kaˈbɛsə], em "cabeça".
- **Fricativas:** São geradas por meio de estreitamento voluntário em algum ponto do trato vocal. Tal estreitamento causa aumento da fricção do ar durante sua passagem, causando a sensação auditiva de ruído. São exemplos: [ʃa], em "chá", e [ʒa], em "já".
- **Nasais:** São geradas por meio do acoplamento das cavidades nasal e oral. Nesse caso, as ressonâncias geradas na primeira se sobrepõem à segunda, gerando um componente adicional nos espectrogramas, qual seja, o formante nasal. São exemplos: [aˈmɪɡʊ], em "amigo", e [kitʃiˈnɛtə], em "kitinete".
- **Vibrantes:** São geradas por meio de repetidas vibrações dos articuladores, po-

dendo ser, no caso do português, a úvula ou a ponta da língua. Estas vibrações ocorrem devido a rápidas variações da obstrução à passagem do ar ao longo do trato vocal. São exemplos: [ro'naldo], em "Ronaldo", e [ru'a] em "rua".

- **Tapes:** São geradas de modo similar às anteriores, contudo há apenas um único período de obstrução à passagem do ar. São exemplos: [pra'to], em "prato", e [car'ta] em "carta".
- **Laterais:** São geradas por meio de bloqueio no centro do trato vocal, fazendo com que o ar se propague pelas laterais. Por exemplo: [pa'ʎasɔ], em "palhaço", e [ga'lɛrɔ], em "galera".
- **Aproximantes:** As aproximantes são conhecidas como semi-vogais, sendo representadas pelos grafemas "i" e "u". São produzidas por meio da aproximação de dois articuladores, mas sem o estreitamento do trato vocal. São exemplos: [cai'ʃa], em "caixa", e [au'finete], em "alfinete".
- **Africadas:** É uma espécie de som que envolve mais de um modo de articulação. Num primeiro momento há uma obstrução do trato vocal, para depois ocorrer o seu estreitamento, fazendo com que o ar se propague por ele de modo turbulento. Sendo assim, as consoantes africadas tem características das oclusivas e das fricativas. São exemplos: [ti'tulo], em "título", e [di'vida], em "dívida".

É possível, assim como para as vogais, realizar a extração de suas formantes por meio de LPC (*Linear Predictive Coding*). Contudo, no caso das consoantes há a possibilidade da ocorrência de antiformantes. Conforme definido em (BARBOSA; MADUREIRA, 2015), antiformantes são formantes geradas num tubo acústico que se encontram numa posição que não permite que seus efeitos de ressonância se exteriorizem pelos lábios. Isso acontece frequentemente com as vogais nasais, motivo pelo qual é aconselhável não utilizá-las em análises de identificação de locutor.

2.2 Análise da intoxicação alcoólica por meio da voz: Aspectos históricos

Historicamente, o marco relevante para as análises forenses de presença de intoxicação alcoólica por meio da voz é o acidente ocorrido com o navio petroleiro Exxon Valdez, em 24 de março de 1989. A embarcação encalhou numa enseada em região próxima ao Alasca, causando o despejamento de aproximadamente 40 milhões de litros de óleo no mar. Este evento ainda figura entre os maiores desastres ambientais já ocorridos.

Durante as investigações do sinistro surgiu a hipótese de embriaguez do capitão da embarcação. Assim, os áudios das conversas ocorridas na cabine foram encaminhados à perícia. Para iniciar os estudos, os peritos separaram os estudos em 5 períodos, conforme a Tabela 2.1.

Tabela 2.1: Variações encontradas pelos peritos nas falas do comandante do petroleiro Exxon Valdez, conforme apresentado em Johnson (JOHNSON; PISONI; BERNACKI, 1990). Adaptada de (HEIGL, Fall 2011).

	-33 horas	-1 hora	0 hora	+1 hora	+9 horas
Sea [si]	[s]	[ʃ]	[ʃ]	[ʃ]	[s]
E Exxon	[s]	-	[ʃ]	-	-
Efeitos suprasegmentais	não	sim	sim	sim	não

Foi observado, principalmente, o aumento das taxas de ruído devido a palatização na produção de palavras com [s] nos áudios adquiridos no período de aproximadamente uma hora antes do acidente, quando comparados com outras ocorrências de [s] que se deram em até trinta horas anteriores ao sinistro. Além disso, foram observados os efeitos suprasegmentais como, por exemplo, a diminuição das variações de *pitch* e o abaixamento da frequência fundamental.

2.3 O estado da arte

Durante a condução deste trabalho foi buscada vasta bibliografia como o objetivo de encontrar o estado da arte em detecção do nível de alcoolização de indivíduos. Além disso, foram buscados trabalhos em áreas similares, que também serviram de apoio ao estudo desenvolvido ao longo desta dissertação, conforme listado abaixo.

- Identificação de doenças
- Reconhecimento de sotaques;
- Identificação de emoções;
- Identificação de intoxicação alcoólica

2.3.1 Detecção de doenças por meio da voz

Com o robustecimento das técnicas de processamento digital de sinais de voz, tornou-se cada vez mais inessante a busca pela diagnóstica de doenças por meios computacionais.

A maioria dos estudos tem como objetivo a busca pela detecção precoce de doenças envolvendo os órgãos do sistema vocal.

Em (CROVATO, 2004), por exemplo, é apresentada a classificação de seis grupos de doenças crônicas, por meio de redes neurais artificiais. Utilizou-se como características o resultado de um empacotamento *wavelet* da vozes de diversos pacientes selecionados, conforme o constante na Tabela 2.2, que procederam a emissão de uma vogal sustentada. O processo de seleção em tal trabalho atingiu, respectivamente, as seguintes taxas de erro: 12,5%, 4,68%, 12,5%, 0%, 3,12% e 10,93%.

Tabela 2.2: Tipos de grupos e doenças em que foi possível realizar a classificação por redes neurais. Adaptada de (CROVATO, 2004) p.91.

Grupo	Patologias
Laringite crônica	Laringite
Degenerativa	Carcinoma, granuloma e leucoplasia
Mobilidade incorreta	Paralisia unilateral, disfonias unilateral e hipercinética
Alterações orgânicas	Edemas de Reinke e edemas generativos
Crescimentos orgânicos	Pólipos, nódulos e cistos
Normal	Normais

Outra área de estudo é a detecção precoce de doenças neurodegenerativas, como é o caso do mal de Parkinson. Estima-se que aproximadamente 90% dos que sofrem desta doença apresentam alguma desordem em nível vocal (SHIRVAN; TAHAMI, 2011). A Figura 2.7 ilustra a comparação entre uma amostra de voz de uma pessoa saudável e outra diagnosticada com a doença de Parkinson.

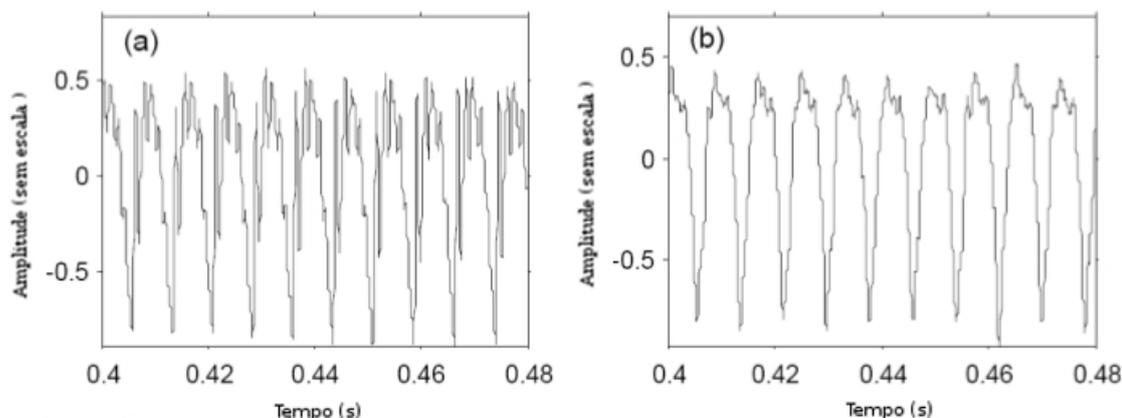


Figura 2.7: Diferença entre as amostras de dois sinais de voz. (a) Voz de uma pessoa saudável. (b) Voz de uma pessoa diagnosticada com mal de Parkinson. Adaptada de (SHIRVAN; TAHAMI, 2011).

Em (SHIRVAN; TAHAMI, 2011), foi estudada a emissão de vogais sustentadas de 32 pessoas, onde 23 delas encontravam-se afetadas pelo mal de Parkinson. Foram extraídas 13 características: média da frequência fundamental, máximo da frequência fundamental, mínimo da frequência fundamental, *jitter* (%), *jitter* (ABS), *jitter* (RAP), *jitter* (DDP), *shimmer*, *shimmer* (dB), *shimmer* (APQ3), *shimmer* (APQ5), *shimmer* (DDA) e HNR (*Harmonics to Noise Ratio*).

Já em (TSANAS et al., 2012) é apresentada a análise de 132 características de disфонia em emissões vocais de vogal sustentada de 43 pacientes de ambos os sexos, entre os quais 33 se apresentavam com mal de Parkinson. Entre os parâmetros de disфонia examinados encontravam-se *jitter*, *shimmer*, medições de ruído glotal. Contudo, também foram consideradas características cepstrais extraídas por meio de coeficientes MFCC. A classificação das vozes foi realizada por meio de SVM (*Support Vector Machine*) e Florestas Randômicas, contudo a primeira se mostrou superior para esta aplicação, alcançando 99% de assertividade.

Em (XIAO, 2012), apresenta-se um estudo da fonação de vogais sustentadas de 31 pessoas de ambos os sexos, onde 23 delas foram previamente diagnosticadas com a doença de Parkinson. O estudo considerou a extração de características de disфонia e realizou a comparação dos seguintes classificadores: SVM, Árvores de Decisão, KNN (*K Nearest Neighbor*) e PNN (*Probabilistic Neural Network*). O SVM se mostrou superior para o tipo de análise em questão, alcançando uma taxa de 96,4% de assertividade.

Em seguida, foi introduzido o algoritmo genético para a tarefa de seleção de parâmetros. Para tanto, utilizou-se como população um vetor binário com dimensões equivalentes ao número de parâmetros extraídos inicialmente. Se um *bit* do vetor é 1, significa que a característica de mesma ordenação que este *bit* é selecionada e, por sua, se o *bit* tiver valor 0, a característica é excluída. Nos testes foram utilizados como probabilidade de *crossover* e de mutação 0,85 e 0,05, respectivamente. O algoritmo convergiu em 20 iterações e alcançou a assertividade média equivalente a 98,4%, considerando como classificador o SVM.

2.3.2 Detecção de sotaques

Em (HUANG; WANG, 2016) é apresentada uma metodologia para classificar a língua materna dos falantes, reconhecendo, a partir do inglês, se o indivíduo se origina de países de língua inglesa, identificando-os, ou de países de idiomas diversos. As características da voz utilizadas para tanto foram *pitch*, timbre e ritmo, coeficientes cepstrais,

bem como coeficientes delta e delta-delta e MFCC, acumulando o montante de 146 descritores. A Figura 2.8 demonstra o esquemático básico proposto neste trabalho para detecção de sotaques.

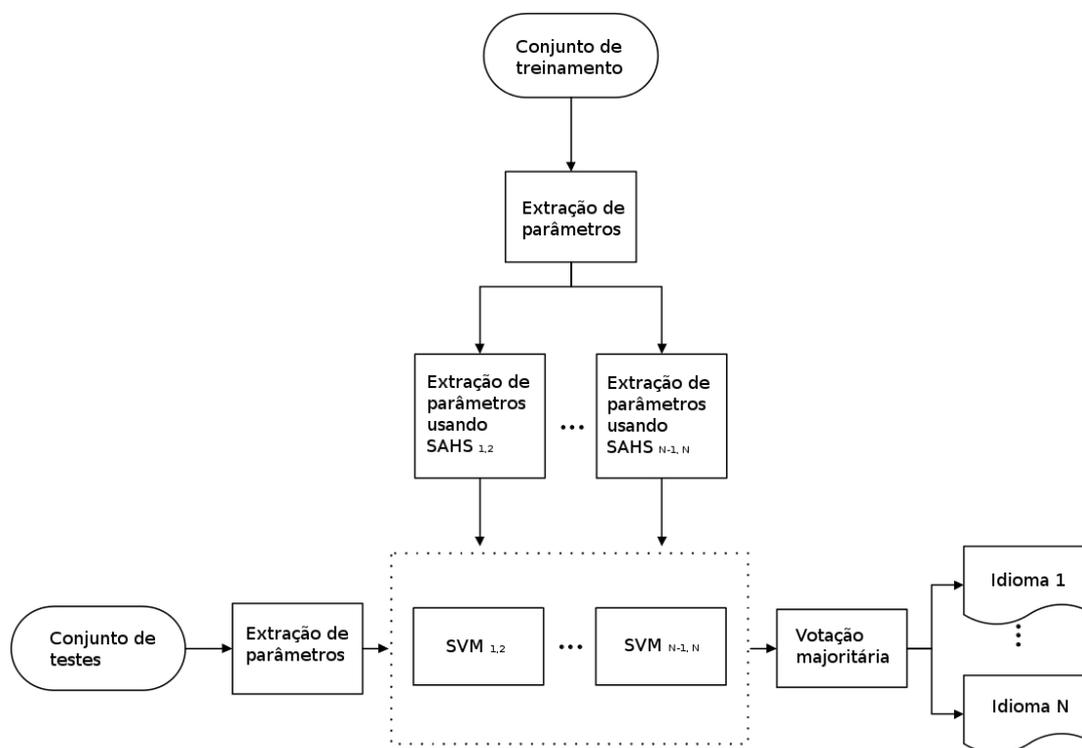


Figura 2.8: Diagrama de blocos proposto para detecção de sotaques. Adaptada de (HUANG; WANG, 2016). p.4.

O sistema implementado é composto de uma fase de treinamento e uma fase de testes. Durante a fase de treinamento todos os 146 descritores são extraídos e, em seguida, passam por um algoritmo de otimização metaheurístico, SAHS (Self-adaptative Harmony Search) (WANG; HUANG, 2010), com o objetivo de selecionar as características mais relevantes para treinar os classificadores SVM. O número de classificadores é dado por $N(N-1)/2$, onde N é o número de amostras da primeira língua a ser treinada.

Na fase de testes todas os descritores são calculados e encaminhados para classificadores SVM implementados. O resultado final é dado por uma estratégia de votação majoritária.

O banco de dados treinado continha gravações de falantes com origens nos seguintes países de língua inglesa e não inglesa: Canadá, USA (*United States of America*), Grã-Bretanha, Nova Zelândia e Austrália e Índia, Coréia, Japão, China e Taiwan. Foi possível obter resultado de até 80,75% e 84,5% de precisão, respectivamente, na seleção de falas de indivíduos que tiveram origem em países de língua inglesa e não-inglesa.

