

UNIVERSIDADE FEDERAL FLUMINENSE
CENTRO TECNOLÓGICO
ESCOLA DE ENGENHARIA
MESTRADO EM ENGENHARIA ELÉTRICA E DE TELECOMUNICAÇÕES

Luis Eduardo Barrientos Sandoval

Comparação de modelos de sinal glotal na síntese de vogais, nos casos de vogal sustentada e de voz cantada, considerando sons na língua espanhola

Niterói-RJ

2018

LUIS EDUARDO BARRIENTOS SANDOVAL

Comparação de modelos de sinal glotal na síntese de vogais, nos casos de vogal sustentada e de voz cantada, considerando sons na língua espanhola

Dissertação apresentada ao Curso de Mestrado em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Mestre. Área de Concentração: Sistemas de Telecomunicações.

Orientador: Prof. D.Sc. EDSON LUIZ CATALDO FERREIRA

Niterói-RJ

2018

LUIS EDUARDO BARRIENTOS SANDOVAL

Comparação de modelos de sinal glotal na síntese de vogais, nos casos de vogal sustentada e de voz cantada, considerando sons na língua espanhola

Dissertação apresentada ao Curso de Mestrado em Engenharia Elétrica e de Telecomunicações da Universidade Federal Fluminense, como requisito parcial para obtenção do Grau de Mestre. Área de Concentração: Sistemas de Telecomunicações.

Aprovada em 11 de maio de 2018.

BANCA EXAMINADORA

Prof. D.Sc. EDSON LUIZ CATALDO FERREIRA - Orientador
UFF

Prof. D.Sc. LENI JOAQUIM DE MATOS
UFF

Prof. D.Sc. AMERICO BARBOSA DA CUNHA JUNIOR
UERJ

Niterói-RJ
2018

Dedicado a Deus.

Agradecimentos

Agradeço a Deus por colocar em meu caminho todas as bênçãos para conquistar este lindo sonho. Toda a glória seja para Ele.

Agradeço à minha namorada Deisy Yurley Basto, pela sua compreensão, respeito, tolerância e por todas as atitudes que a faz merecedora do meu amor.

Agradeço aos meus pais e irmãos, pela confiança na minha capacidade, pelo apoio, motivação e carinho em todos os momentos da minha vida.

Agradeço especialmente ao professor Edson Cataldo pela orientação e disposição para a realização deste projeto, pela confiança e apoio nas situações de dificuldade, pela bondade e sobretudo sua paciência.

Agradeço aos professores Márcio Fortes, Natália e Murilo pela colaboração e o conhecimento oferecido.

Agradeço aos funcionários da secretaria do programa de pós-graduação, Vânia, Patrícia e Maurício pela atenção e gentileza.

Agradeço à Universidade Federal Fluminense e ao PPGEET, por abrir as portas e permitir que me torne Mestre.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela bolsa de estudos e pelo apoio financeiro de todos os recursos necessários para o desenvolvimento deste trabalho de mestrado.

Lista de Figuras

2.1	<i>Compresorium</i> do Johannes Müller para a experimentação laríngea [1].	4
2.2	Sistema da produção da voz humana [3].	4
2.3	Esquemático do sistema pulmonar [5].	5
2.4	Movimento do peito, diafragma, e abdômen para (a) inspiração, (b) expiração [5].	6
2.5	Estrutura da laringe: (a) Vista Midsagittal, (b) vista anterior, (c) vista lateral [5].	6
2.6	Vista das cordas vocais da laringe (a) Cordas vocais abertas, (b) Cordas vocais fechadas [3].	7
2.7	Representação de um ciclo glotal [6].	7
2.8	Aparelho ressonador ou trato vocal [9].	8
2.9	Diagrama de blocos da teoria acústica de Gunnar Fant [3].	9
2.10	Representação do trato vocal dos espectros da fonte sonora e do filtro [11].	10
2.11	Corte sagital para as vogais /i/ e /u/ do espanhol [12].	10
2.12	Espectrograma do fonema vocálico /a/ [11].	11
2.13	Representação de formantes para os fonemas vocálicos da língua espanhola [13].	12
2.14	Triângulo articulatório projetado na cavidade vocal [14].	12
3.1	Forma de onda da fonte fluxo glotal [27].	17
3.2	Fases do movimento glótico [3].	18
3.3	Terminologia usada para descrever a fonte glótica [19].	19
3.4	Pulso glotal do modelo de Rosenberg e sua derivada: $A_v = 1$, $F_0 = 440 Hz$, $\alpha_1 = 40\%$, $\alpha_2 = 16\%$	20
3.5	Formas de pulso glotal do Modelo de Rosenberg [20].	21
3.6	Modelo do pulso glótico adotado por Fant [3].	21
3.7	Parâmetros unificados da fonte glótica [19].	24
4.1	Esquema do trato vocal [27].	27
4.2	(a) Tubo uniforme sem perdas com terminação ideal, (b) Analogia à linha de transmissão elétrica [27].	30
4.3	(a) Resposta em frequência magnitude Log, (b) Posições dos polos para o tubo uniforme sem perdas [27].	32
4.4	Resposta em frequência magnitude Log para o tubo uniforme de paredes moles [27].	34

4.5	Resposta em frequência magnitude Log para o tubo uniforme com paredes moles e perdas por fricção e condução térmica [27].	35
4.6	(a) Radiação de um defletor esférico, (b) Radiação de um defletor plano infinito [27].	35
5.1	Pulso glotal de Rosenberg (em cima) e sua derivada (embaixo) com: $A_v = 3$, $F_0 = 262 Hz$, $\alpha_1 = 40\%$ e $\alpha_2 = 20\%$	38
5.2	Pulso glotal de Rosenberg (em cima) e sua derivada (embaixo) com: $A_v = 1$, $F_0 = 262 Hz$, $\alpha_1 = 50\%$ e $\alpha_2 = 50\%$	39
5.3	Pulso glotal de Rosenberg (em cima) e sua derivada (embaixo) com: $A_v = 1$, $F_0 = 440 Hz$, $\alpha_1 = 30\%$ e $\alpha_2 = 50\%$	39
5.4	Pulso glotal de LF (em cima) e sua derivada (embaixo) com: $A_v = 3$, $F_0 = 262 Hz$, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$ e $Q_a = 0.1$	40
5.5	Pulso glotal de LF (em cima) e sua derivada (embaixo) com: $A_v = 1$, $F_0 = 262 Hz$, $O_q = 1$, $\alpha_m = 0.5$ e $Q_a = 0.1$	41
5.6	Variações temporais do pulso glotal (em cima) e a sua derivada (embaixo) para quatro valores de O_q no caso em que A_v , F_0 , α_m e Q_a sejam mantidos constantes.	42
5.7	Variações espectrais do pulso glotal para quatro valores de O_q no caso em que A_v , F_0 , α_m e Q_a sejam mantidos constantes.	42
5.8	Variações temporais do pulso glotal (em cima) e a sua derivada (embaixo) para quatro valores de α_m no caso em que A_v , F_0 , O_q e Q_a sejam mantidos constantes.	43
5.9	Variações espectrais do pulso glotal para quatro valores de α_m no caso em que A_v , F_0 , O_q e Q_a sejam mantidos constantes.	43
5.10	Fase de retorno do fluxo glotal dos modelos de Rosenberg e LF para três valores de Q_a	44
5.11	Fase de retorno da derivada do fluxo glotal dos modelos de Rosenberg e LF para três valores de Q_a	44
5.12	Variações temporais do pulso glotal (em cima) e sua derivada (embaixo) para quatro valores de α_1 no caso em que A_v , F_0 e α_2 são mantidos constantes.	45
5.13	Variações espectrais do pulso glotal para quatro valores de α_1 no caso em que A_v , F_0 e α_2 são mantidos constantes.	45
5.14	Variações temporais do pulso glotal (em cima) e sua derivada (embaixo) para quatro valores de α_2 no caso em que A_v , F_0 e α_1 sejam mantidos constantes.	46
5.15	Variações espectrais do pulso glotal para quatro valores de α_2 no caso em que A_v , F_0 e α_1 são mantidos constantes.	46
5.16	Sinal glotal do modelo LF gerado com os parâmetros: $A_v = 1$, $F_0 = 262 Hz$, $O_q = 0.8$, $\alpha_m = 0.7$ e $Q_a = 0.1$	47
5.17	Resposta em frequência do trato vocal para as vogais do espanhol cantadas no caso de um baixo e uso do modelo LF. (a) vogal a, (b) vogal e, (c) vogal i, (d) vogal o, (e) vogal u.	48