2.3.3 Detecção de intoxicação por meio do álcool

O álcool é uma droga psicotrópica pois atua no sistema nervoso central, provocando mudanças no comportamento dos indivíduos, as quais se caracterizam por duas fases distintas. Na primeira, podem aparecer os efeitos estimulantes como euforia e desinibição. Com o tempo, aparecem os efeitos depressores como falta de coordenação motora, descontrole e sono. Estes efeitos se refletem na voz, o que traz uma perspectiva de percepção por meios automáticos.

Em (BARFÜßER; SCHIEL, 2010), é demonstrado que a alcoolização tem efeito significativo em diversas características da fala/voz das pessoas de ambos os sexos, entre elas o aumento do número e do tempo de pausas, do número de sons não-usuais e de interrupções.

Em (BAUMEISTER; SCHIEL, 2010), é apresentado o resultado do estudo do efeito do álcool nas frequências fundamentais da voz. Ficou constatado que as vozes dos indivíduos de ambos os sexos são afetadas pela contaminação alcoólica e que o tempo médio de fala das palavras é sensivelmente aumentado, o que também já tinha sido afirmado em (CHIN; LARGE; PISONI, 1996). Em estudo similar, constatou-se a perda da fluência na fala dos indivíduos intoxicados pelo álcool (HOLLIEN et al., 2001).

Em (LEVIT et al., 2001), utilizaram-se as características de prosódia, em nível segmental, para construir uma rede neural tendo separado o sinal em *frames* de 10 ms. Para a construção de seu modelo foram considerados os parâmetros de frequência fundamental, cruzamento por zero, energia e *jitter*. Alcançou-se o resultado de 69% na tarefa de discriminação das falas intoxicadas e não intoxicadas pelo álcool.

Já em (BONE et al., 2011) foram utilizados parâmetros hierárquicos normalizados de cada falante, juntamente com vetores GMM (*Gaussian Mixture Model*), para analisar a intoxicação alcoólica. Utilizando como classificador o SVM, concluiu-se com acurácia de 65,9% na definição da segregação das condições de sobriedade e intoxicação.

2.4 Considerações Finais

Nesta seção foi apresentada parte dos conceitos necessários para a compreensão da pesquisa relatada neste texto. Entre os conceitos revisados estão o básico de fonética acústica e a teoria acústica da fala. Além disso, aduziu-se o histórico do início dos exames de identificação de intoxicação alcoólica e o estado da arte.

Em relação ao estado da arte, foram expostos trabalhos que apresentam similaridade com o tema desta pesquisa. Apesar da extensa pesquisa realizada, não foram encontrados trabalhos similares ao proposto neste texto. Contudo, o trabalho de identificação de emoções por meio da voz, utilizando algoritmo genético, apresenta alguma similaridade.

3 *Corpus* de falas alcoolizadas

Neste capítulo é apresentado o banco de vozes que serviu de subsídio ao desenvolvimento deste trabalho. A seção 3.1 introduz o *corpus* e suas características principais. A seção 3.2 indica quais foram os critérios para a configuração das palavras escolhidas para compor o banco de vozes. Já a seção 3.3 aduz ao leitor os critérios que foram observados para a gravação das falas, bem como para a geração dos metadados. Por fim, a seção 3.4 apresenta as considerações finais.

3.1 Introdução

Obviamente, a alcoolização traz diversos efeitos físicos e psicológicos que variam desde fadiga, excitação, estresse e desinibição até a perda de coordenação motora. Tais efeitos estimularam os pesquisadores a desenvolverem diversos estudos ao longo dos anos. Contudo, havia a dificuldade de se analisar os efeitos do álcool na fala devido à inexistência de um *corpus* amplo, que contivesse diversas vozes de pessoas de ambos os sexos em condições de intoxicação e de sobriedade.

Tais problemas se reduziram sensivelmente com o surgimento do *corpus* de vozes *Alcohol Language Corpus* (ALC), que é uma parte do banco de vozes *Bavarian Archive for Speech Signals* (BAS), fornecido gratuitamente para pesquisas pela Universidade de Munique. Além de gravações de vozes intoxicadas pelo álcool, também é fornecido um conjunto de metadados, que fornece as características dos voluntários que forneceram suas falas, entre as quais o nível de intoxicação alcoólica ao qual foram submetidos.

O ALC é composto por diversos arquivos de áudio de falas de 162 pessoas, de ambos os gêneros em percentuais iguais, nos estados sóbrio e alcoolizado. Os voluntários tem idades variando entre 21 e 64 anos, de nacionalidade alemã, de diversas regiões daquele país. Sendo assim, o idioma padrão deste *corpus* é o alemão. Em (SCHIEL et al., 2010), são apresentados os principais motivos para a elaboração de tal *corpus*, que foram:

- **Ciências Forenses da fala:** Os estudos produzidos anteriormente a este *corpus* teriam apresentado investigações parcialmente inconsistentes a respeito das descobertas sobre como a intoxicação afeta a fala. A maioria destes estudos tinha

analisado menos de 40 falantes do sexo masculino, em condições de laboratório. Além disso, geralmente a intoxicação era medida por etilômetros ou estimada a partir da ingestão de bebidas. No ALC os 162 falantes foram submetidos a testes por meio de etilômetros e por meio de análise sanguínea;

- **Ciências biométricas:** A comunidade científica tem buscado incessantemente técnicas e parâmetros que objetivam a realização da identificação de indivíduos por meio da voz, bem como a identificação de idade, sexo, interesse e emoções dos falantes. Contudo, o comportamento da variável intoxicação alcoólica geralmente é desconhecido. Assim, tal *corpus* foi implementado para contribuir para tal investigação;
- **Detecção alcoólica em ambiente automotivo:** Sabendo que a intoxicação alcoólica é uma das maiores causas de acidentes de trânsito e que as medições existentes de intoxicação alcoólica somente são realizadas após a ocorrência de algum sinistro, considerou-se a possibilidade de implementação de algo preventivo. Assim surgiu o conceito deste *corpus*, que observa as falas de controle típicas para aplicações automobilísticas, de modo a possibilitar o treinamento de detecção automática de alcoolização.

3.2 Desenvolvimento do *corpus*

Durante a construção do *corpus*, a mensuração da graduação alcoólica a que estavam submetidos os indivíduos voluntários foi realizada por meio exames de sangue. Alguns estudos de análise de pessoas alcoolizadas utilizam os etilômetros em faixas, conforme determinado pela lei. Contudo, os valores disponibilizados pelos etilômetros tendem a se correlacionar com o nível de álcool no sangue, mas, obviamente, não são 100% confiáveis.

Durante um teste piloto foram realizados estudos comparativos entre as mensurações dos níveis de alcoolemia por meio de etilômetro e de exame de sangue com 152 pessoas embriagadas. Tal teste encontrou uma correlação de Pearson de 0,89. Sendo assim, os desenvolvedores do ALC decidiram pela aplicação de testes de alcoolemia por meio sanguíneo para todos os experimentos na geração deste *corpus*. Contudo, também disponibilizaram as medições por meio de etilômetros.

Tal *corpus* de dados tinha como preceito realizar a coleta de vozes do maior número de pessoas possível para que pudesse reproduzir resultados significativos. Uma das

maiores barreiras enfrentadas durante a produção deste *corpus* era como garantir a fala espontânea. A maioria dos estudos requisitava estímulos por meio de leitura em papel. O ALC contém gravações de diversos tipos de falas, conforme listado na Tabela 3.1. Todas as gravações foram realizadas em ambiente automobilístico, em dois automóveis.

Tabela 3.1: Tipo de falas existentes no *corpus* ALC. Adaptada de (SCHIEL et al., 2010)

Tipo de Falas	Tipo do item falado	Intoxicado/controle	Sóbrio
Fala lida	cadeia numérica	5	10
	trava línguas	5	10
	comando lido	4	9
	endereço	5	10
	soletração	1	1
Fala espontânea	descrição de figuras	2	4
	resposta de perguntas	1	1
	comando espontâneo	5	10
	diálogo	2	5
Total:		30	60

Durante o projeto do *corpus*, foram previstas falas espontâneas e de comando. Neste contexto, foram utilizados fonemas que a literatura científica apontava como sendo afetados pela intoxicação alcoólica. São utilizados endereços de rua que são considerados de difícil pronúncia.

Os endereços proferidos são reais. Foram selecionados a partir de um banco de dados geográfico e apresentam pronúncia complexa, como por exemplo: *Schwester-Hermenegildis-Strasse*. Também foram escolhidos nomes de cidades para soletração. Quanto às descrições, o voluntário tinha até sessenta segundos para apresentar uma resposta às perguntas como, por exemplo, "Conte-me sobre suas últimas férias".

3.3 Procedimento de gravação dos discursos

Conforme disposto em (SCHIEL et al., 2010), as falas contidas no *corpus* foram proferidas por voluntários, que assinaram um termo de consentimento para o uso científico e técnico do discurso gravado, sob a condição de que o conteúdo de áudio não pode ser associado a dados pessoais. Todo o processo de teste foi avaliado e supervisionado pelo Instituto de Medicina Legal da Universidade de Munich. Um detalhe importante sobre a metodologia estabelecida: Antes de cada teste real, cada voluntário escolheu

a taxa de concentração de álcool no sangue que desejou alcançar durante o teste de intoxicação, que variava entre 0,3% e 1.5%.

Após esta escolha da taxa de alcoolização que o voluntário desejava alcançar, foi utilizada a Equação 3.1 para estimar o volume (v) de álcool que ele deveria ingerir, considerando os dados de massa (m), concentração (c) e uma taxa de redução (r), definida pela Equação 3.2.

$$V = cmr \quad (3.1)$$

$$r = \frac{\rho_b g}{f_m} = 1,31875g \quad (3.2)$$

A constante g é definida pelas Equações 3.3 e 3.4, onde: t = idade (anos), h = altura (cm) e m = massa (kg).

$$g_{female} = 0.03 - 0.07t + 0.1069h + 0.2466m \quad (3.3)$$

$$g_{male} = 2.447 - 0.09516t + 0.1074h + 0.3362m \quad (3.4)$$

Após os cálculos, os voluntários foram submetidos às quantidades estipuladas de vinho ou cerveja. Após decorrido o prazo de 20 minutos, foram submetidos a mensuração do nível de alcoolização por meio de exame sanguíneo e etilômetro. Em seguida iniciaram-se as gravações das falas.

Como curiosidade, em (SCHIEL, 2011) foi descrito que a quantidade média de palavras faladas pelas mulheres nos estados sóbrio e alcoolizada é a mesma. Já os homens proferem menos palavras quando intoxicados pelo álcool, mas produzem áreas de silêncio menores, quando comparadas com as áreas deixadas nas falas femininas. A Figura 3.1 apresenta o histograma do nível de alcoolização a que foram submetidos os dois gêneros durante a produção do *corpus*.

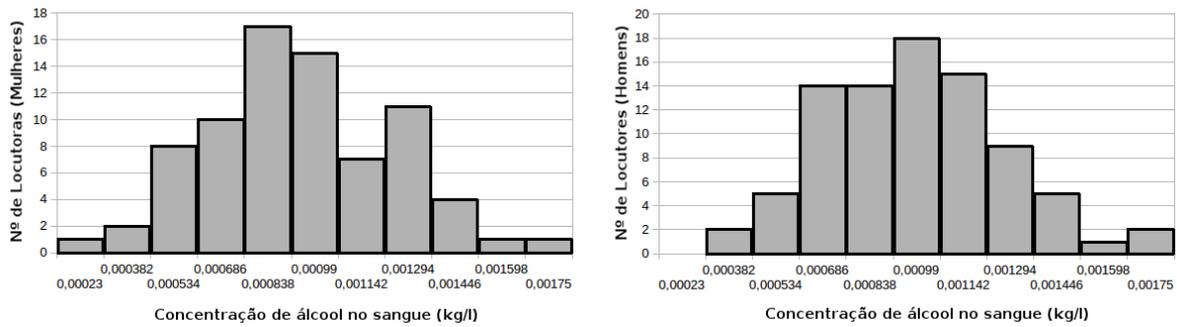


Figura 3.1: Histograma apontando os níveis de alcoolização a que foram submetidos os indivíduos durante a formação do *corpus* de voz.

3.4 Considerações Finais

Neste capítulo foi introduzido o *corpus* de vozes utilizado neste trabalho. Foram demonstradas suas características, assim como aduzidos os critérios que foram observados no processo de geração dos metadados e gravações dos arquivos das falas intoxicadas e sóbrias.

4 Algoritmos genéticos

Este capítulo é composto de oito seções, onde são expostos os princípios de algoritmos genéticos. A seção 4.1 apresenta, a título de considerações iniciais, as principais características e a função desse algoritmo heurístico. Na seção 4.2 são apresentadas as principais técnicas para a configuração da população inicial de um algoritmo genético. A seção 4.3 apresenta a definição de como são realizadas as avaliações durante o processamento do algoritmo. Na sequência, a seção 4.4 aduz as características dos processos de seleção nos algoritmos genéticos. Em seguida, na seção 4.5 detalham-se as metodologias para a realização dos cruzamentos genéticos. Após isso, a seção 4.6 explica o processo de mutação genética. A seção 4.7, por sua vez, define os parâmetros necessários para uma boa configuração dos algoritmos genéticos. Por fim, na seção 4.8, exprimem-se as considerações finais.

4.1 Considerações iniciais

Os algoritmos genéticos são métodos heurísticos de otimização de processos que imitam o sistema de adaptação e de evolução das espécies. Esta comparação pode ser resumida no princípio do ciclo de vida constante na natureza, onde os seres vivos nascem, crescem, reproduzem-se, envelhecem e morrem. Esta metodologia foi introduzida em (HOLLAND, 1975), bem como em (GOLBERG, 1989).

Matematicamente, os algoritmos genéticos são mecanismos de busca da melhor solução global para otimização de processos por meio da procura da melhor solução em algum máximo local (HINÇAL, 2008). Essa procura parte de uma população criada, muitas vezes aleatoriamente, que é imediatamente considerada candidata a melhor solução.

Em seguida, tal população passa por operadores genéticos e uma nova solução ótima pode ser alcançada. Caso contrário, o algoritmo manter-se-á em sua busca iterativa. Cada iteração no processo gera uma nova população, que é chamada de geração.

Os algoritmos genéticos se diferenciam de outros métodos de busca e de otimização tradicionais nos seguintes aspectos (GOLBERG, 1989):

- Trabalham com a codificação dos dados e não com eles próprios;

- Consideram uma população de resultados e não uma solução unívoca;
- Não requerem informações auxiliares do espaço de busca, utilizando-se somente de resultados da função objetivo;
- Usam regras probabilísticas para realização das buscas, não utilizando qualquer modelo determinístico.

A função do algoritmo é manter a população que tem potencial de ser uma solução ótima para o problema por meio de competição. A Figura 4.1 ilustra um diagrama de blocos genérico de um algoritmo genético.