5.18	Resposta em frequência do trato vocal para as vogais do espanhol cantadas no caso de um baixo e uso do modelo de Rosenberg. (a) vogal a, (b) vogal e, (c) vogal i, (d) vogal o, (e) vogal u.	49
5.19	Resposta em frequência do trato vocal para as vogais do espanhol cantadas. (a) vogal a, (b) vogal e, (c) vogal i, (d) vogal o, (e) vogal u.	50
5.20	Resposta em frequência do trato vocal para as vogais do espanhol cantadas. (a) vogal a, (b) vogal e, (c) vogal i, (d) vogal o, (e) vogal u.	52
5.21	Simulação da fonte glotal com <i>jitter</i> , usando o modelo de Rosenberg com parâmetros: $A_v = 7$, $F_0 = 98 \text{ Hz}$, $\alpha_1 = 60\%$, $\alpha_2 = 20\%$	56
5.22	Simulação da fonte glotal com <i>jitter</i> , usando o modelo de LF com parâmetros: $A_v = 7$, $F_0 = 98 \text{ Hz}$, $O_q = 0.8$, $\alpha_m = 0.75$, $Q_a = 0.3$	56

Lista de Tabelas

2.1	Notas musicais para a escala musical temperada.	13
2.2	Classificação vocal dos cantores líricos.	14
4.1	Analogias entre quantidades acústicas e elétricas.	30
5.1	Formantes do filtro do trato vocal para o caso de um cantor com classificação vocal de baixo.	49
5.2	Larguras de Banda para os cinco formantes do trato vocal no caso de um cantor com classificação vocal de baixo.	50
5.3	Formantes do filtro do trato vocal para o caso de um cantor com classificação vocal de tenor.	51
5.4	Larguras de Banda dos cinco formantes do trato vocal para o caso de um cantor com classificação vocal de tenor.	51
5.5	Formantes do filtro do trato vocal para o caso de uma cantora com classificação vocal de mezzo-soprano.	52
5.6	Larguras de Banda dos cinco primeiros formantes do trato vocal para o caso de uma cantora com classificação vocal de mezzo-soprano.	53

Sumário

Agradecimentos	v
Lista de Figuras	viii
Lista de Tabelas	ix
Resumo	xiii
Abstract	xiv
1 Introdução	1
2 A voz humana	3
2.1 Introdução	3
2.2 Um pouco da história, da Fisiologia e da Morfologia da produção da voz	3
2.3 Sistema de produção da voz	4
2.3.1 A respiração	5
2.3.2 A fonação	6
2.3.3 A ressonância	8
2.4 A teoria fonte-filtro	9
2.5 Os fonemas	9
2.6 Fonemas vocálicos	10
2.7 Voz cantada	12
3 Modelagem da fonte glotal	16
3.1 Introdução	16
3.2 Fluxo glotal	16
3.3 Fonte glotal	17
3.3.1 <i>Jitter</i>	17
3.4 Pulso glotal	17
3.5 Modelos do pulso glotal	18
3.5.1 Modelo glotal de Rosenberg	20
3.5.2 Modelo glotal de Fant	21

3.5.3	Modelo glotal de Liljencrants-Fant	22
3.6	Unificação dos modelos temporais da fonte glotal	23
3.6.1	Modelo glotal de Rosenberg com parâmetros unificados	24
3.6.2	Modelo glotal de Liljencrants-Fant com parâmetros unificados	24
4	Modelagem do trato vocal humano	26
4.1	Introdução	26
4.2	Geração e propagação do som	26
4.3	Propagação do som no trato vocal	27
4.4	Tubo uniforme sem perdas	29
4.5	Efeitos de perda no trato vocal	32
4.5.1	Perdas devido a paredes moles	33
4.5.2	Perdas devido a fricção viscosa e condução térmica	34
4.5.3	Perdas devido a radiação nos lábios	35
5	Resultados	37
5.1	Introdução	37
5.2	Síntese da voz	37
5.3	Síntese do pulso glotal com o modelo de Rosenberg	38
5.4	Síntese do pulso glotal com o modelo de Liljencrants-Fant	40
5.5	Análise de sensibilidade dos parâmetros para o modelo LF	41
5.6	Análise de sensibilidade dos parâmetros do modelo Rosenberg	44
5.7	Síntese do sinal glotal	46
5.8	Síntese das vogais cantadas com fonte glotal pelos modelo de Rosenberg e LF	47
5.9	Efeitos produzidos pela forma de onda dos pulsos da fonte glotal na voz cantada	53
5.10	Análise de sensibilidade da voz cantada com fonte glotal pelo modelo LF	54
5.11	Análise de sensibilidade da voz cantada com fonte glotal pelo modelo Rosenberg	54
5.12	Síntese de uma sequência de vogais em diferentes notas musicais	55
5.13	Síntese das vogais cantadas com <i>jitter</i> na fonte glotal	55
6	Conclusões e sugestões para trabalhos futuros	57
6.1	Conclusões	57
6.2	Sugestões para trabalhos futuros	59
	Referências Bibliográficas	60
A	Algoritmo do pulso glotal com o modelo de Rosenberg	63
B	Algoritmo do pulso glotal com o modelo de LF	66
C	Algoritmo para gerar a excitação glotal	69

D	Algoritmo para a síntese das vogais cantadas com fonte glotal pelo modelo de Rosenberg	72
E	Algoritmo para a síntese das vogais cantadas com fonte glotal pelo modelo LF	77
F	Algoritmo para a síntese de uma sequência de vogais em diferentes notas musicais pelo modelo de Rosenberg	82
G	Algoritmo para a síntese de uma sequência de vogais em diferentes notas musicais pelo modelo de LF	86
H	Algoritmo para a síntese das vogais cantadas com <i>jitter</i> na fonte glotal pelo modelo de Rosenberg	91
I	Algoritmo para a síntese das vogais cantadas com <i>jitter</i> na fonte glotal pelo modelo de LF	95

Resumo

Essa dissertação aborda um estudo comparativo dos modelos de sinal glotal, na síntese de vogais sustentadas e, também, cantadas com cobertura, no caso de cantores líricos em suas diferentes categorias vocais. Para realizar as comparações, uma série de algoritmos foi implementada na plataforma MATLAB, considerando o modelo fonte-filtro que é muito utilizado em sistemas de processamento para descrever o mecanismo da produção da fala. Os modelos de Rosenberg e o de Liljencrants-Fant (LF) foram escolhidos para representar diferentes naturezas dos pulsos glotais. Os efeitos desses pulsos glotais na síntese de vogais foram considerados no domínio do tempo e da frequência, através da análise de sensibilidade dos parâmetros envolvidos nos modelos do pulso glotal e, também, mediante diferentes formas de onda do pulso glotal. Foram obtidos resultados importantes com os sons produzidos, em relação à naturalidade e inteligibilidade das sínteses. Foram consideradas as frequências formantes e as larguras de banda associadas que determinam as vogais da língua espanhola, além de diferentes timbres vocais. Foi ainda adicionada nas sínteses, a característica importante que faz com que um cantor seja escutado, mesmo que uma melodiosa orquestra o acompanhe. Por último, sínteses de vogais cantadas com *jitter* na fonte glotal, pequenas variações aleatórias em torno da média dos intervalos glotais, foram obtidas em diferentes medidas. Essa variabilidade aleatória na frequência fundamental é usada como uma das ferramentas para diferenciar vozes normais de vozes com características de patologia.

Palavras-chave: modelos de pulso glotal, sinal glotal, síntese de vogais, produção da voz, canto lírico, *jitter*.

Abstract

This dissertation approaches a comparative study with glottal signal models, in the synthesis of sustained vowels and, mainly, sung vowels with cover, in the case of lyric singers in their different vocal categories. To perform this comparisons, a series of algorithms were implemented in MATLAB platform, considering the source-filter model that is widely used in processing systems to describe the mechanism of speech production. The Rosenberg and Liljencrants-Fant (LF) models were chosen to represent different natures of the glottal pulses. The effects of these glottal pulses in the vowel synthesis were considered in time and frequency domain, through the sensitivity analysis of the parameters involved in the glottal pulse models and also through different glottal pulse waveforms. Important results were obtained with the sounds produced, in relation to the naturalness and intelligibility of the synthesis. The formant frequencies and the associated bandwidths that determine the vowels of the Spanish language were considered, as well as different vocal tones. It was also added in the synthesis, the important characteristic which permits that a singer be heard even if an orchestra is accompanying him. Finally, syntheses of vowels sung with jitter in the glottal source, little random variation around the mean values of the glottal intervals were obtained for different measures. This random variability in fundamental frequency is used as one of the tools to identify normal voices and with pathological characteristics.

Keywords: glottal pulse models, glottal signal, vowel synthesis, voice production, lyric sung, jitter.

Capítulo 1

Introdução

A voz tem sido objeto de estudo ao longo da história da humanidade devido a sua importância na comunicação de ideias e sentimentos, porém, muitos pesquisadores têm achado a forma de descrevê-la do ponto de vista fisiológico e anatômico. Na literatura, descreve-se a produção da voz humana como a ação coordenada de sistemas tais como o respiratório, fonador e ressonador. Cada um desses sistemas, além de ter uma função vital, realiza uma tarefa na produção da voz. Todos esses sons gerados são classificados como vozeados ou não vozeados. Os sons vozeados ou sonoros são produzidos pela oscilação das cordas vocais, como as vogais e algumas consoantes, e os não vozeados são gerados por outros fenômenos.

A principal fonte de som para a fala é originada na laringe, especialmente na glote, e é causada por uma coluna de ar proveniente dos pulmões, que passa através das cordas vocais, produzindo-se o som. Essa fonte de som é considerada pela teoria Fonte-Filtro, apresentada por Fant (1960) [21, 23], como a fonte de excitação glotal. A teoria propõe que a convolução da fonte de excitação com um sistema de filtros lineares, conectados em série, modelam a produção da fala. O pulso glotal, pulso de ar correspondente a um ciclo glotal, está descrito através de modelos glotais, e os utilizados neste trabalho são o modelo glotal de Rosenberg e o modelo de Liljencrants-Fant (LF). Esses modelos glotais consideram o sinal glotal, formado por pulsos glotais, como periódico, embora, na vida real, ele apresente flutuações do período fundamental entre 0.1% e 1%, chamadas de *jitter*, para a voz normal, e flutuações superiores a esses valores como indícios de patologias.

Uns dos filtros lineares enunciados na teoria Fonte-Filtro modela o comportamento da propagação do som no trato vocal. A propriedade de ressonância do trato vocal modifica o som produzido na fonte glotal graças aos articuladores. Aquelas frequências harmônicas da fonte glotal são amplificadas e atenuadas e se encontram perto ou longe das frequências de ressonância do trato vocal, respectivamente. Por outro lado, as perdas de som no trato vocal são modeladas pelas larguras de banda do filtro.

As frequências de ressonância do trato vocal, também chamadas de formantes, são de primordial importância nos sons vozeados, pois determinam a qualidade vocal e dão a maior contribuição ao timbre pessoal da voz. No caso dos cantores, os valores destas frequências formantes contém características relevantes que identificam o cantor, dependendo de sua categoria vocal e o mecanismo usado para cobrir as notas de fonação.

Por outra parte, a síntese de fala é um procedimento de geração automática de linguagem falada por computador. Também é referido como comunicação de texto por voz. Nesse processo, um texto em fala normal é transformado em comunicação de voz. O objetivo da síntese de fala é produzir uma máquina que tenha uma voz compreensiva e natural para comunicar [32]. O dispositivo que sintetiza a voz é chamado de sintetizador; o sintetizador pode ser implementado como um hardware ou software. Os sintetizadores de voz são utilizados, em sua maioria, por pessoas cegas ou com baixa visão para acessar as informações exibidas na tela, assim como, no análise de patologias da voz.

Atualmente existem diferentes técnicas de síntese de fala, que mediante o uso da fonética, a linguística de cada linguagem e através de modelos matemáticos, recriam o processo da produção da voz humana. Existem quatro técnicas básicas de sínteses de fala chamadas de: articulatória, formantes, modelo oculto de Markov e concatenação [33]. A técnica usada em este trabalho é a síntese de formantes, que usa o modelo fonte-filtro da fala para fornecer um número infinito de sons. A eficiência ou precisão de qualquer sintetizador de fala é analisada pela sua naturalidade e inteligibilidade da fala gerada, que no caso da síntese de formantes é obtida pela posição das frequências formantes e as larguras de banda associadas que determinam a vogal articulada e sua qualidade [34], no entanto, a principal desvantagem é que o timbre de voz soa mais artificial.

O desafio principal desta dissertação é sintetizar pelo método de formantes as vogais do espanhol, com a característica espectral própria da voz cantada conhecida como o formante do cantor, apresentado por Johan Sundberg (1988)[31], e que diferencia a voz cantada da voz falada. O tipo de canto escolhido para a síntese das vogais corresponde ao canto lírico e se sintetizam para as categorias vocais de baixo, tenor e mezzo-soprano.

Por outro lado, se analisam os efeitos gerados pela variação de parâmetros, dos modelos glotal de Rosenberg e de LF, identificando o modelo glotal que permita alcançar maior naturalidade à síntese das vogais e, também, revelar os efeitos causados pela variação dos seus parâmetros e na forma de onda do pulso glotal. Além disso, se analisa o comportamento da síntese das vogais cantadas ao adicionar o *jitter* à fonte glotal por um método simples, e dessa forma, comparar os sons produzidos, tanto com características determinísticas quanto aleatórias.