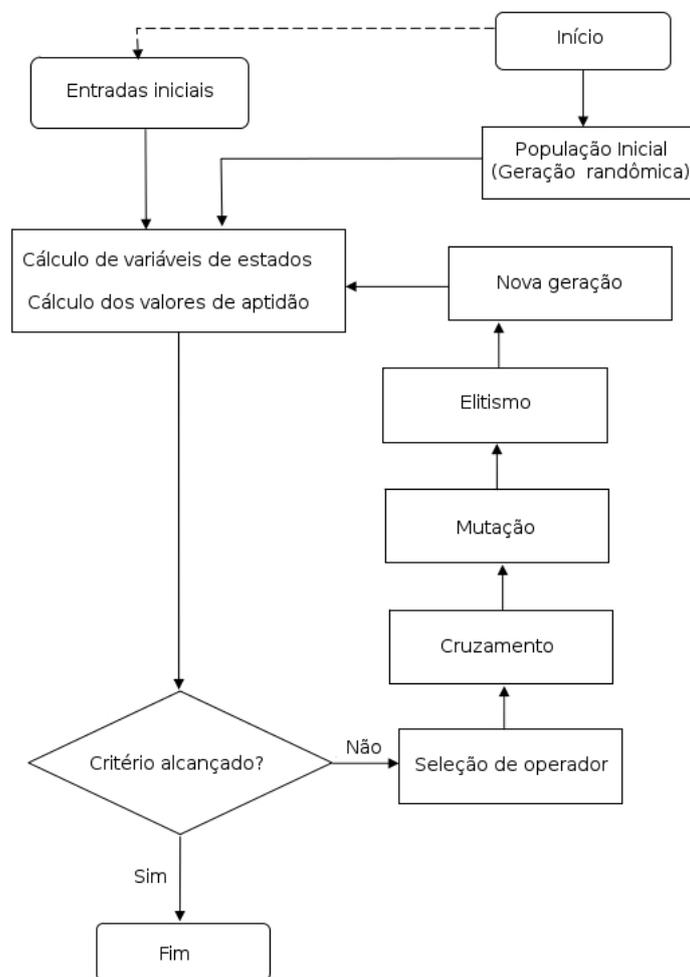


Figura 4.1: Diagrama de blocos genérico de um algoritmo genético. Adaptada de (HINÇAL, 2008), p.20.

4.2 Inicialização

O primeiro passo na configuração de um algoritmo genético é a definição de uma população inicial. A caracterização dessa população, obviamente, vai depender do tipo

de problema envolvido. Por se tratar de um processo iterativo, a escolha adequada da técnica de inicialização pode antecipar ou postergar a solução requerida pelas condições de parada.

Uma população consiste em um agrupamento de indivíduos que são comumente chamados de cromossomos. Na prática, cada cromossomo é definido por um vetor com N parâmetros, onde cada um destes é um gene. Cada um destes genes pode ser codificado por meio de números reais ou binários. A Figura 4.2 ilustra dois exemplos de cromossomos, indicando seus respectivos genes e seus tipos de codificações.

Cromossomo A:	1	0	0	1	1	1	1	1
Cromossomo B:	1,45	3,20	0,88	-0,15	1,45	0,33	7,12	2,13
	Gene	Gene	Gene	Gene	Gene	Gene	Gene	Gene

Figura 4.2: Exemplos de cromossomos e seus genes. Os genes do "Cromossomo A" estão codificados em binário, enquanto os do "Cromossomo B" estão codificados em real.

Conforme o autor de (KAZIMIPOUR; LI; QIN, 2014), os tipos de inicialização de população são classificados devido aos aspectos de Aleatoriedade, Composição ou Generalidade e são divididos conforme a Figura 4.3. A aleatoriedade diz respeito a completa imprevisibilidade, incompressibilidade e irregularidade das sequências das populações. O conceito de composição observa a quantidade de procedimentos autônomos que estão envolvidos em uma técnica. Já quando se fala em generalidade de um inicializador de população refere-se à variedade dos domínios que podem ser aplicados. As subseções a seguir descreverão cada uma das suas subclassificações.

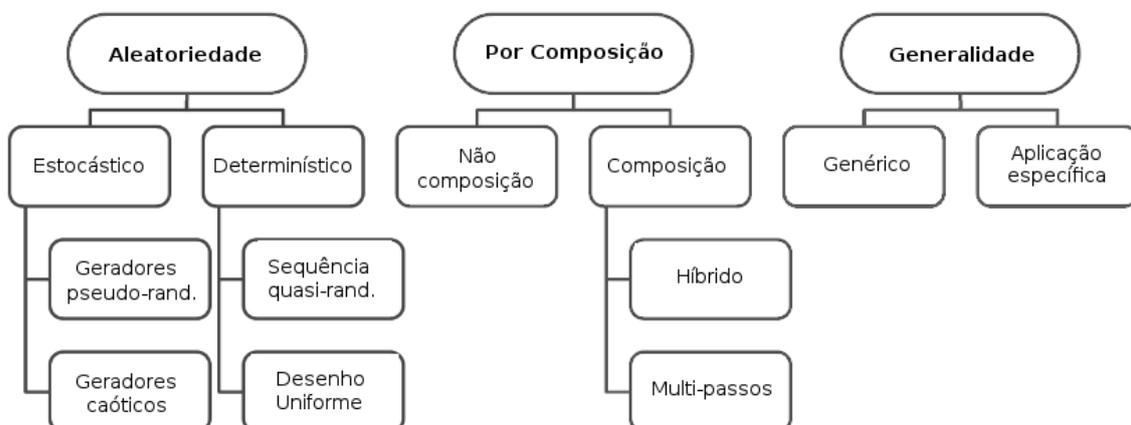


Figura 4.3: Classificação dos tipos de iniciação de população. Adaptada de (KAZIMIPOUR; LI; QIN, 2014).

4.2.1 Técnicas estocásticas

São aquelas em que os inicializadores da população são dependentes de modelos randômicos e as sementes iniciais são fornecidas unicamente por fontes externas ao processo. Se dividem em dois subgrupos:

- **Geradores de números pseudo-randômicos:** São aqueles geradores de números aleatórios implementados em máquinas computacionais, e que são, geralmente, baseados em tempo de ciclo. Assim estas funções são, na realidade, determinísticas e têm frequência e probabilidade de ocorrência rastreáveis. Contudo, são comumente utilizadas devido a sua simplicidade para implementação em qualquer linguagem de programação;
- **Geradores de números caóticos:** São aqueles geradores de números configurados por meio de um mapa caótico, construído por um conjunto de parâmetros definidos pelo usuário.

4.2.2 Técnicas determinísticas

São aquelas técnicas de iniciação projetadas para fornecer pontos ao longo de todo o universo de pesquisa. A literatura científica se refere aos geradores determinísticos como técnicas de baixa discrepância, ou seja, técnicas com alta uniformidade. Se dividem em dois subgrupos:

- **Sequência quasi-randômica:** Apesar da descrição em seu nome, as sequências não são aleatórias nem semi-aleatórias, como aquelas descritas na subseção 4.2.1. Esta técnica busca sequências de baixa discrepância se apoiando nos limites teóricos daquelas máximas discrepâncias conhecidas, assumindo que há uma correlação positiva entre esses valores da população inicial e aquele da solução ótima.
- **Desenho experimental uniforme:** Esta técnica busca realizar o preenchimento dos espaços à procura de pontos para ser uniformemente dispersos em um determinado intervalo.

4.2.3 Técnicas de Não-composição

São todas as técnicas que produzem populações em apenas um único passo, podendo o procedimento ser estocástico, determinístico, genérico ou específico.

4.2.4 Técnicas de Composição

Em oposição às técnicas de não-composição, as de composição produzem populações em mais de um passo. Estas, por sua vez, podem ser divididas em dois grupos:

- **Técnicas híbridas:** São aquelas em que cada passo pode ser aplicado separadamente, como uma técnica não-composicional.
- **Técnicas de multi-passos:** São técnicas sequenciais inseparáveis, onde nenhum dos passos pode ser utilizado de forma autônoma. Essas técnicas geralmente refinam nos passos subsequentes a população gerada anteriormente.

4.2.5 Técnicas genéricas

São aquelas técnicas de inicialização de população que podem ser aplicadas diretamente a todo e qualquer tipo de problemas de otimização.

4.2.6 Técnicas para aplicações específicas

São técnicas especialmente projetadas para lidar com um grupo seletivo de aplicações. Em seu projeto, busca-se explorar o conhecimento do domínio para evitar a visita de regiões desnecessárias, o que aumenta sensivelmente a velocidade de convergência dos algoritmos.

4.3 Avaliação

A avaliação de uma população é realizada por meio de valores da função objetivo do processo. A validação é o processo que expõe cada elemento da população à função objetivo e, ao final, realiza uma ordenação de acordo com a aptidão desta função (SILVA, 2005).

4.4 Seleção

Trata-se de uma simulação do processo de reprodução assexuada e é aplicada após o cálculo da aptidão de todos os indivíduos da população, pois este mecanismo proporciona maior prioridade aos indivíduos mais aptos. Contudo, uma seleção muito forte pode significar que indivíduos sub-ótimos, altamente aptos, assumirão o controle da população, reduzindo a diversidade necessária para novos progressos, o que resultará numa baixa velocidade de convergência à solução ótima (MITCHELL, 1998). Os principais métodos de seleção referenciados pela literatura científica são descritos a seguir.

4.4.1 Seleção por roleta

Neste tipo de operador os cromossomos são representados como se estivessem em segmentos consecutivos de uma roleta imaginária, ocupando um espaço proporcional ao seu valor de aptidão. Toda vez que a roleta para de girar, o cromossomo selecionado é copiado para a próxima população. O problema deste tipo de abordagem é que se houver divergências muito grandes entre indivíduos, havendo um muito maior que os outros, os menores terão poucas chances de serem selecionados.

A Figura 4.4 ilustra o funcionamento desta metodologia de seleção. À esquerda estão indicados os cromossomos com suas respectivas taxas de aptidão, que por sua vez, são proporcionais à área da roleta. A cada giro, o ponteiro da roleta seleciona para cópia o cromossomo apto e o passa para a direita.

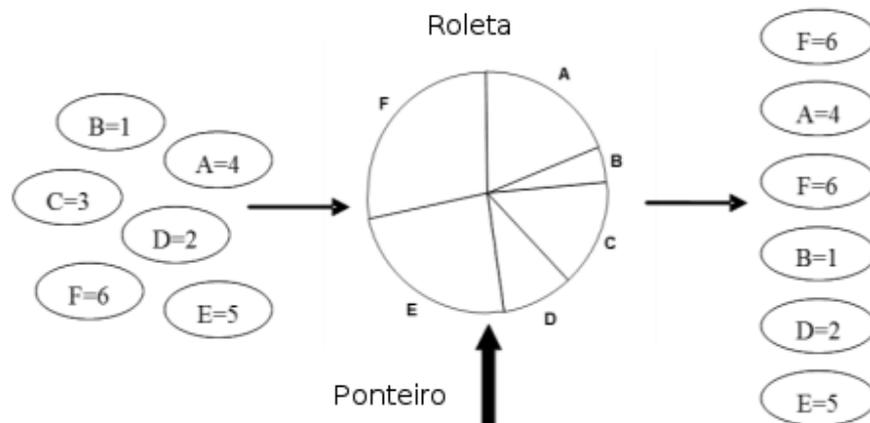


Figura 4.4: Exemplo típico da seleção por roleta. Adaptada a partir de (NORAINI; GERAGHTY, 2011).

4.4.2 Seleção por amostragem estocástica universal

A seleção por meio da amostragem estocástica universal ocorre de forma similar àquela que ocorre por meio de roleta. Os cromossomos são dispostos de forma adjacente e ocupam áreas proporcionais aos valores dados pela aptidão individual.

A grande diferença em relação à seleção por roleta é que, ao invés de um único ponteiro, é configurada uma quantidade N de ponteiros igualmente espaçados. Desta forma, uma quantidade N de indivíduos são selecionados a cada vez que a roleta girar, podendo o mesmo indivíduo ser selecionado mais de uma vez. Não há impedimento. A Figura 4.5 ilustra a técnica.

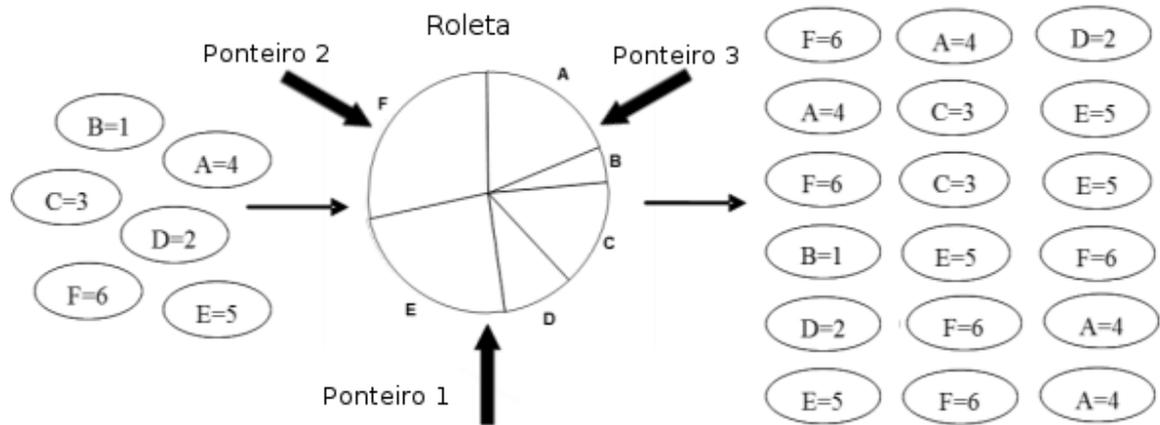


Figura 4.5: Exemplo típico da seleção por amostragem espectral universal. Adaptada a partir de (NORAINI; GERAGHTY, 2011).

4.4.3 Seleção por ranking

Esta modalidade também é similar à roleta. A grande diferença é que a probabilidade de seleção desta estratégia considera a posição relativa de cada cromossomo em toda a população, tomando como critério a ordenação crescente dos níveis de aptidão, que são ajustados por meio de uma pressão de seleção.

A técnica não é influenciada por indivíduos extremamente aptos, o que impede a precocidade da convergência do algoritmo (NORAINI; GERAGHTY, 2011). A função matemática utilizada para realizar o ajuste dos níveis de aptidão é representada pela Equação 4.1, onde AA é a aptidão ajustada, pos é a posição em relação a ordenação crescente, ps é a pressão de seleção e, por último, n é a quantidade de cromossomos presentes na população.

$$AA(pos) = 2 - ps + \frac{2(ps - 1)(pos - 1)}{n - 1} \quad (4.1)$$

A Figura 4.6 ilustra a modificação imposta pela pressão de seleção, tomando como referência os dados equivalentes aos constantes na Figura 4.4, calculados por meio da Equação 4.1 e apresentados na Tabela 4.1.

4.4.4 Seleção por torneio

Esta técnica é extremamente simples e de rápida execução. Um número de cromossomos qualquer, previamente estabelecido, é selecionado aleatoriamente para comparação e o

que tiver a maior aptidão é devidamente repassado à próxima população. Geralmente se pode alterar a pressão para influenciar a convergência da pesquisa genética. Quanto maior é o tamanho do torneio, maior é a pressão de seleção (XIE, 2008).

Tabela 4.1: Valores de aptidão ajustados por meio da pressão de seleção.

Comossomo	Ranking	Aptidão	Aptidão Ajustada PS = 1,25
B	1	1	0,75
D	2	2	0,85
C	3	3	0,95
A	4	4	1,05
E	5	5	1,15
F	6	6	1,25

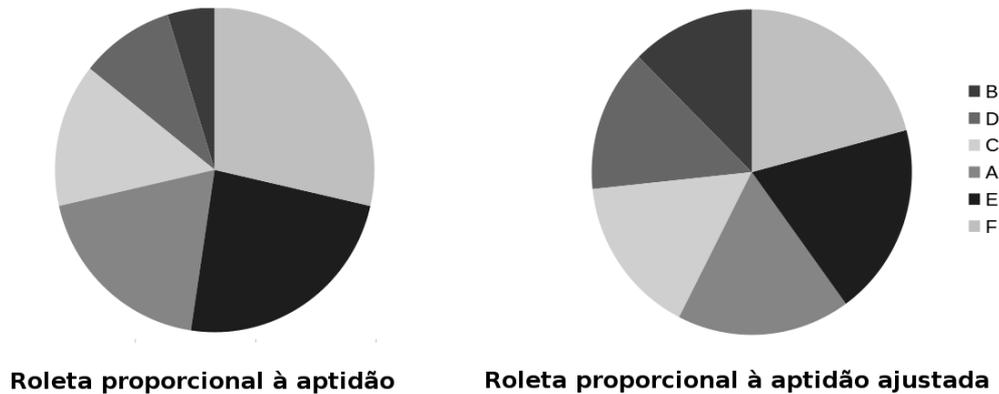


Figura 4.6: Exemplo de modificação da roleta, baseado nos dados de ordenação e de aptidão ajustada calculada por meio da Equação 4.1 e demonstrada na Tabela 4.1.

4.4.5 Seleção por NSGA-II

O NSGA-II (*Non-dominated Sorting Genetic Algorithm*) é uma técnica muito utilizada para resolução de problemas com múltiplos objetivos que foi proposta em (DEB et al., 2002). Tal algoritmo é a segunda versão daquela demonstrada em (SRINIVAS; DEB, 1994).

Conforme o contido em (GODINEZ; ESPINOSA; MONTES, 2010), este processo realiza um torneio por meio de grupos, onde o critério de julgamento é baseado nos critérios de posição e distância que cada indivíduo traz para o torneio. O NSGA-II possibilita o tratamento de vários objetivos, sem a necessidade da definição de fatores de peso (KUMAR; DAS; SHARMA, 2008).

A técnica pode ser resumida da seguinte forma: Com a população previamente definida, o algoritmo identifica as frentes de dominância e avalia as suas respectivas distâncias de aglomeração. Em seguida, usando técnicas de recombinação e mutação, gera novos descendentes. Após isso, os pais e os filhos são reagrupados e classificados frente às fronteiras de dominância e eliminam-se aqueles com a distância à frente de dominância. A Figura 4.7 ilustra o diagrama esquemático do funcionamento do algoritmo NSGA-II.

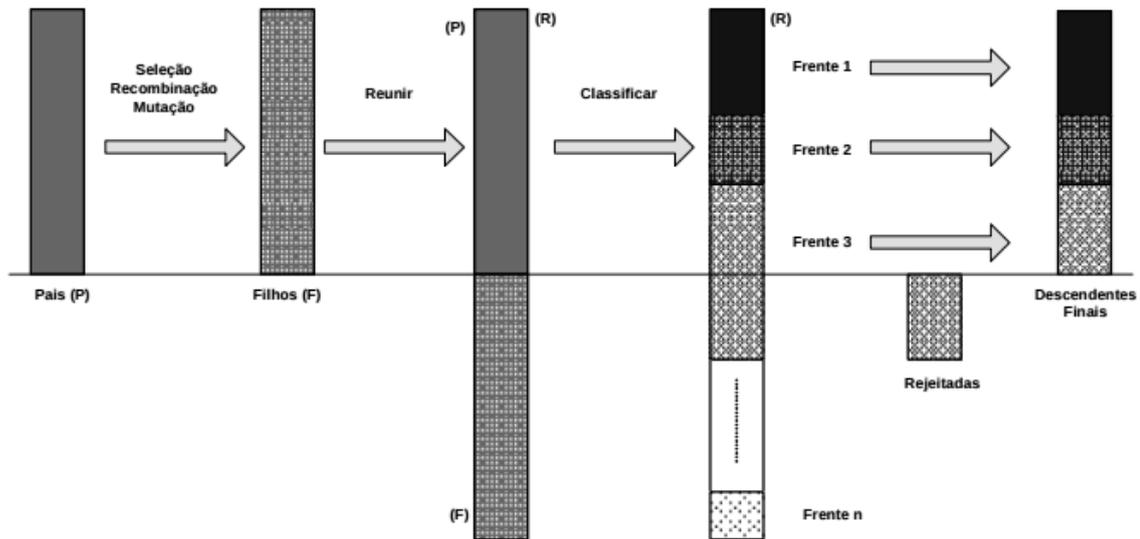


Figura 4.7: Diagrama esquemático do funcionamento do algoritmo NSGA-II. Adaptado de (ZINI, 2009), p.77.

4.5 Cruzamento

Trata-se de uma forma de recombinação de cromossomos, fazendo com que alelos que antes se apresentavam num mesmo cromossomo sejam separados. Tal mecanismo direciona o processo de busca a novos máximos locais que ainda não foram devidamente visitados. O processo não produz material genético novo (FERREIRA, 2007).

Tabela 4.2: Material genético dos pais selecionados para os exemplos de tipos de cruzamentos das Tabelas a seguir.

Pais	Cromossomos							
A	1	0	1	1	1	0	1	1
B	0	1	0	1	0	0	0	1

4.5.1 Cruzamento de um ponto

Trata-se de uma técnica onde uma posição aleatória é definida e a partir deste ponto o material genético dos cromossomos dos pais é trocado. A Tabela 4.3 ilustra o resultado da recombinação por esta metodologia e os pais apresentados na Tabela 4.2.

Tabela 4.3: Exemplo de cruzamento de um ponto considerando.

Filhos	Cromossomos							
A	1	0	1	1	0	0	0	1
B	0	1	0	1	1	0	1	1

4.5.2 Cruzamento de múltiplos pontos

No cruzamento de múltiplos pontos, uma quantidade K de pontos é selecionada aleatoriamente e o operador divide os cromossomos em " $k+1$ ". Estes números são considerados os intervalos intergenes que deverão ser recombinados. A Tabela 4.3 ilustra o resultado da recombinação por esta metodologia, considerando os pais apresentados na Tabela 4.2, em que $K=3$ e que as posições intergenes sorteadas sejam $k_1 = 1$, $k_2 = 4$ e $k_3 = 5$.

Tabela 4.4: Exemplo de cruzamento de múltiplos pontos.

Filhos	Cromossomos							
C	1	1	0	1	1	0	0	1
D	0	1	0	1	0	0	1	1

4.5.3 Cruzamento uniforme

No caso de cruzamentos uniformes, uma máscara é gerada de forma aleatória para cada par de cromossomos pais para definir quais genes serão recombinados. A Tabela 4.6 apresenta o resultado da recombinação por esta metodologia, considerando os pais apresentados na Tabela 4.2 e a máscara contida na Tabela 4.5.

Tabela 4.5: Máscara gerada aleatoriamente

Máscara	1	0	1	0	1	1	0	0
---------	---	---	---	---	---	---	---	---

4.6 Mutação

A mutação é a operação responsável por manter a diversidade de uma população durante o processo de reprodução, escolhendo aleatoriamente um gene e modificando seu alelo. Ao contrário do cruzamento, a mutação produz material genético novo, sendo um criador ou um destruidor de boa genética. Alguns autores estabelecem a mutação como um processo associado à reprodução. Desta forma pode ser considerada como uma cópia danificada de conteúdo genético durante a reorganização genética (FERREIRA, 2007).

Tabela 4.6: Exemplo de cruzamento uniforme.

Filhos	Cromossomos							
E	0	0	0	1	0	0	1	1
F	1	1	1	1	1	0	0	1

A sua função, do ponto de vista matemático, é garantir que o processo de busca da solução ótima não fique preso num mínimo local (ABDOUN; ABOUCHABAKA; TAJANI, 2012). A seguir são apresentadas as formas mais usuais de mutação constantes nos processos de algoritmo genético.

4.6.1 Mutação de um ponto

Na mutação de um ponto, um alelo do cromossomo é selecionado aleatoriamente e os seus valores são invertidos. A Tabela 4.7 ilustra esta metodologia.

Tabela 4.7: Exemplo de mutação de um ponto.

Cromossomo original	1	0	1	0	1	1	0	0
Cromossomo mutado	1	0	1	0	0	1	0	0

4.6.2 Mutação de múltiplos pontos

Na mutação de múltiplos pontos, uma quantidade K de alelos é selecionada para ocorrer a alteração dos valores. Em seguida são sorteados, de forma randômica, os endereços dos alelos que deverão ser invertidos. A Tabela 4.8 ilustra esta metodologia para uma quantidade $K = 3$ e $k_1 = 1$, $k_2 = 4$ e $k_3 = 6$.

Tabela 4.8: Exemplo de mutação de múltiplos pontos.

Cromossomo original	1	0	1	0	1	1	0	0
Cromossomo mutado	1	1	1	0	0	1	1	0

4.6.3 Mutação uniforme

Na mutação uniforme, uma máscara é gerada para cada cromossomo para definir quais alelos devem ter seus valores invertidos. A Tabela 4.9 ilustra esta metodologia para a máscara definida na Tabela 4.5.

Tabela 4.9: Exemplo de mutação uniforme.

Cromossomo original	1	0	1	0	1	1	0	0
Cromossomo mutado	0	1	0	0	0	0	1	0

4.7 Parametrização

Para o desempenho de um algoritmo genético é necessário selecionar os parâmetros de forma adequada para se evitar dificuldades de convergência na procura da solução ótima. Entre os problemas que podem ser encontrados estão a baixa velocidade de busca e a generalização excessiva da população, que podem aumentar em muito o tempo de processamento. O contrário destas características pode trazer um resultado sub-ótimo. A seguir serão apresentados alguns critérios para a escolha dos principais parâmetros para uma configuração adequada dos algoritmos genéticos.

4.7.1 Tamanho da população

Em (HAUPT, 2000) concluiu-se que uma pequena população inicial com aplicação de uma alta taxa de mutação pode funcionar melhor em muitas das aplicações. Para tanto foram consideradas populações variando entre 4 e 60 indivíduos e as taxas de mutação variando entre 0,01 e 0,49.

Contudo em (GOTSHALL; RYLANDER, 2002) concluiu-se que quanto maior a população inicial, maior a chance de um cromossomo representar a solução ótima. Foi observado que o aumento do tamanho da população inicial influencia no aumento do número de gerações para a convergência à solução ótima. O autor considerou que este aumento no número de gerações era o necessário para estabelecer a eliminação dos cromossomos mutados.

Em (ROEVA; FIDANOVA; PAPRZYCKI, 2013) foi apresentado um estudo da influência do tamanho da população inicial nos algoritmos genéticos. Trabalhou-se com um número fixo de gerações. O tamanho da população inicial variou entre 5 e 200 indivíduos.

Em tal trabalho foi encontrado que a população ótima foi de 100, considerando os resultados da função objetivo estabelecida e em relação ao tempo de processamento computacional. Além disso, foi observado que pequenas populações convergem para soluções com baixa precisão e que grandes populações apenas aumentam o custo computacional.

4.7.2 Taxas de cruzamento e mutação

Quanto maior a taxa de cruzamento, mais rapidamente novos atributos são introduzidos na população, mas pode-se perder indivíduos mais aptos. Caso essa taxa seja muito

pequena, a busca pode se tornar muito lenta. Uma baixa taxa de mutação previne que a busca fique estagnada em regiões do espaço de busca. Caso essa taxa seja muito alta, a busca se torna aleatória. Em (DEJONG, 1975) é sugerida uma taxa de mutação equivalente a 0,001 e uma taxa de cruzamento equivalente a 0,6.

4.7.3 Critério de parada

Determinar o processo de busca da solução ótima é uma ação difícil que geralmente exige o compromisso da satisfatoriedade da precisão alcançada com a responsabilidade do uso dos recursos computacionais. Neste sentido, existem vários critérios possíveis de serem utilizados, conforme apresentado em (ZIELINSKI; PETERS; LAUR, 2005).

4.7.3.1 Critérios baseados na exaustão

São aqueles critérios de parada baseados no estabelecimento de uma quantidade fixa de gerações, ou de número de avaliações da função objetivo ou de tempo de processamento computacional.

4.7.3.2 Critérios baseados no progresso

São aqueles critérios de parada baseados na melhoria de algum parâmetro. Caso haja uma diminuição substancial do progresso ao longo das gerações, o processo de busca da solução otimizada deve ser interrompido. Podem ser utilizados como parâmetros tanto o valor quanto o valor médio da função objetivo, além da quantidade de gerações em que não houve melhoria do desempenho, por exemplo.

4.7.3.3 Critérios baseados no movimento

São aqueles critérios de parada baseados no movimento dos indivíduos da população. Podem ser avaliados por meio da análise do movimento da população em relação ao valor médio da função objetivo ou em relação a um limiar para a quantidade específica de gerações. Também podem utilizar o desvio padrão de todas as distâncias calculadas ou ainda a diferença entre a distância do melhor para o pior indivíduo.

4.7.3.4 Critérios baseados na distribuição

São aqueles critérios de parada baseados na distribuição estatística da população. Entende-se, neste caso que a função objetiva é alcançada quando os indivíduos da

população estão próximos, ou seja, com baixa dispersão. As variantes desta metodologia usam a distância do melhor cromossomo a todos os outros dispostos na população ou a um grupo seletivo pré-determinado.

4.7.3.5 Critérios combinados

São aqueles critérios de parada que são baseados em algumas das diversas metodologias listadas anteriormente.

4.8 Considerações finais

Nesta seção foram introduzidos os princípios básicos dos algoritmos genéticos. Foram demonstradas suas características essenciais, suas vantagens, as principais operações genéticas bem como os elementos para a realização de sua parametrização.

5 Técnica acústica proposta para detecção de intoxicação alcoólica

Neste capítulo é apresentada a metodologia acústica proposta para a detecção da intoxicação alcoólica por meio das falas de pessoas embriagadas. A seção 5.1 apresenta as considerações iniciais. A seção 5.2 apresenta as técnicas necessárias para a extração das características que foram utilizadas neste trabalho. A seção 5.3 comenta como se dá a seleção das características por meio de algoritmo genético e sua parametrização. A seção 5.4 apresenta como se deu o processo de classificação. A seção 5.5 demonstra a avaliação de alguns resultados de trabalhos similares. A seção 5.6 apresenta a avaliação dos efeitos do gênero na classificação. A seção 5.7 demonstra uma tentativa de validação de resultados do AG por meio de uma técnica recursiva de eliminação de parâmetros. Por fim, na seção 5.8 são apresentadas as considerações finais.