Em conformidade com o anterior, as contribuições mais importantes que apresenta esta dissertação se centram em torno dos programas de sínteses de vogais cantadas, pelo mecanismo de cobertura no canto lírico, na plataforma Matlab, assim como os análises apresentados sobre os efeitos produzidos pela variação dos parâmetros dos modelos de pulso glotal e os efeitos produzidos pela presença do *jitter* na voz cantada. Esses programas de síntese de voz cantada podem ser modificados em trabalhos futuros para realizar análises sobre os efeitos produzidos por outros modelos de fonte glotal e pelo efeito de um *jitter* por métodos mais complexos.

Capítulo 2

A voz humana

2.1 Introdução

A voz é a base da comunicação humana através da qual transmitem-se informações, expressam-se sentimentos e ideias. Devido a sua excepcional importância, pesquisadores e pensadores, ao longo da história da humanidade, realizaram estudos com o objetivo de conhecer sua fisiologia.

No decorrer das pesquisas muitos experimentos foram realizados com laringes artificiais e de cadáveres de animais e humanos, com o objetivo de construir teorias e aportes sobre as características morfológicas da laringe e determinar a forma como a voz humana é produzida.

Neste capítulo, apresenta-se uma descrição do processo da produção da voz humana e seu comportamento, para o caso da produção de uma vogal sustentada e cantada pelo método de cobertura, além de compreender os antecedentes históricos mais importantes da voz e a função do aparelho respiratório, fonador e ressonador na produção da voz, não só da língua portuguesa mas, também, da língua espanhola.

2.2 Um pouco da história, da Fisiologia e da Morfologia da produção da voz

Notáveis pesquisas quanto à fisiologia e à morfologia da laringe, realizadas a partir do século XVII, permitiram evidenciar o fato de que dois tecidos moles juntos (as cordas ou pregas vocais) podiam produzir um som a partir da excitação do fluxo de ar. Por outro lado, indicou-se a existência de importantes cartilagens que constituíam a laringe (aritenóides e cricóides) e da função de seus músculos internos, como os encarregados de abrir e fechar a glote [2]. Em torno dos anos 1600, o médico francês Claude Perrault, considerou a voz como um ruído produzido pela saída violenta do ar que, por sua passagem ao exterior, esfrega as duas membranas da glote. Além disso, revelou que a geração dos diferentes tons variam com a longitude e tensão das pregas vocais.

No começo do século XIX, o físico francês Henri Dutrochet e o médico, também francês, Françoise Magendie, mostraram que o tom da voz é modificado pela intervenção do músculo tireoaritenódeo e, além disso, que o fechamento da glote depende da ação do nervo laríngeo superior e da abertura pelo nervo

laríngeo inferior [1].

No final do século XIX, experimentos realizados pelo professor de fisiologia Johannes Müller, forneceu dados específicos sobre a produção da voz humana baseando-se em testes com tecidos elásticos membranosos e laringes, obtidos de cadáveres de animais e humanos, mediante seu aparelho *compresorium* ilustrado na Fig. 2.1.

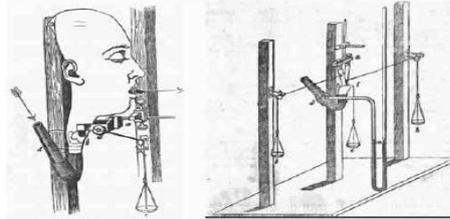


Figura 2.1: *Compresorium* do Johannes Müller para a experimentação laríngea [1].

As bases das teorias fornecidas pelo professor Johannes Müller ainda são vigentes e têm permitido o desenvolvimento de estudos posteriores na fisiologia da voz humana. No entanto, a necessidade de obter informação do comportamento da laringe em sua situação fisiológica normal, foi resolvida pelo professor espanhol Manuel Garcia, que conseguiu observar, em detalhes, as características da vibração da mucosa das pregas vocais com o uso de um espelho laríngeo (chamado hoje em dia laringoscopia indireta), chegando a diagnósticos corretos.

2.3 Sistema de produção da voz

A Fig. 2.2 ilustra os órgãos do sistema da produção da voz humana.

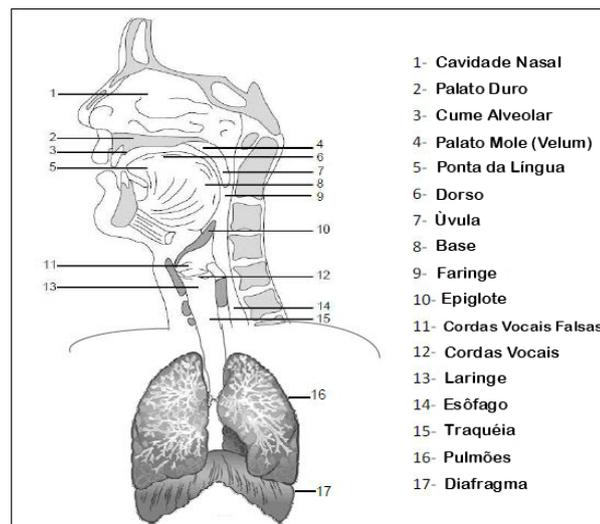


Figura 2.2: Sistema da produção da voz humana [3].

A produção da voz não ocorre somente na laringe, pois, com respeito à fisiologia moderna, ela ocorre pela ação coordenada de várias estruturas do corpo humano. Estas estruturas compõem o sistema

fonador ilustrado na Fig. 2.2, o qual conforme [4] é composto pelos músculos de diferentes regiões do corpo, elementos do aparelho respiratório e digestivo.

O fundamento para produzir uma voz efetiva é baseado na coordenação de três ações feitas pelo sistema fonador: a respiração, a fonação e a ressonância.

2.3.1 A respiração

Uma das funções mais importantes que o corpo humano desempenha para garantir a produção da voz é a respiração, a qual não poderia ocorrer sem a intervenção do sistema pulmonar. Como se pode ver no diagrama esquemático da Fig. 2.3, a maioria dos componentes deste sistema são os pulmões, a traqueia, a glote, e o trato vocal (situado entre a glote e a boca).

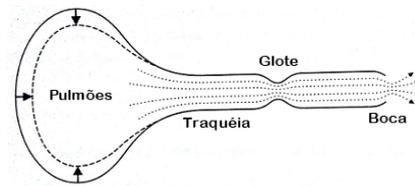


Figura 2.3: Esquemático do sistema pulmonar [5].

Para Ingo Titze [5], o tórax e o abdômen, junto com os pulmões, atuam como um fole para produzir um fluxo de ar. A glote atua como uma válvula reguladora do fluxo. Durante a inspiração (inalação) os pulmões expandem-se fazendo com que o ar flua da boca em direção a eles, com a glote relativamente aberta. Durante a expiração (exalação) os pulmões contraem-se, empurrando o ar em direção à boca.

Dito de outro modo, a respiração se dá em dois momentos: inspiração e expiração. Na inspiração, a glote permanece aberta, seguida por uma entrada de ar que ocorre pela contração da musculatura do diafragma e dos músculos intercostais externos, fazendo com que os pulmões se encham de ar, causando um pouco de pressão pulmonar negativa e um aumento de volume da caixa torácica, como se observa na Fig. 2.4 (a). Posteriormente, a parte expiratória ilustrada na Fig. 2.4 (b) acontece pelo relaxamento do diafragma, a contração dos músculos intercostais internos e abdominais e pelo retrocesso elástico dos pulmões que diminuem o volume da caixa torácica e expulsam (pela pressão exercida) o ar contido nos pulmões.

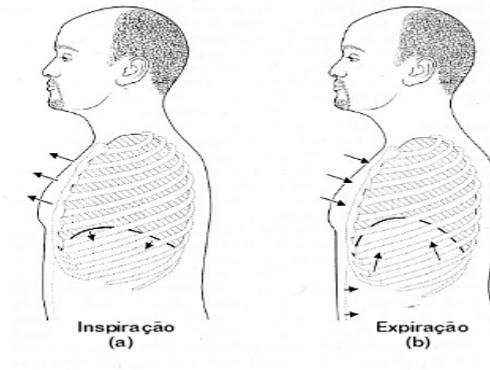


Figura 2.4: Movimento do peito, diafragma, e abdômen para (a) inspiração, (b) expiração [5].

2.3.2 A fonação

Para compreender a fisiologia da fonação é importante revisar, em primeiro lugar, a anatomia básica da laringe, encarregada de produzir a principal fonte de som para a fala.

A Fig. 2.5 mostra a laringe, situada entre a faringe e traqueia, e a maior parte de sua estrutura, na forma de cartilagem, com apenas um osso (hioide) que flutua em relação ao esqueleto.

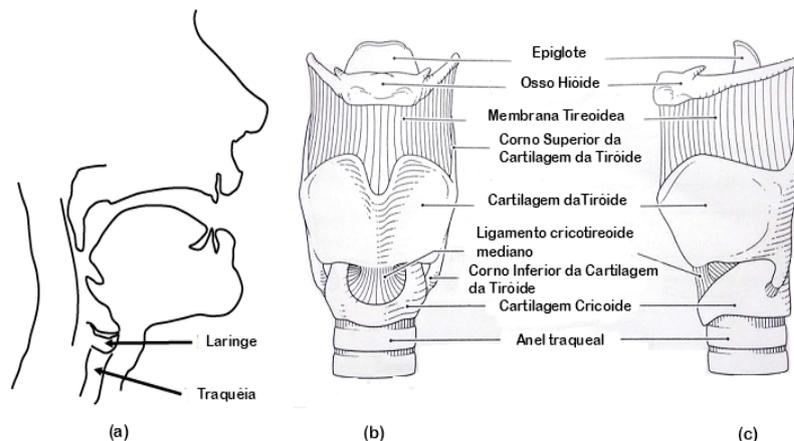


Figura 2.5: Estrutura da laringe: (a) Vista Midsagittal, (b) vista anterior, (c) vista lateral [5].

Os elementos que compõem a estrutura da laringe são: dois anéis traqueais, as cartilagens tireoidianas e cricóides e os ossos hioídes. Alguns ligamentos e membranas, também são ilustrados para indicar como as cartilagens e o osso hioíde estão interligados [5].

As cartilagens laríngeas (tireóide, cricóide, aritenóide e epiglote) são interligadas através de músculos (intrínsecos) para permitir o movimento da laringe e assim, desempenhar o processo de fonação, proteger o trato respiratório na deglutição e controlar o fluxo de ar na respiração, embora as cartilagens também atuam como ressonadores acústicos na produção da voz. Estas funções são coordenadas pelo cérebro através de conexões nervosas específicas. Por outro lado, é importante clarificar que esta estrutura laríngea está conectada às estruturas circundantes através de músculos (extrínsecos).

Segundo [3], as cordas vocais são os componentes funcionais mais relevantes da laringe porque funcionam como um gerador de sons, ou seja, sua função é oscilar com a passagem do ar. Elas incluem três camadas distintas que trabalham em conjunto para promover tal oscilação: a camada interna é o corpo, composto pelo músculo tiroaritenóideo, seguido por uma camada de ligamento vocal e, finalmente, cobertas por uma membrana mucosa que é essencial para a ondulação das cordas vocais durante a produção do som. Na Fig. 2.6, observa-se uma abertura de forma triangular que fica entre elas, chamada glote.

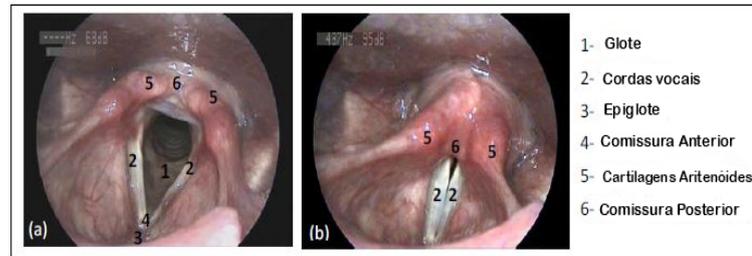


Figura 2.6: Vista das cordas vocais da laringe (a) Cordas vocais abertas, (b) Cordas vocais fechadas [3].

No processo de respiração, a glote se abre permitindo a livre passagem do ar, não ocorrendo a oscilação das cordas vocais e, portanto, sem produção sonora. Para ocorrer a fonação, a glote fecha-se (adução) e uma coluna de ar, expelida dos pulmões, quebra a resistência que as cordas vocais criam para essa passagem, produzindo um som ou zumbido. Cada vez que isso acontece, um ciclo glotal é cumprido tal como é apresentado na Fig. 2.7.

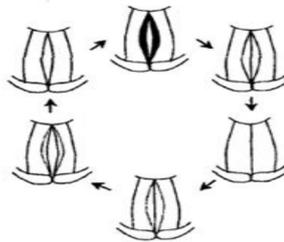


Figura 2.7: Representação de um ciclo glotal [6].

Quando as pregas se encontram em adução, a pressão do ar que existe abaixo da glote (pressão subglótica) se torna maior do que a pressão do ar acima dela (pressão supraglótica). Essa força faz com que as cordas vocais comecem a se separar, pouco a pouco, começando com a parte inferior, até que fique apenas o contato da parte superior e, finalmente, elas se separem completamente. Após a separação total, elas começam a se unir novamente por baixo e, devido ao efeito Bernoulli, uma espécie de sucção faz com que toda a superfície entre em contato novamente, possibilitando o fechamento completo das cordas vocais. A teoria mais aceita que explica a fonação é a aerodinâmica mioelástica (Van den Berg, 1958), completada com a teoria Muco-ondulatória de Perelló (1962): mioelástica pela configuração das cordas e aerodinâmica pelo efeito do ar [6].

Os sons gerados como resultado da oscilação das cordas vocais são chamados de vozeados ou sonoros, e aqueles produzidos sem oscilação (as cordas vocais permanecem abertas apenas e a excitação

glotal é barulhenta) são referidos como não vozeados [3].

O comprimento das cordas vocais varia em conformidade com o gênero e a idade do ser humano, sendo mais curtas e finas em recém-nascidos (em torno de 3 mm), e mais compridas em adultos ($9\text{ mm} - 13\text{ mm}$ e $15\text{ mm} - 20\text{ mm}$ para os sexos feminino e masculino, respectivamente).

A oscilação das cordas vocais contribuem para a produção de diferentes sons que acontecem com uma certa frequência. Esta frequência chamada de fundamental, denotada por F_0 , é normalmente maior em crianças e mulheres e é inversamente proporcional ao comprimento das cordas.