5.1 Considerações iniciais

Nas seções anteriores, foram revisados alguns conceitos de fonética forense. Observou-se que o trato vocal realiza um papel preponderante, funcionando como tubo acústico ressonante conforme o posicionamento dos órgãos envolvidos na fala.

Foi observado que as vozes dos indivíduos submetidos à intoxicação alcoólica, na média, ocupam um espaço acústico das vogais diferente daqueles indivíduos que se encontram em estado de sobriedade. Comportamento similar é esperado para a produção das consoantes quando se comparam as falas de indivíduos alcoolizados e sóbrios. Como exposto anteriormente, há a tendência de palatização de sons fricativos quando o falante se encontra sob o efeito do álcool.

Considerando o contexto discriminado, este trabalho apresenta uma técnica para detecção da intoxicação alcoólica por meio da análise acústica das falas de indivíduos alcoolizados. Contudo, diferentemente das abordagens constantes nos métodos acústicos estabelecidos por estado da arte, tal metodologia consiste na análise única e exclusivamente de parâmetros acústicos de forma segmental, que são classificados por meio de Algoritmos Genéticos e Regressão Logística. O diagrama de blocos básico do modelo proposto é similar ao constante na Figura 5.1.

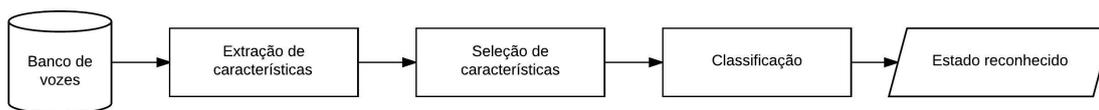


Figura 5.1: Diagrama de blocos básico do modelo proposto de sistemas de análise de voz.

5.2 Extração do vetor de características

Nesta seção será discutida a segunda etapa do diagrama de blocos disponibilizado na Figura 5.1, a extração de características das falas. A proposta deste trabalho é a obtenção desses parâmetros de forma segmental.

5.2.1 Janelamento

A primeira parametrização necessária para tanto é o estabelecimento da janela sob a qual serão calculadas as características acústicas das vozes, bem como a definição do tempo de deslocamento da mesma. Durante a revisão bibliográfica foram buscados estudos que proporcionassem o embasamento para a escolha, a fim de garantir que a gama de valores cepstrais calculados pudesse representar as falas dos indivíduos alcoolizados da melhor maneira possível.

A janela é uma função matemática que aplica zero a todo e qualquer outro valor fora do intervalo de interesse. Nesse sentido, a técnica de janelamento nada mais é que a realização da multiplicação da função randômica do sinal a se processar com essa função matemática. A Figura 5.2 ilustra alguns tipos de janela constantes na literatura científica.

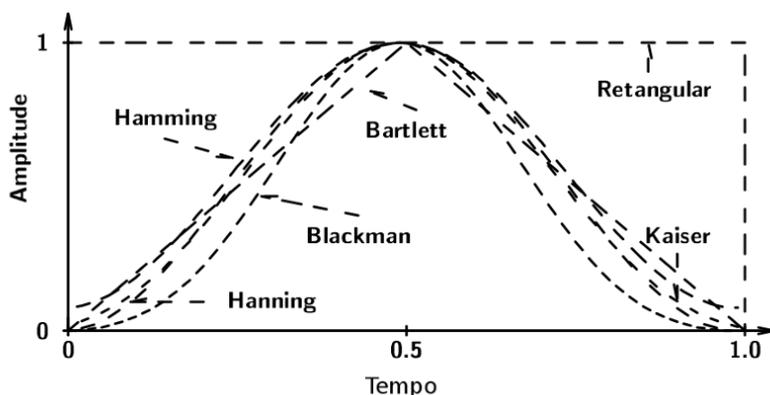


Figura 5.2: Imagem demonstrativa dos tipos de janela existentes na literatura científica. Adaptada a partir de (DENG; O'SHAUGHNESSY, 2003).

Inicialmente convém informar que as janelas quadrada e triangular não foram consideradas como opções viáveis porque seus lóbulos centrais são menores e os laterais maiores, quando comparados aos lóbulos gerados pelas técnicas de Hanning, Hamming e Blackman. Estas, por sua vez, tiveram seus desempenhos analisadas em (PODDER et al., 2014) quando aplicadas em filtros digitais. Foram avaliadas a largura de banda, a resposta de fase, a largura de transição de lóbulos, além da resposta em frequência. Demonstrou-se que a janela de Blackman foi superior às outras em todos os critérios. A Figura 5.3 ilustra a comparação das respostas em frequência dos janelamentos de Hamming, Hanning e Blackman.

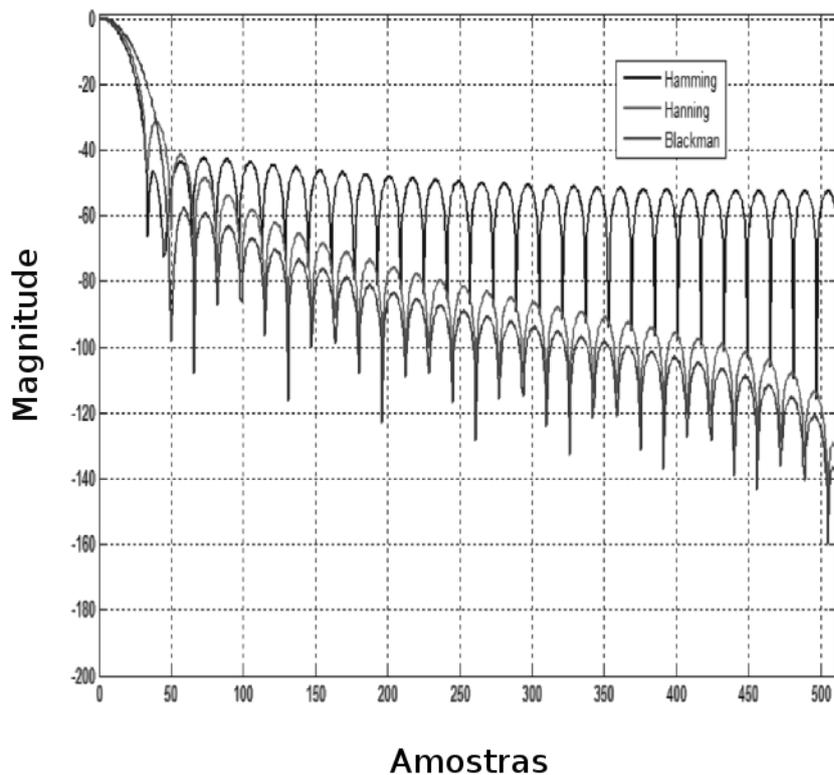


Figura 5.3: Comparação das respostas em frequência dos janelamentos de Hamming, Hanning e Blackman. Adaptada a partir de (PODDER et al., 2014).

Outra parametrização importante é o tamanho da janela. Em (LEE; CHO, 2016), o autor cita que intervalo de tempo mais utilizado para o janelamento é aquele de 25 ms, com deslocamento padrão de 10 ms. Já em (PALIWAL; LYONS; WÓJCICKI, 2010), discutiu-se tal tema e concluiu-se, baseando-se em dados objetivos de reconhecimento de consoantes, que uma janela entre 15 e 35 ms para análise de tempo é o ideal. Contudo, este estudo indicou que, em sistemas de reconhecimento automático de falantes, é possível a utilização de intervalos de 15 a 60 ms. Na Figura 5.4 são demonstrados espectrogramas de uma locução na língua inglesa e é possível observar o efeito de diversas larguras de janela.

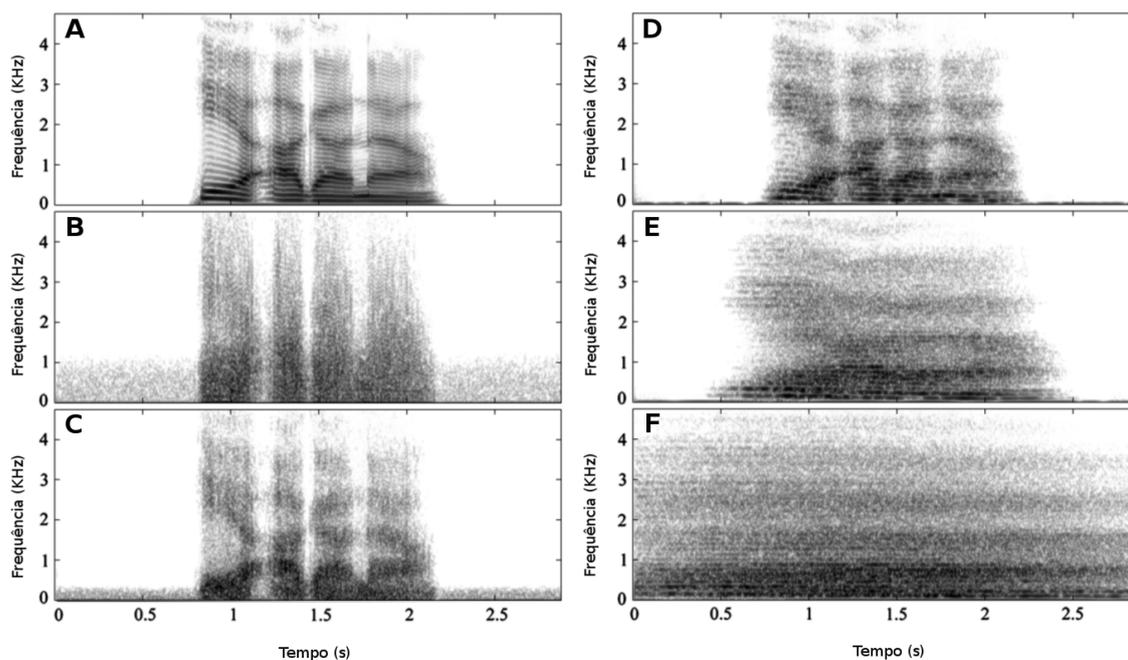


Figura 5.4: Espectrograma da fala da língua inglesa *"Hear aga now"* proferida por um indivíduo do sexo masculino. (A) Fala original ou sem perdas. (B) Fala processada com janela de 2 ms. (C) Fala processada com janela de 8 ms. (D) Fala processada com janela de 32 ms. (E) Fala processada com janela de 128 ms. (F) Fala processada com janela de 512 ms. Adaptada a partir de (PALIWAL; WÓJCICKI, 2008).

Já pelo exposto em (DENG; O'SHAUGHNESSY, 2003), é possível depreender que a escolha da janela é um compromisso com aquilo de necessitamos processar. A análise de sons vocálicos longos pode permitir janelas de até 100 ms com pouca perda de detalhamento. Contudo, sons plosivos necessitam de janelamento curto, em torno de 10 ms. A janela sugerida pelo autor está num intervalo entre 20 e 30 ms, com deslocamento em torno de 10 ms.

Corroborando o exposto nesta subseção, foi definido neste trabalho que seria utilizada a técnica de Blackman para o janelamento, devido a esta sofrer menos influência dos lóbulos laterais, bem como ser mais eficiente em relação ao seu central quando comparada com outras técnicas. Além disso, considerando o conteúdo exposto em (DENG; O'SHAUGHNESSY, 2003), bem como no Capítulo 2 desta dissertação, que demonstrou que as pessoas intoxicadas pelo álcool tendem a realizar a palatização do [s], o que se percebe sonoramente por meio da produção de [ʃ], há uma nítida necessidade de realizar a maior quantidade de amostras possíveis. Sendo assim, foi estabelecido que a janela utilizada neste trabalho seria fixada em 30 ms com o deslocamento temporal equivalente a 10 ms.

5.2.2 Características cepstrais

O GFCC é uma característica cepstral que foi definida pela primeira vez em (ZHANG; ABDULLA, 2005). O cálculo de seus valores ocorre por meio de um banco de filtros conhecido como *Gammatone*. A saída à excitação deste banco é conhecido como Co-cleograma. O objetivo é simular as características da membrana basilar da cóclea do ouvido humano.

Conforme apresentado em (QI et al., 2013), a equação 5.1 define a resposta em frequência do filtro *Gammatone*.

$$g(t) = at^{n-1}e^{2\pi bt} \cos(2\pi f_c t + \phi) \quad (5.1)$$

Onde, f_c é a frequência central do filtro. b é o fator de atenuação do filtro e está relacionado com a largura de banda, definido pela Equação 5.2. A Figura 5.5 demonstra um exemplo de resposta em frequência do filtro *Gamma*.

$$b = 25.17 \left(\frac{4.37 f_c}{1000} + 1 \right) \quad (5.2)$$

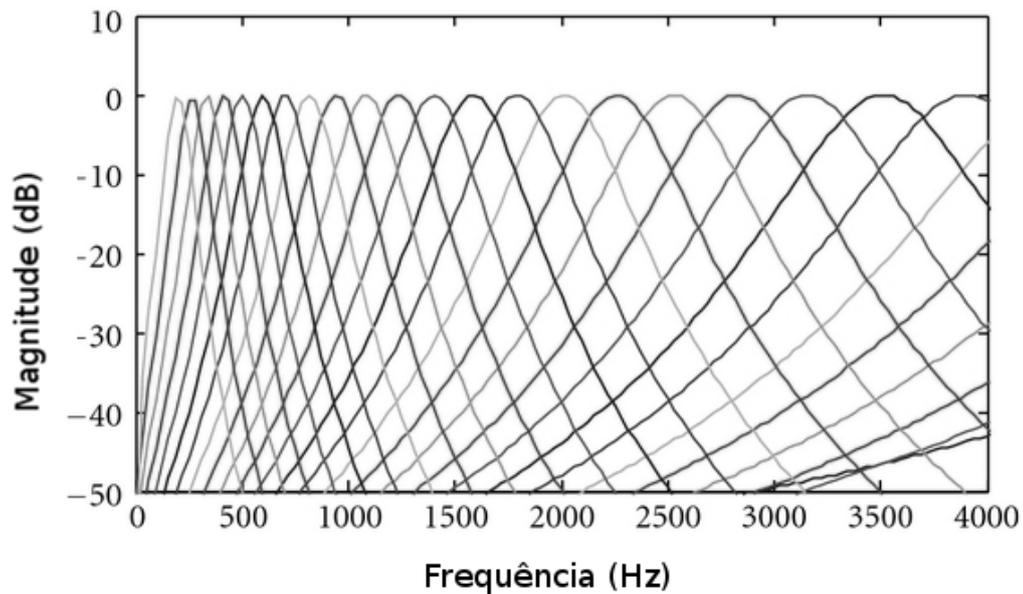


Figura 5.5: Exemplo de resposta em frequência de canais do filtro Gamma. Adaptada a partir de (ZOGHLAMI; LACHIRI, 2012).