2.3.3 A ressonância

Para Ventura [7], tanto na voz cantada como na falada, a ressonância vocal consiste na modificação do som produzido pelas cordas vocais, obtida através das cavidades de ressonância, onde ocorrem efeitos de atenuação e amplificação do sinal sonoro em regiões específicas de frequência. No canto, esta ressonância é de uma forma geral elevada, existindo uma maior concentração de energia sonora na parte superior do trato vocal. O objetivo de utilizar uma ressonância equilibrada tem a ver com o fato de aliviar a sobrecarga muscular da laringe. Muitos cantores utilizam um certo grau de nasalidade para dissipar a energia sonora sem sobrecarregar a laringe.

Graças aos articuladores (língua, lábios, palato duro, palato mole, dentes, vélum, mandíbula inferior e laringe) o som proveniente das pregas vocais é dirigido para o aparelho ressonador (ou trato vocal). As manobras feitas para modificar a forma do trato vocal durante a fonação se denomina articulação [8].

O aparelho ressonador (ou trato vocal) é composto pela caixa de ressonância inferior (faringe, traqueia, brônquios e pulmões) e pela caixa de ressonância superior (cavidade bucal e nasal), como pode ser observado na Fig. 2.8.

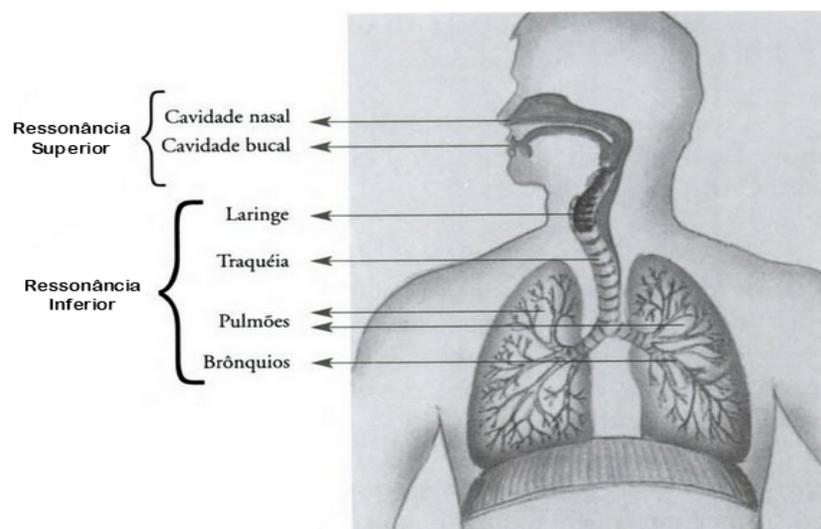


Figura 2.8: Aparelho ressonador ou trato vocal [9].

Os ressonadores, além de amplificar ou atenuar o som gerado pelas pregas vocais, dão brilho e passam facilmente de grave para agudo. No canto, os ressonadores são muito utilizados e conhecidos como ressonadores faciais e de peito.

2.4 A teoria fonte-filtro

A teoria acústica apresentada por Gunnar Fant [21, 23], em 1960, consiste em modelar o mecanismo de produção da fala como a convolução entre uma fonte de excitação e um sistema de filtros lineares conectados em série. Neste caso, o fluxo de ar, que faz vibrar as cordas vocais, é considerado a fonte de excitação e a ressonância vocal, que muda o som (onda sonora) com o tempo, é representado por um filtro, assim como os lábios que irradiam a voz. Esta teoria apresenta duas fortes suposições:

1. A fonte e o filtro são considerados independentes entre si.
2. No domínio do tempo, o processo da produção da fala pode ser representado pela convolução de seus elementos (A fonte glotal, o filtro do trato vocal e o filtro da radiação dos lábios e narinas).

A primeira suposição implica que a fonte glotal é igual ao fluxo glotal o qual, na realidade, não é perfeitamente válido porque existe uma interação entre fonte-filtro, e o fluxo glotal depende em certo grau das variações da impedância do trato vocal [3].

Essa teoria, contudo, é amplamente utilizada em sistema de processamento da fala, e é considerada viável na maioria dos casos de interesse. Este modelo é simplificado como a relação entre a fonte glotal, o trato vocal e a irradiação pelos lábios/narinas, como é representado na Fig. 2.9.



Figura 2.9: Diagrama de blocos da teoria acústica de Gunnar Fant [3].

A teoria fonte-filtro do processo da produção da fala pode ser expresso no domínio do tempo por:

$$s[n] = g[n] * t[n] * l[n], \quad (2.1)$$

onde $s[n]$ é o sinal da fala, $g[n]$ é o sinal da fonte de excitação, $t[n]$ é a resposta ao impulso do trato vocal e $l[n]$ as características de irradiação dos lábios/narinas.

2.5 Os fonemas

O fonema é a imagem mental do som da fala utilizado para transmitir ideias e comunicar informação. O fonema pode ser representado através de um conjunto finito de símbolos básicos ou à concatenação deles.

Este conjunto de símbolos possui, normalmente, de 30 a 50 elementos, que podem ser divididos basicamente em 4 classes: Vogais, ditongos, semivogais e consoantes [10]. Neste trabalho, apenas será especificada a classe das vogais da língua espanhola.

A fonologia estuda os sons da fala do ponto de vista da função que realiza dentro de um sistema linguístico particular, portanto, ela estabelece os sons que ocorrem em uma língua e os símbolos básicos ou

fonemas, que são particulares de uma língua. Enquanto que a fonética estuda os sons da fala do ponto de vista da sua realização física, prestando atenção aos mecanismos utilizados pelo ser humano para produzir e receber o som.

2.6 Fonemas vocálicos

A estrutura da onda sonora (produzida pela oscilação das cordas vocais) é modificada pela ação dos filtros que, em conjunto, causam essa modificação e produzem os sons, mais especificamente os fonemas vocálicos e consonânticos das línguas [11]. A Fig. 2.10 é uma representação do trato vocal dos espectros da fonte sonora e do filtro.

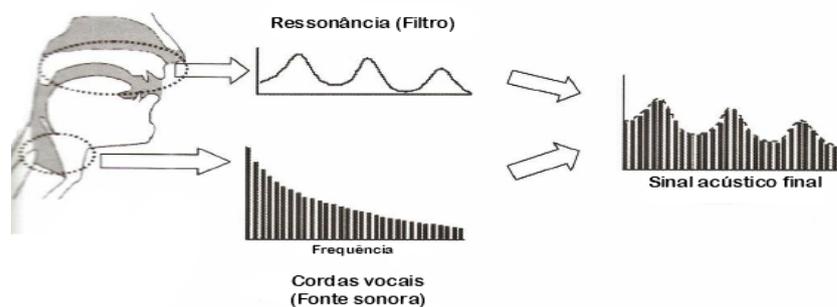


Figura 2.10: Representação do trato vocal dos espectros da fonte sonora e do filtro [11].

Qualquer fonema vocálico pode ser produzido pelo trato vocal, apenas ao encontrar-se as cordas vocais em adução e a livre passagem de pulsos de ar, não interrompidos, causam as oscilações das cordas vocais. Todas as vogais são sonoras; porém, quando sussurradas, são surdas.

Estes são os fonemas básicos que desempenham o papel de núcleo das sílabas, isso significa que, em toda sílaba há, necessariamente, uma única vogal. As diferentes vogais do espanhol (a, e, i, o, u) dispõem de uma articulação apropriada para sua produção, assim como se observa na Fig. 2.11.



Figura 2.11: Corte sagital para as vogais /i/ e /u/ do espanhol [12].

A partir de uma perspectiva acústica, as oscilações das cordas vocais são consideradas uma fonte de produção de ondas complexas, que são constituídas por uma combinação de ondas senoidais mais simples, com frequências e amplitudes diferentes. A frequência mais baixa (ou básica) da onda complexa é chamada de frequência fundamental da voz ou primeiro harmônico (F_0) e os fatores físicos que a regulam

são: a massa, o comprimento e a tensão das pregas vocais, sendo esta última, controlada pelos músculos da laringe. As outras frequências que constituem a onda complexa, múltiplas da F_0 , são denominadas frequências harmônicas. Por exemplo, se a F_0 da onda é de 80 Hz , então, o segundo harmônico é de 160 Hz , o terceiro de 240 Hz e assim sucessivamente [11].

A cavidade bucal e a nasal, em conjunto, são as maiores modificadoras da onda sonora produzida na laringe. Elas exercem o papel de filtro acústico e dão formato aos sons como as vogais. Esse processo de filtragem resulta nos formantes, que são definidos como as frequências ressonantes e caracterizam o timbre de um som [11].

Geralmente, a análise acústica dos fonemas vocálicos é feita através dos dois primeiros formantes, que resultam da ação da faringe e da cavidade bucal, sendo assim, os dois principais ressoadores do aparelho fonador.

Na Fig. 2.12, apresenta-se o espectrograma do fonema vocálico $/a/$, que permite observar seus formantes, como as áreas mais escuras da linha horizontal.

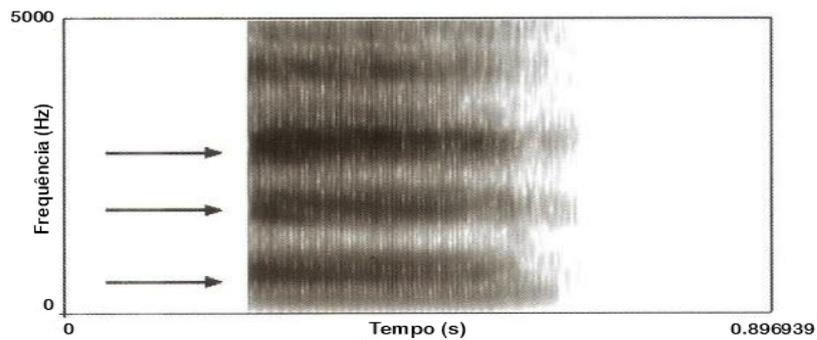


Figura 2.12: Espectrograma do fonema vocálico $/a/$ [11].

Além de ter mais frequências formantes, a natureza de uma vogal depende, principalmente, da posição das duas primeiras frequências formantes, sendo que estas contêm as características mais relevantes para sua identificação e análise espectral.

O espaço articulatório das vogais expõe, em uma escala numérica, o ponto que uma determinada vogal ocupa de acordo com os valores de suas duas primeiras frequências formantes. Os fonemas vocálicos da língua espanhola limitam dentro desse espaço articulatório assim: o $/i/$ e o $/u/$ ocupam as áreas limítrofes (à direita e à esquerda superior), o $/a/$ ocupa a área central inferior desse espaço entre $/i/ - /u/$ e as outras vogais do espanhol situam-se em posições intermediárias entre $/i/ - /a/$ e $/u/ - /a/$, criando um triângulo articulatório.

Na Fig. 2.13, pode-se observar o posicionamento das vogais utilizadas no espanhol, onde $/i/$, $/u/$ e $/a/$ ficam nas pontas do triângulo, e os valores da primeira frequência formante F_1 estão no eixo vertical e da segunda frequência formante F_2 no eixo horizontal.

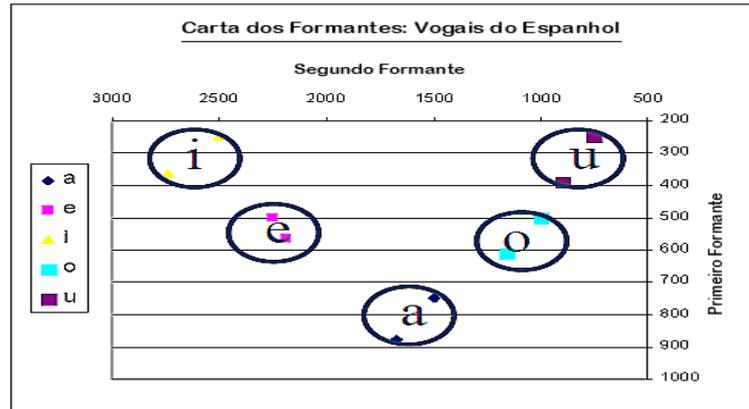


Figura 2.13: Representação de formantes para os fonemas vocálicos da língua espanhola [13].

O triângulo articulatório projetado na cavidade vocal da Fig. 2.14 permite observar que, os formantes dos fonemas vocálicos têm relação direta com sua classificação linguística. As duas primeiras frequências formantes das vogais, F_1 e F_2 , estão relacionadas com o posicionamento do dorso da língua, portanto, a frequência formante F_1 está associada com o levantamento e o abaixamento do dorso da língua, e a frequência formante F_2 está associada com o avanço e o recuo do dorso da língua.

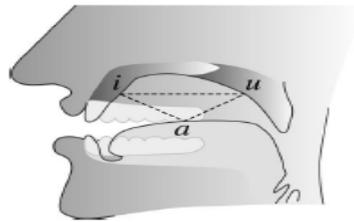


Figura 2.14: Triângulo articulatório projetado na cavidade vocal [14].

A F_1 apresenta uma frequência alta na medida que o dorso da língua se posiciona na parte baixa da cavidade vocal, por exemplo, quando produz o fonema vocálico /a/, e obtém uma frequência baixa na medida que a língua se posiciona na parte alta da cavidade vocal. A F_2 apresenta uma medida alta em frequência quando o dorso da língua é posicionado na frente da cavidade vocal, neste caso, ao produzir o fonema /i/, e uma medida baixa em frequência com o recuo do dorso da língua, por exemplo, ao produzir o fonema /u/.

2.7 Voz cantada

As diferentes notas musicais e suas frequências equivalentes aproximadas para a escala musical temperada são apresentadas na Tab. 2.1.