A largura de banda retangular equivalente, ERB (*Equivalent Rectangular Bandwidth*), é definida como sendo uma medida psicoacústica da largura de banda de um filtro auditório em cada ponto da cóclea humana (MOINUDDIN; KANTHI, 2014). Ela pode ser representada matematicamente em função de b , conforme a Equação 5.3. A Figura 5.6 ilustra a relação entre a largura de banda de um filtro genérico e a ERB.

$$ERB(f_c) = \frac{b}{1,019} \quad (5.3)$$

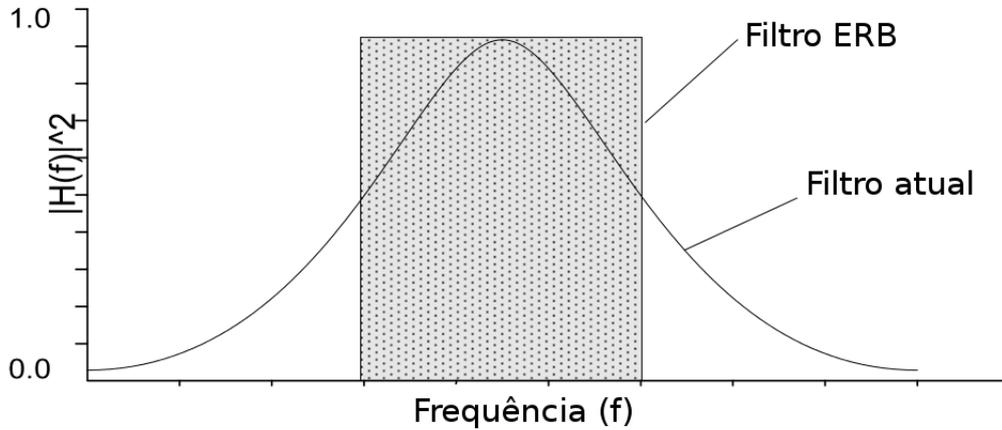


Figura 5.6: Relação entre a largura de banda de um filtro genérico e a ERB. Adaptada a partir de (ABDULLA, 2012).

Por fim, a Equação 5.4 define a função para a extração dos parâmetros GFCC, onde M é o número total de canais (QI et al., 2013).

$$gfcc(n; u) = \left(\frac{2}{M} \right)^{0.5} \sum_{i=0}^{M-1} \left\{ \frac{1}{3} \log(y(n; i)) \cos \left[\frac{\pi u}{2M} (2i - 1) \right] \right\} \quad (5.4)$$

Ao longo dos anos a comunidade científica tem demonstrado que o GFCC tem resultados similares ao MFCC, característica cepstral tida como o estado da arte e amplamente utilizada em diversos estudos da voz. A grande vantagem da primeira em relação à segunda é a sua maior robustez a ruídos. Tal fato foi demonstrado em (SHAO; WANG, 2008), (JIANG; LIU; ZU, 2014), (MOINUDDIN; KANTHI, 2014), entre outros estudos.

Em (ZHAO; WANG, 2013), foi investigado o porquê de o GFCC apresentar maior invariância aos ruídos que o MFCC. Foi concluído que esta superioridade deve se dar

devido à retificação implementada pela raiz cúbica. As diferenças básicas entre MFCC e GFCC podem ser visualizadas na Tabela 5.1.

Tabela 5.1: Comparação entre MFCC e GFCC.

Categoria	MFCC	GFCC
Pré-ênfase	Sim	Não
Nº. bandas frequenciais	26	64
Filtragem cepstral	Sim	Não
Retificação não -linear	Logaritmica	Cúbica
Invariante à escala	Sim	Não
Representação	Espectro Mel	Variante do cochleograma

5.2.3 Características estáticas, dinâmicas e estatísticas

A terminologia estabelece a diferenciação entre características estáticas, dinâmicas e estatísticas. A primeira é aquela extraída como efeito direto do janelamento. Por exemplo, todo e qualquer parâmetro extraído por meio de GFCC.

Uma característica dinâmica é aquela que tem o objetivo de capturar a velocidade e a aceleração das mudanças do sinal de voz. As Equações 5.5 e 5.6 apresentam, respetivamente as funções Delta e Delta-deltas.

$$\Delta F(n.u) = \frac{\sum_{\kappa=1}^K \kappa(F(n+k, u) - F(n-k, u))}{2 \sum_{\kappa=1}^K \kappa^2} \quad (5.5)$$

$$\Delta\Delta F(n.u) = \frac{\sum_{\kappa=1}^K \kappa(\Delta F(n+k, u) - \Delta F(n-k, u))}{2 \sum_{\kappa=1}^K \kappa^2} \quad (5.6)$$

Por sua vez, as características estatísticas são aquelas que derivam do estudo dos vetores de parâmetros por meio de técnicas de análise de dados estocásticos. Entre estes parâmetros estão os cálculos de média aritmética, média quadrática, desvio padrão, variância, assimetria e curtose.

5.2.4 Definição dos vetores de características

A geração das características é a primeira tarefa a ser realizada em processos de análise da voz. Estes dados gerados são armazenados vetorialmente considerando uma janela temporal previamente definida.

Considerando o exposto na seção 5.2.1, ficou estabelecido que os dados cepstrais seriam calculados por meio de um janelamento do tipo Blackman, com espaçamento e deslocamento temporais equivalentes a 30 ms e 10 ms, respectivamente. Em cada uma destas amostragens realizadas foi extraído um vetor contendo 15 coeficientes GFCC e a energia de 40 ERB.

Além disso, foram extraídos os coeficientes Delta e Delta-deltas de todos os parâmetros previamente adquiridos, utilizando para tanto o formalismo matemático apresentado pelas Equações 5.5 e 5.6. Em seguida, tais características dinâmicas foram adicionadas ao vetor original.

A quantidade de amostras para cada arquivo de áudio analisado foi, obviamente, dependente do seu comprimento total. Cada vetor, correspondente a cada uma destas amostras, foi empilhado sequencialmente formando uma matriz de comprimento $m \times n$, que são, respectivamente definidos pelas Equações 5.7 e 5.8, onde T_{arq} é o comprimento do arquivo em segundos, T_a é o intervalo de tempo em que ocorre cada amostra (em segundos), N_{gfcc} é a quantidade de coeficientes GFCC, N_{ERB} é a quantidade coeficientes de ERBs e N_{din} é a quantidade de variáveis dinâmicas calculadas.

$$m = \frac{T_{arq}}{T_a} \quad (5.7)$$

$$n = (N_{gfcc} + N_{ERB})(1 + N_{din}) \quad (5.8)$$

Como exemplo, supondo um arquivo de comprimento equivalente a 10 segundos e considerando $N_{gfcc} = 15$, $N_{ERB} = 40$ e $N_{din} = 2$, geraríamos uma matriz com 1000 linhas e 165 colunas.

Na sequência, foram calculados os seguintes parâmetros estatísticos: média aritmética, variância, assimetria e curtose. Sendo assim, cada coluna da matriz gerada passou pela

avaliação destas funções, e foram listadas em sequência, formando um vetor único por arquivo. A Equação 5.9 define o comprimento do vetor, onde N_{st} é a quantidade de parâmetros estatísticos a serem calculados e n foi definido pela Equação 5.8.

$$l = nN_{st} \tag{5.9}$$

Considerando então os dados do caso dessa dissertação, onde $n = 165$ e $N_{st} = 4$, cada arquivo de áudio do *corpus* de vozes foi discriminado por meio de um vetor de comprimento l equivalente a 660 parâmetros.

5.3 Seleção das características por meio de Algoritmo Genético

Para a configuração do algoritmo genético foram utilizados os preceitos estabelecidos no Capítulo 4. O objetivo da aplicação do algoritmo genético para o processo de classificação das falas das pessoas alcoolizadas foram:

- Garantir a maior taxa de classificação;
- Garantir a menor quantidade de características.

A proposta de algoritmo estabelecida neste texto é similar àquela apresentada em (PUNCH et al., 1993), onde seu autor utilizou vetores binários para servir de multiplicador para selecionar dimensões em um classificador KNN, em caso de valor 1, ou não, em caso de zero.

Para a realização do processo de seleção de características, e a consequente avaliação do modelo proposto, foi aleatoriamente escolhido um texto que citava, no idioma alemão, “+491763582901”. Todos os voluntários que submeteram suas vozes ao *corpus* leram, vez a vez, esse discurso em voz alta. Os arquivos de áudio tinham, tipicamente, cerca de 10 segundos.

Cada um dos arquivos de áudio foi processado conforme especificado na seção 5.2 e seus vetores característicos foram devidamente agrupados matricialmente. Na sequência, passaram por um pré-processamento com o objetivo de re-escalonar a matriz.

Foi gerado um vetor \mathbf{V} , de módulo unitário, com comprimento equivalente a cada um dos vetores característicos dos arquivos de áudio do *corpus*. O AG irá trabalhar sobre este vetor \mathbf{V} , que será responsável por selecionar as características por meio de acionamento de uma característica, se seu valor é 1, ou dispensando-a se 0. Sendo assim, do ponto de vista do algoritmo genético, o vetor \mathbf{V} que levará aos objetivos da aplicação é o cromossomo que levará ao valor ótimo em função de avaliação, qual seja, a máxima taxa de classificação e com o acionamento do menor número de características.

Como se depreende dos objetivos citados no início dessa subseção, existem duas funções de avaliação, uma necessitando ser minimizada e outra maximizada. Esta é uma característica de processos multiobjetivos. A primeira delas é uma função que é dependente da quantidade de índices acionados do vetor \mathbf{V} . A segunda é o retorno do processo de classificação das falas do *corpus* de vozes realizado com Regressão Logística, considerando como saída a assertividade da classificação, calculado por meio de *Leave-One-Out Cross-Validation* (LOOCV).

Sendo assim, o algoritmo genético deverá ser parametrizado de tal forma que possibilite alcançar, ao final do processo, a solução otimizada que satisfaça as duas condições.

5.3.1 Parametrização do algoritmo genético

No Capítulo 4 foram discutidos os princípios de algoritmos genéticos. Mais especificamente, na seção 4.7, foram vistos suas variáveis de parametrização. A seguir, são apresentadas as variáveis utilizadas para a realização da seleção das características para a execução da tarefa de classificação das falas de indivíduos intoxicados pelo álcool.

5.3.1.1 Definição da população inicial

Tomando-se por base o que foi demonstrado em (ROEVA; FIDANOVA; PAPRZYCKI, 2013), e já indicado na subseção 4.7.1, definiu-se que a população inicial seria composta por 100 cromossomos. Tal conteúdo genético foi iniciado com todos os cromossomos sendo compostos por alelos de valor zero.

5.3.1.2 Definição do critério de seleção

O critério escolhido para a realização da tarefa de seleção foi o NSGA-II, devido ao seu bom comportamento em processos multiobjetivos.

5.3.1.3 Definição da taxa de cruzamento

Para a configuração da taxa de cruzamento, foi considerado o exposto em (OLIVEIRA et al., 2002), que optou por uma taxa de 0,8 para tarefa similar de seleção de características por meio de algoritmo genético. Isso significa que 80% dos indivíduos tem chance de realizar um cruzamento.

5.3.1.4 Definição da taxa de mutação

Optou-se por valores para taxa de mutação equivalente a 0,05, o que significa que 0,5% dos indivíduos tem a chance de sofrer mutação.

5.3.1.5 Critério de parada

O critério estabelecido para a parada foi do tipo combinado, baseado na exaustão, no progresso e no movimento. Em relação à exaustão foi inserida condição de parada quando ocorrerem 150 gerações. Em relação ao progresso, o processo deve parar se houver diferenças menores que 0,00010 entre cada iteração. Em relação ao movimento, a rotina será interrompida caso 1/3 das 15 últimas iterações tenham soluções iguais ou piores que os melhores resultados anteriores.

5.4 O processo de classificação

A rotina para a realização do teste do processo de classificação foi posta em execução e seu resultado pode ser verificado através dos gráficos contidos nas Figuras 5.7 e 5.8. O algoritmo rodou por 103 gerações, tendo parado devido ao critério de movimento estabelecido. Ao final, atingiu a marca de 81,18% na taxa de classificação das falas constantes no *corpus* com 65 características acústicas.

5.5 Comparação com trabalhos similares

A Tabela 5.2 apresenta o resultado consolidado das classificações das falas alcoolizadas que foram publicados na literatura científica, que utilizaram o mesmo *corpus* de falantes e características acústicas para realizar a segregação das falas intoxicadas pelo álcool.

Em (MONTACIÉ; CARATY, 2011) foi apresentado um classificador baseado em fonemas. Utilizaram-se para tanto características acústicas e alcançou-se a taxa de 64,5% na classificação das falas alcoolizadas. Este valor foi alcançado utilizando-se 20% do banco de dados para treinamento e teste.

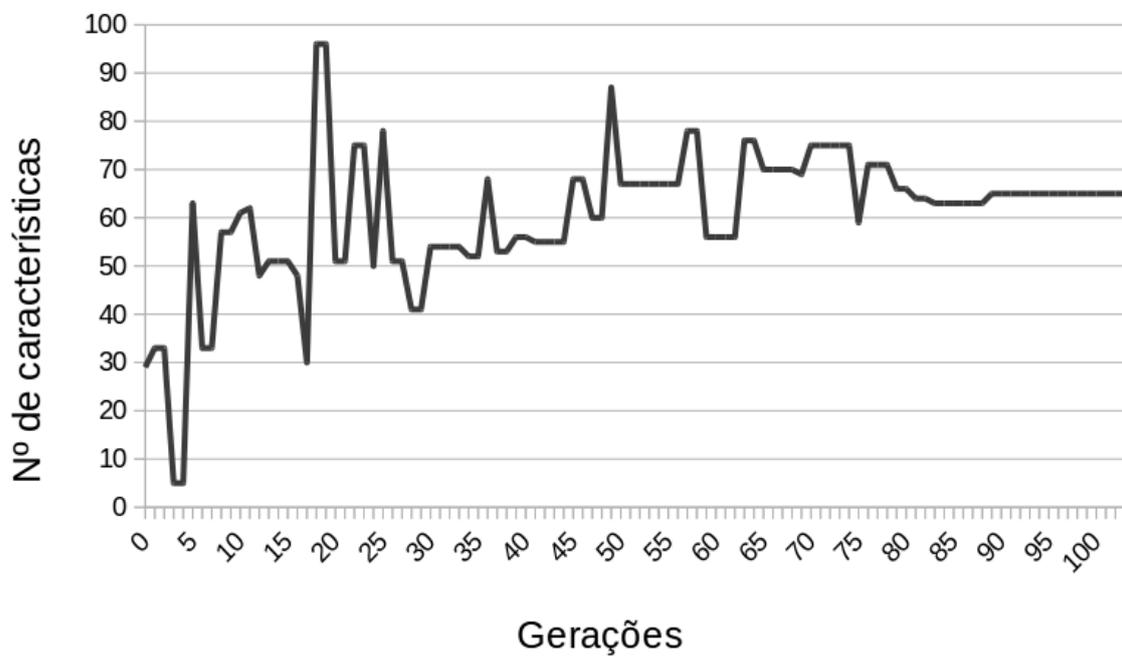


Figura 5.7: Ilustração da evolução do processo de redução de características do classificador implementado.

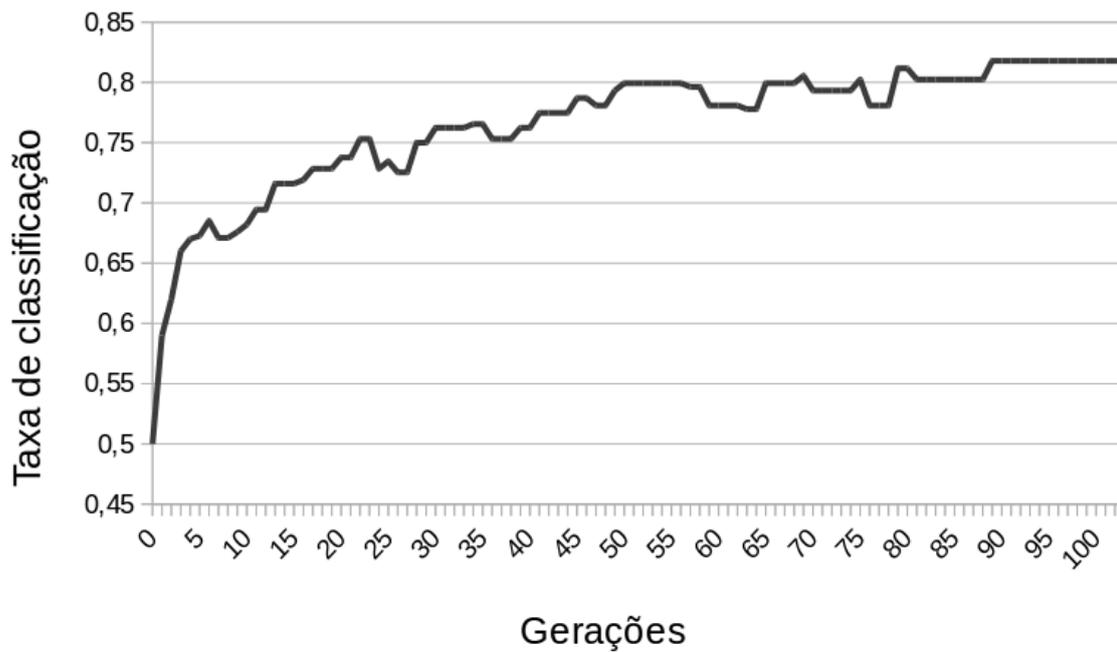


Figura 5.8: Ilustração da evolução da performance do classificador implementado.

Em (GAJŠEK; DOBRIŠEK; MIHELIČ, 2011), foram utilizadas algumas adaptações de HMM (*Hidden Markov Model*) para realizar a classificação da voz, após a extração dos parâmetros acústicos. Em seguida esses processos passaram por um classificador baseado em voto majoritário. Alcançou-se a taxa de 67,47% na classificação dessas falas.

Em (SCHULLER et al., 2011) foram utilizadas características acústicas cepstrais, bem como características estatísticas. Tais dados passaram por um classificador SVM, que retonou uma taxa de classificação equivalente a 65,9%.

Em (BONE et al., 2011) foram treinadas 512 características GMM (*Gaussian Mixture Model*) em cima de 39 características MFCC. Em seguida passou-se por um classificador SVM alcançando-se uma taxa de classificação equivalente a 70,5%.

Tabela 5.2: Classificadores propostos em outros trabalhos similares

#	Trabalho	
1	(MONTACIÉ; CARATY, 2011)	
2	(GAJŠEK; DOBRIŠEK; MIHELIČ, 2011)	
3	(SCHULLER et al., 2011)	
4	(BONE et al., 2011)	
#	Atributos	Resultados
1	MFCC, LPCC, <i>jitter</i> , <i>shirmer</i> e outros.	64,5%
2	MFCC, energia, coeficientes delta.	67,5%
3	MFCC, energia, coeficientes delta e outros.	65,9%
4	MFCC, <i>jitter</i> , <i>shirmer</i> e outros.	70,5%

5.6 Avaliação dos efeitos do gênero na classificação

Com o objetivo de constatar o comportamento do gênero nas taxas de classificação, o *corpus* foi dividido em dois, conforme o critério sexo do falante. O estudo foi conduzido utilizando as características consideradas ótimas pelo algoritmo genético gerado anteriormente. A tabela 5.3 apresenta os resultados obtidos. Os valores foram obtidos considerando validação cruzada LOOCV.

Demonstrou-se que as falas das mulheres são mais sensíveis ao álcool. Tal fato já era esperado e já tinha sido reportado em (SCHIEL, 2011). Segundo o estudo apresentado em (MUMENTHALER et al., 1999), a maior sensibilidade das mulheres ao álcool é

devido à metabolização mais lenta nas mulheres, bem como da menor quantidade de água presente em seus corpos.

Tabela 5.3: Avaliação do efeito do gênero na classificação.

Gênero	Performance do classificador
Mulheres	79,87%
Homens	71,18%

5.7 Validação dos dados

Para validar o desempenho do algoritmo genético, realizou-se o estudo das características por meio de RFE (*Recursive Feature Elimination*). Tal técnica é descrita em (PENG et al., 2009) e corresponde a um procedimento iterativo que passa por uma fase de treinamento do classificador, pelo cálculo do *ranking* de todas as características, para depois remover aqueles que, de acordo com a classificação neste *ranking*, sejam menos adequados.

Foi realizada uma rotina computacional que buscou as K características mais adequadas para a realização da classificação. A busca foi realizada variando o valor de K entre 5 e 200. O resultado pode ser visto na Figura 5.9.



Figura 5.9: Ilustração dos resultados de busca das melhores características por RFE.

Observou-se que a seleção das características por meio de RFE levou a uma taxa de classificação equivalente a 85,49%, utilizando-se de 98 parâmetros. Tal técnica mostrou-se mais eficiente para a redução dessas características que aquela implementada em

algoritmo genético. Entretanto, tal técnica é um algoritmo iterativo de busca das melhores características a serem aplicadas, que geralmente é utilizado em casos em que existem poucos parâmetros a serem pesquisados (CHEN; JEONG, 2007). Isso leva a crer que o resultado obtido aqui pode não se replicar em casos com *corpus* de vozes mais extensos e complexos.

Contudo, a técnica apresentada nesta dissertação se mostrou eficiente, pois alcançou uma taxa de classificação alta, quando comparada aos estudos anteriores, e conseguiu reduzir sensivelmente a quantidade de características selecionadas. Não foram encontrados estudos que obtiveram performance similar a este utilizando a mesma base de dados e que tivessem utilizado somente parâmetros cepstrais.

5.8 Considerações Finais

Ao longo deste capítulo foi demonstrada a metodologia acústica proposta para a detecção da alcoolização por meio de algoritmo genético. Foram detalhados cada um dos blocos básicos que compõe a proposta, desde o bloco de extração de características até o bloco que estabelece a classificação por meio de algoritmo genético. A taxa de classificação atingiu a marca de 81,11% e a redução de características deu-se de 660 para 65, alcançando o objetivo inicial do AG proposto: alcançar a maior taxa de classificação com o menor quantidade de características selecionadas nos vetores de entrada.

Contudo, com o objetivo de avaliar a aplicação do algoritmo genético, foi realizado um teste com RFE que encontrou uma taxa de classificação melhor que a aplicação realizada com AG. Contudo, o RFE é uma técnica de busca que tem sido utilizada nos meios científicos para selecionar características mais aptas em aplicações com poucos parâmetros. Sendo assim, talvez o resultado ocorrido neste estudo não se repita para aplicações com dimensionalidade mais alta.

6 CONCLUSÕES

O álcool é uma droga depressora que atua no sistema nervoso central e provoca uma série de alterações comportamentais e físicas nas pessoas intoxicados por ele. A soma desses efeitos pode influenciar sobremaneira na forma que o discurso de uma pessoa embriagada é percebido por um ouvinte. Essa sensação acústica vai depender tanto da sensibilidade desse ouvinte, como do grau de embriaguez em que se encontra o locutor, contudo, as mudanças causadas pela alcoolemia são percebidas na voz por meio de técnicas de análise espectral.

Apoiada sobre o exposto acima, esta dissertação apresenta a implementação de um classificador de vozes que tem o objetivo de segregar as falas de pessoas alcoolizadas daquelas falas de pessoas que se encontram no estado de sobriedade. Tal modelo de análise se baseia na análise cepstral das falas e sua subsequente seleção por meio de algoritmos genéticos, que foi implementado com dupla função: garantir a máxima taxa de classificação e reduzir o número de características ao mínimo possível para atingir a máxima taxa de classificação.

O algoritmo genético foi setado com um critério de parada combinado pelos critérios de exaustão, progresso e movimento. As iterações foram interrompidas na 103^a geração devido a este último critério.

O processo de classificação implementado neste trabalho conseguiu reduzir em mais de 90% a quantidade de características analisadas. Além disso, foi alcançada a marca de 81,18%, nas taxas de classificação considerando LOOCV. Assim, a hipótese inicial de que a seleção de características por meio de algoritmos genéticos iria reduzir de forma substancial as características não relevantes se tornou demonstrada.

Para avaliar a aplicação do AG foi realizado um teste com RFE. Tal teste performou melhor que o algoritmo genético e alcançou uma classificação equivalente a 85,49% com 98 parâmetros. Contudo, acredita-se que como esse algoritmo de busca teve uma performance melhor que o algoritmo genético devido à base de dados ser pequena.

Apesar do exposto acima, não foram encontrados na comunidade científica quaisquer estudos que tivessem obtido uma performance similar aos obtidos durante esta pesquisa com o mesmo *corpus* e que tenham utilizado somente parâmetros acústicos. Além disso, entre os parâmetros acústicos que já foram utilizados para estudos de intoxicação alcoólica por meio das falas de pessoas alcoolizadas, não há qualquer menção à utilização de GFCC. Tais taxas de classificação podem estar sendo alcançadas mais devido à técnica cepstral utilizada, que propriamente pela metodologia de seleção por algoritmo genético.

6.1 Recomendações para pesquisas futuras

Entre as possibilidades de trabalhos futuros para suceder os estudos realizados nessa dissertação sugere-se a realização de um banco de vozes para o português do Brasil, construído com representantes voluntários de todas as regiões do país. O ideal seria que tal *corpus* contivesse gravações dos voluntários em diversos níveis de alcoolização, desde a sobriedade.

Considerando a existência de um *corpus* conforme o sugerido acima, poderia-se considerar a viabilidade da realização de um estudo vislumbrando alcançar como resultado final um etilômetro vocal.

Além disso, há a necessidade de se realizar mais estudos a respeito da utilização do parâmetro cepstral GFCC e sua aplicabilidade em sistemas de detecção de alcoolização por meio da voz.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDOUN, O.; ABOUCHABAKA, J.; TAJANI, C. Analyzing the performance of mutation operators to solve the travelling salesman problem. *arXiv preprint arXiv:1203.3099*, Cornell University Library, 2012. Disponível em: <<https://arxiv.org/pdf/1203.3099.pdf>>. Acesso em: 12/11/2016.

ABDULLA, W. H. *Auditory based feature vectors for speech recognition systems*. 2012. Disponível em: <<http://cs1t.riit.tsinghua.edu.cn/mediawiki/images/3/30/AuditoryBasedFeatureVectors.pdf>>. Acesso em: 06/12/2016.

ANAGNOSTOPOULOS, C.-N.; ILIOU, T.; GIANNOUKOS, I. Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. *Artificial Intelligence Review*, Springer, v. 43, n. 2, p. 155–177, 2012.

ARAÚJO, C. *O princípio do nemo tenetur se detegere no crime de embriaguez ao volante: peça de resistência no vale tudo probatório*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio Grande do Sul, 2015.

BARBOSA, P. A.; MADUREIRA, S. *Manual de Fonética acústica experimental: Aplicações a dados do português*. [S.l.]: Cortez Editora, 2015. 591 p.

BARFÜßER, S.; SCHIEL, F. Disfluencies in alcoholized speech. Citeseer, 2010.

BAUMEISTER, B.; SCHIEL, F. On the effect of alcoholization on fundamental frequency. Department of Phonetics, University of Trier, 2010.

BONE, D. et al. Intoxicated speech detection by fusion of speaker normalized hierarchical features and gmm supervectors. In: *INTERSPEECH*. [S.l.: s.n.], 2011. p. 3217–3220.

BRANDÃO, A. *Modelagem Acústica da Produção da voz Utilizando Técnicas de Visualização de Imagens Médicas Associadas a Métodos Numéricos*. Tese (Doutorado) — Universidade Federal Fluminense-Programa de Pós-Graduação em Engenharia Mecânica (PGMEC), Niterói, RJ, Brazil, 2011.

BRASIL. Decreto-lei n.º 5.452, de 1º de maio de 1943. aprova a consolidação das leis do trabalho. *Diário Oficial da República do Brasil*, Brasília, DF, 09 ago. 1943. Disponível em: <https://www.planalto.gov.br/ccivil_03/Decreto-Lei/De15452.htm>. Acesso em: 10/12/2016.

BRASIL. Lei nº 9.503, de 23 de setembro de 1997. institui o código de trânsito brasileiro. *Diário Oficial da República do Brasil*, Brasília, DF, 24 set. 1997. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/L9503.htm>. Acesso em: 10/12/2016.

BRASIL. Decreto nº 6.488, de 19 de junho de 2008. regulamenta os arts. 276 e 306 da lei nº 9.503, de 23 de setembro de 1997 - código de trânsito brasileiro, disciplinando a margem de tolerância de álcool no sangue e a equivalência entre os distintos testes de alcoolemia para efeitos de crime de trânsito. *Diário Oficial da República do Brasil*, Brasília, DF, 20 jun. 2008. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6488.htm>. Acesso em: 10/12/2016.

BRASIL. Lei nº 11.705, de 19 de junho de 2008. altera a lei no 9.503, de 23 de setembro de 1997, que ‘institui o código de trânsito brasileiro’, e a lei no 9.294, de 15 de julho de 1996, que dispõe sobre as restrições ao uso e à propaganda de produtos fumíferos, bebidas alcoólicas, medicamentos, terapias e defensivos agrícolas, nos termos do § 4o do art. 220 da constituição federal, para inibir o consumo de bebida alcoólica por condutor de veículo automotor, e dá outras providências. *Diário Oficial da República do Brasil*, Brasília, DF, 20 jun. 2008. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/lei/111705.htm>. Acesso em: 10/12/2016.

BRASIL. Lei nº 12.760, de 20 de dezembro de 2012. altera a lei no 9.503, de 23 de setembro de 1997, que institui o código de trânsito brasileiro. *Diário Oficial da República do Brasil*, Brasília, DF, 21 dez. 2012. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112760.htm>. Acesso em: 10/12/2016.

BRASIL. Lei nº 13.146, de 6 de julho de 2015. institui a lei brasileira de inclusão da pessoa com deficiência (estatuto da pessoa com deficiência). *Diário Oficial da República do Brasil*, Brasília, DF, 07 jun. 2015. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/Lei/L13146.htm>. Acesso em: 10/12/2016.

BRASIL. Lei nº 13.281, de 4 de maio de 2016. altera a lei nº 9.503, de 23 de setembro de 1997 (código de trânsito brasileiro), e a lei nº 13.146, de 6 de julho de 2015. *Diário Oficial da República do Brasil*, Brasília, DF, 05 mai. 2016. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/lei/113281.htm>. Acesso em: 10/12/2016.

BRASIL. Câmara dos Deputados. Projeto de lei da câmara nº 144, de 2015. altera dispositivos da lei nº 9.503, de 23 de setembro de 1997 - código de trânsito brasileiro, para dispor sobre crimes cometidos na direção de veículos automotores. Brasília, DF,

2015. Disponível em: <<https://www25.senado.leg.br/web/atividade/materias/-/materia/123299>>. Acesso em: 10/12/2016.

BRASIL. Superior Tribunal de Justiça. Recurso especial nº 1.111.566 - df (2009/0025086-2). terceira turma. relator: Ministro marco aurélio bellizze. r.p/acórdão: Ministro adilson vieira macabu (desembargador convocado do tj/rj)(28/03/2012). processual penal. provas. averiguação do índice de alcoolemia em condutores de veículos. vedação à autoincriminação. determinação de elemento objetivo do tipo penal. exame pericial. prova que só pode ser realizada por meios técnicos adequados. decreto regulamentador que prevê expressamente a metodologia de apuração do índice de concentração de álcool no sangue. princípio da legalidade. Brasília, DF, 28 de mar. 2012. Disponível em: <<http://www.stj.jus.br/>>. Acesso em: 10/12/2016.

CHEN, X.-w.; JEONG, J. C. Enhanced recursive feature elimination. In: IEEE. *Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on*. [S.l.], 2007. p. 429–435.

CHIN, S. B.; LARGE, N. R.; PISONI, D. B. Effects of alcohol on the production of words in context: a first report. *Indiana University, Research on Spoken Language Processing, Progress Report*, v. 1997, n. 21, p. 403–20, 1996.

CROVATO, C. D. P. *Classificação de sinais de voz utilizando a transformada wavelet packet e redes neurais artificiais*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2004.

DEB, K. et al. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, IEEE, v. 6, n. 2, p. 182–197, 2002.

DEJONG, K. *An analysis of the behavior of a class of genetic adaptive systems*. Tese (Doutorado) — University of Michigan, 1975.

DENG, L.; O'SHAUGHNESSY, D. *Speech processing: a dynamic and optimization-oriented approach*. [S.l.]: CRC Press, 2003.

ECKSCHMIDT, F. *Efeitos do consumo de álcool combinado a bebidas energéticas (AmED) entre motoristas de caminhão no Estado de São Paulo: uma combinação de risco*. Dissertação (Mestrado) — Universidade de São Paulo, 2015.

FARRÚS, M.; HERNANDO, J.; EJARQUE, P. Jitter and shimmer measurements for speaker recognition. In: *Interspeech*. [S.l.: s.n.], 2007. p. 778–781.

FERREIRA, A. B. *Avaliação de operadores de algoritmos genéticos em otimização multidimensional*. Tese (Doutorado) — Universidade Estadual Paulista, 2007.

- GAJŠEK, R.; DOBRIŠEK, S.; MIHELIC, F. University of ljubljana system for interspeech 2011 speaker state challenge. In: *Twelfth Annual Conference of the International Speech Communication Association*. [S.l.: s.n.], 2011.
- GODINEZ, A. C.; ESPINOSA, L. E. M.; MONTES, E. M. An experimental comparison of multiobjective algorithms: Nsga-ii and omopso. In: IEEE. *Electronics, Robotics and Automotive Mechanics Conference (CERMA), 2010*. [S.l.], 2010. p. 28–33.
- GOLBERG, D. E. Genetic algorithms in search, optimization, and machine learning. *Addion wesley*, v. 1989, p. 102, 1989.
- GOTSHALL, S.; RYLANDER, B. Optimal population size and the genetic algorithm. *Population*, Citeseer, v. 100, n. 400, p. 900, 2002.
- HAUPT, R. L. Optimum population size and mutation rate for a simple real genetic algorithm that optimizes array factors. In: IEEE. *Antennas and Propagation Society International Symposium, 2000. IEEE*. [S.l.], 2000. v. 2, p. 1034–1037.
- HEIGL, B. [s] under the influence of alcohol. *Working Papers in Linguistics*, v. 8, Fall 2011. Disponível em: <http://www.gmu.edu/org/lingclub/WP/texts/8_Heigl2.pdf>. Acesso em: 07/10/2016.
- HINÇAL, O. *Optimization of multireservoir systems by genetic algorithm*. Tese (Doutorado) — Middle East Technical University, 2008.
- HINGSON, R.; WINTER, M. Epidemiology and consequences of drinking and driving. *Alcohol Research and Health*, US National Institute on on Alcohol Abuse and Alcoholism., v. 27, n. 1, p. 63–78, 2003.
- HOLLAND, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. [S.l.]: U Michigan Press, 1975.
- HOLLIEN, H. et al. Effects of ethanol intoxication on speech suprasegmentals. *The Journal of the Acoustical Society of America*, Acoustical Society of America, v. 110, n. 6, p. 3198–3206, 2001.
- HUANG, Y.-F.; WANG, W.-S. First language recognition of spoken english using audio features. researchgate, 2016. Disponível em: <https://www.researchgate.net/publication/303074844_First_Language_Recognition_of_Spoken_English_Using_Audio_Features>. Acesso em: 27/11/2016.

- JIANG, Y.; LIU, R.; ZU, Y. Auditory feature for monaural speech segregation. In: *2nd International conference on information, electronics and computer (ICIEAC)*. [S.l.: s.n.], 2014.
- JOHNSON, K.; PISONI, D. B.; BERNACKI, R. H. Do voice recordings reveal whether a person is intoxicated? a case study. *Phonetica*, Karger Publishers, v. 47, n. 3-4, p. 215–237, 1990.
- KAZIMIPOUR, B.; LI, X.; QIN, A. A review of population initialization techniques for evolutionary algorithms. In: IEEE. *2014 IEEE Congress on Evolutionary Computation (CEC)*. [S.l.], 2014. p. 2585–2592.
- KIKUCHI, L. *Vogais Altas e Glides no Português Brasileiro e no Inglês Britânico*. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, 2001.
- KUMAR, Y.; DAS, B.; SHARMA, J. Multiobjective, multiconstraint service restoration of electric power distribution system with priority customers. *IEEE Transactions on Power delivery*, IEEE, v. 23, n. 1, p. 261–270, 2008.
- LEE, B.; CHO, K.-H. Brain-inspired speech segmentation for automatic speech recognition using the speech envelope as a temporal reference. *Scientific Reports*, Nature Publishing Group, v. 6, 2016.
- LEVIT, M. et al. Use of prosodic speech characteristics for automated detection of alcohol intoxication. In: *ISCA Tutorial and Research Workshop (ITRW) on Prosody in Speech Recognition and Understanding*. [S.l.: s.n.], 2001.
- MCLOUGHLIN, I. *Applied speech and audio processing: with Matlab examples*. [S.l.]: Cambridge University Press, 2009. 2459 p. ISBN 978-0-521-51954-0.
- MITCHELL, M. *An introduction to genetic algorithms*. [S.l.]: MIT press, 1998.
- MODELLI, M. E. d. S.; PRATESI, R.; TAUIL, P. L. Alcolemia em vítimas fatais de acidentes de trânsito no distrito federal, brasil. Faculdade de Saúde Pública da Universidade de São Paulo, 2008.
- MOINUDDIN, M.; KANTHI, A. N. Speaker identification based on gfcc using gmm. *International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN*, p. 2349–2163, 2014.
- MONTACIÉ, C.; CARATY, M.-J. Combining multiple phoneme-based classifiers with audio feature-based classifier for the detection of alcohol intoxication. In: *Twelfth An-*

nual Conference of the International Speech Communication Association. [S.l.: s.n.], 2011.

MUMENTHALER, M. S. et al. Gender differences in moderate drinking effects. *Alcohol research and health*, v. 23, n. 1, p. 55–64, 1999.

NEY, H. Digital processing of speech and image signals. *Skript zur Vorlesung. RWTH Aachen, Sommersemester*, 2003. Disponível em: <http://www-i6.informatik.rwth-aachen.de/web/Teaching/Lectures/WS06_07/Signal/skript.pdf>. Acesso em: 21/11/2016.

NORAINI, M. R.; GERAGHTY, J. *Genetic algorithm performance with different selection strategies in solving TSP*. 2011.

OLIVEIRA, L. S. et al. Feature selection using multi-objective genetic algorithms for handwritten digit recognition. In: IEEE. *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. [S.l.], 2002. v. 1, p. 568–571.

PALIWAL, K.; WÓJCICKI, K. Effect of analysis window duration on speech intelligibility. *IEEE signal processing letters*, IEEE, v. 15, p. 785–788, 2008.

PALIWAL, K. K.; LYONS, J. G.; WÓJCICKI, K. K. Preference for 20-40 ms window duration in speech analysis. In: IEEE. *Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference on*. [S.l.], 2010. p. 1–4.

PENG, S. et al. A new implementation of recursive feature elimination algorithm for gene selection from microarray data. In: IEEE. *Computer Science and Information Engineering, 2009 WRI World Congress on*. [S.l.], 2009. v. 3, p. 665–669.

PODDER, P. et al. Comparative performance analysis of hamming, hanning and blackman window. *International Journal of Computer Applications*, Foundation of Computer Science, v. 96, n. 18, 2014.

PROAKIS, J. G.; MANOLAKIS, D. G. *Digital signal processing: principles, algorithms, and application*. 3. ed. [S.l.]: Prentice-Hall, Inc., 1996. 1033 p. ISBN 0-13-39438-9.

PUNCH, W. F. et al. Further research on feature selection and classification using genetic algorithms. In: *ICGA*. [S.l.: s.n.], 1993. p. 557–564.

QI, J. et al. Auditory features based on gammatone filters for robust speech recognition. In: IEEE. *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*. [S.l.], 2013. p. 305–308.

- ROEVA, O.; FIDANOVA, S.; PAPRZYCKI, M. Influence of the population size on the genetic algorithm performance in case of cultivation process modelling. In: IEEE. *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on*. [S.l.], 2013. p. 371–376.
- ROSE, P. *Forensic speaker identification*. [S.l.]: CRC Press, 2002. 384 p.
- SANTOS, G. B. d. *Análise fonético-acústica das vogais orais e nasais do português: Brasil e Portugal*. Tese (Doutorado) — Universidade Federal de Goiás, 2013.
- SCHIEL, F. Perception of alcoholic intoxication in speech. In: *Proceedings of the Interspeech*. [S.l.: s.n.], 2011. p. 3281–3284.
- SCHIEL, F. et al. Alcohol language corpus—a publicly available large corpus of alcoholized speech. In: *IAFPA Annual Conference, Trier, Germany*. [S.l.: s.n.], 2010.
- SCHIEL, F.; HEINRICH, C.; NEUMEYER, V. Rhythm and formant features for automatic alcohol detection. In: *Interspeech*. [S.l.: s.n.], 2010. p. 458–461.
- SCHULLER, B. et al. The interspeech 2011 speaker state challenge. In: *INTERSPEECH*. [S.l.: s.n.], 2011. p. 3201–3204.
- SHAHAMAT, H.; POUYAN, A. A. Feature selection using genetic algorithm for classification of schizophrenia using fmri data. *Journal of AI and Data Mining*, v. 3, n. 1, p. 30–37, 2015.
- SHAO, Y.; WANG, D. Robust speaker identification using auditory features and computational auditory scene analysis. In: IEEE. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.], 2008. p. 1589–1592.
- SHIRVAN, R. A.; TAHAMI, E. Voice analysis for detecting parkinson’s disease using genetic algorithm and knn classification method. In: IEEE. *Biomedical Engineering (ICBME), 2011 18th Iranian Conference of*. [S.l.], 2011. p. 278–283.
- SILVA, A. H. P. *Língua Portuguesa I: fonética e fonologia*. [S.l.]: IESDE BRASIL SA, 2007.
- SILVA, A. J. M. *Implementação de um Algoritmo Genético utilizando o modelo de ilhas*. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, 2005.
- SRINIVAS, N.; DEB, K. Multi-objective function optimization using non-dominated sorting genetic algorithms. *Evolutionary Computation*, v. 2, n. 3, p. 221–248, 1994.

- TAN, F. *Improving feature selection techniques for machine learning*. Tese (Doutorado) — Georgia State University, 2007.
- TRASCASA, N. G. *Start-and-End Point Detection at the Input of Speech Recognition Application*. Dissertação (Mestrado) — Universitat Politècnica de Catalunya, 2013.
- TSANAS, A. et al. Novel speech signal processing algorithms for high-accuracy classification of parkinson’s disease. *IEEE Transactions on Biomedical Engineering*, IEEE, v. 59, n. 5, p. 1264–1271, 2012.
- VALE, S. M. M. do. Emissão vocal. uma visão física, fisiológica e psicológica das pregas vocais. Universidade Católica Portuguesa, 2012. Disponível em: <http://www.meloteca.com/pdfartigos/susana-vale_emissao-vocal.pdf>. Acesso em: 21/11/2016.
- WANG, C.-M.; HUANG, Y.-F. Self-adaptive harmony search algorithm for optimization. *Expert Systems with Applications*, Elsevier, v. 37, n. 4, p. 2826–2837, 2010.
- XIAO, H. Diagnosis of parkinson’s disease using genetic algorithm and support vector machine with acoustic characteristics. In: IEEE. *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*. [S.l.], 2012. p. 1072–1076.
- XIE, H. *An analysis of selection in genetic programming*. Tese (Doutorado) — Victoria University of Wellington, 2008.
- ZAMALLOA, M. et al. Feature selection based on genetic algorithms for speaker recognition. In: IEEE. *2006 IEEE Odyssey-The Speaker and Language Recognition Workshop*. [S.l.], 2006. p. 1–8.
- ZAMALLOA, M. et al. Comparing genetic algorithms to principal component analysis and linear discriminant analysis in reducing feature dimensionality for speaker recognition. In: ACM. *Proceedings of the 10th annual conference on Genetic and evolutionary computation*. [S.l.], 2008. p. 1153–1154.
- ZHANG, T. et al. A ga-svm feature selection model based on high performance computing techniques. In: IEEE. *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. [S.l.], 2009. p. 2653–2658.
- ZHANG, Y.; ABDULLA, W. H. *Gammatone auditory filterbank and independent component analysis for speaker identification systems*. [S.l.], 2005.

ZHAO, X.; WANG, D. Analyzing noise robustness of mfcc and gfcc features in speaker identification. In: IEEE. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.], 2013. p. 7204–7208.

ZIELINSKI, K.; PETERS, D.; LAUR, R. Stopping criteria for single-objective optimization. In: CITESEER. *Proceedings of the Third International Conference on Computational Intelligence, Robotics and Autonomous Systems*. [S.l.], 2005.

ZINI, É. d. O. C. *Algoritmo Genético Especializado na Resolução de Problemas com Variáveis Contínuas e Altamente Restritos*. Dissertação (Mestrado) — Universidade Estadual Paulista, 2009.

ZOGLAMI, N.; LACHIRI, Z. Application of perceptual filtering models to noisy speech signals enhancement. *Journal of Electrical and Computer Engineering*, Hindawi Publishing Corp., v. 2012, p. 25, 2012.