									<i>La</i> 0	<i>La</i> #0	<i>Si</i> 0
									28	29	31
<i>Do</i> 1	<i>Do</i> #1	<i>Re</i> 1	<i>Re</i> #1	<i>Mi</i> 1	<i>Fa</i> 1	<i>Fa</i> #1	<i>Sol</i> 1	<i>Sol</i> #1	<i>La</i> 1	<i>La</i> #1	<i>Si</i> 1
33	35	37	39	41	44	46	49	52	55	58	62
<i>Do</i> 2	<i>Do</i> #2	<i>Re</i> 2	<i>Re</i> #2	<i>Mi</i> 2	<i>Fa</i> 2	<i>Fa</i> #2	<i>Sol</i> 2	<i>Sol</i> #2	<i>La</i> 2	<i>La</i> #2	<i>Si</i> 2
65	69	73	78	82	87	92	98	104	110	117	123
<i>Do</i> 3	<i>Do</i> #3	<i>Re</i> 3	<i>Re</i> #3	<i>Mi</i> 3	<i>Fa</i> 3	<i>Fa</i> #3	<i>Sol</i> 3	<i>Sol</i> #3	<i>La</i> 3	<i>La</i> #3	<i>Si</i> 3
131	139	147	156	167	175	185	196	208	220	233	247
<i>Do</i> 4	<i>Do</i> #4	<i>Re</i> 4	<i>Re</i> #4	<i>Mi</i> 4	<i>Fa</i> 4	<i>Fa</i> #4	<i>Sol</i> 4	<i>Sol</i> #4	<i>La</i> 4	<i>La</i> #4	<i>Si</i> 4
262	277	294	311	330	349	370	392	415	440	466	494
<i>Do</i> 5	<i>Do</i> #5	<i>Re</i> 5	<i>Re</i> #5	<i>Mi</i> 5	<i>Fa</i> 5	<i>Fa</i> #5	<i>Sol</i> 5	<i>Sol</i> #5	<i>La</i> 5	<i>La</i> #5	<i>Si</i> 5
523	554	587	622	659	698	740	784	831	880	932	988
<i>Do</i> 6	<i>Do</i> #6	<i>Re</i> 6	<i>Re</i> #6	<i>Mi</i> 6	<i>Fa</i> 6	<i>Fa</i> #6	<i>Sol</i> 6	<i>Sol</i> #6	<i>La</i> 6	<i>La</i> #6	<i>Si</i> 6
1.047	1.109	1.175	1.245	1.319	1.397	1.480	1.568	1.661	1.760	1.865	1.976
<i>Do</i> 7	<i>Do</i> #7	<i>Re</i> 7	<i>Re</i> #7	<i>Mi</i> 7	<i>Fa</i> 7	<i>Fa</i> #7	<i>Sol</i> 7	<i>Sol</i> #7	<i>La</i> 7	<i>La</i> #7	<i>Si</i> 7
2.093	2.218	2.349	2.489	2.637	2.794	2.960	3.136	3.322	3.520	3.729	3.951
<i>Do</i> 8											
4.186											

Tabela 2.1: Notas musicais para a escala musical temperada.

Esses valores são usados para representar as variações da intensidade do som musical e facilitar a composição de melodias. Como pode-se observar, existem sete notas da escala natural: Dó, Ré, Mi, Fá, Sol, Lá, Si, e cinco notas alteradas *Do*#1, *Re*#1, *Fa*#1, *Sol*#1, e *La*#1 (# lê-se sustenido). O intervalo entre duas notas se enumera de acordo com o número de notas da escala natural que as separa, incluindo as duas dos extremos, por exemplo, o intervalo *Do*1 - *Do*2 é conhecido de oitava: 1 *Do*1, 2 *Re*1, 3 *Mi*1, 4 *Fa*1, 5 *sol*1, 6 *La*1, 7 *Si*1, 8 *Do*2.

Entre cada nota e a seguinte, seja natural ou alterada, haverá sempre um semitom (ST), a metade de um tom (T). Para encontrar a proporção correspondente a qualquer outro intervalo da escala temperada, apenas tem-se que elevar $\sqrt[12]{2}$ para o número de semitons (ST) que esse intervalo contém. Para dar um exemplo, para passar da frequência de um *Do*1 a um *Fa*#1, multiplica-se a frequência do *Do*1 por $(\sqrt[12]{2})^6$, pois há seis semitons que os separam: 1: *Do*1 - *Do*#1, 2: *Do*#1 - *Re*1, 3: *Re*1 - *Re*#1, 4: *Re*#1 - *Mi*1, 5: *Mi*1 - *Fa*1, 6: *Fa*1 - *Fa*#1.

Uma das características que diferenciam a voz falada da voz cantada é a extensão vocal usada em cada caso. Segundo [29], a extensão vocal é o número total de notas, desde a mais grave à mais aguda, que podem ser emitidas sem esforço e com boa qualidade vocal. Na fala normal, esta extensão vocal usada é curta (usam-se em torno de 4 ou 5 notas de acordo com a identidade da pessoa), enquanto que no canto, a extensão usada é duas oitavas (16 notas) em um cantor lírico e uma oitava e meia (12 notas) em cantores populares com boa qualidade.

De acordo com a extensão vocal, a voz do cantor pode ser classificada em uma categoria vocal. Na voz feminina são conhecidas as categorias de contralto, mezzo-soprano, e soprano; na voz masculina, as categorias de baixo, barítono e tenor. Esta classificação vocal é mostrada na Tab. 2.2 [29].

CLASSIFICAÇÃO VOCAL					
HOMEM			MULHER		
Categoria vocal	Extensão	Nota de passo	Categoria vocal	Extensão	Nota de passo
Baixo	Re2 - Fa4	Si3	Contralto	Fa3 - Sol5	Si4 / Do5
Barítono	Sol2 - La4	Re4	Mezzo-Soprano	Sol3 - La5	Re5
Tenor	Si2 - Do5	Mi4	Soprano	Do4 - Re6 / Sol6	Mi5
Do4 = Do Central = 262 Hz					

Tabela 2.2: Classificação vocal dos cantores líricos.

O cantor consegue chegar aos tons mais agudos de sua extensão vocal, através de dois mecanismos da técnica vocal do canto:

1. Aumentando a pressão subglótica e, portanto, a tensão muscular no músculo vocal. Este mecanismo é usado pelos cantores populares.
2. Descer a laringe e dilatar a faringe inferior, gerando um alongamento das cordas vocais com menor variação de pressão subglótica. Este mecanismo denominado cobertura dos sons agudos é usado pelos cantores líricos.

Em um estudo feito por J.Sundberg [31], em 1988, se mostrou que existe uma diferença considerável em relação às características espectrais dos fonemas vocálicos cantados por cantores masculinos e altos. A diferença mais aparente é chamada de **formante do cantor**. Este formante é um pico do envelope do espectro que aparece na proximidade de 3 kHz (2.2 kHz e 3.8 kHz) em todos os espectros de vogais cantados com cobertura por cantores masculinos e, também, por altos e pertence às características típicas de uma vogal cantada. O formante do cantor é gerado a partir do agrupamento das formantes superiores (terceiro, quarto e quinto) ao ter as suas frequências próximas uma da outra.

O nível do pico do envelope do espectro varia dependendo da classificação vocal. Ele é menor para um baixo e mais alto para um tenor. Os sopranos, por outro lado, têm um nível de espectro muito menor desse pico e, em seguida, as outras categorias. No caso da soprano, este pico não é mais que um terceiro ou quarto formante. O nível deste pico também varia dependendo da intensidade da fonação, efeito derivado da fonte glotal.

Quanto à frequência central, o pico varia dependendo da categoria de voz. Em cantores baixos, concentra-se em torno de 2.2 kHz ; em barítonos, em torno de 2.7 kHz ; em tenores, cerca de 3.2 kHz e em altos, cerca de 2.8 kHz . Essas diferenças de frequência parecem contribuir significativamente para as diferenças tímbricas entre essas categorias de voz.

A presença do formante do cantor no espectro de um som de vogal ajuda a ouvir a voz do cantor através de um alto acompanhamento orquestral. No espectro do som de uma orquestra sinfônica, os parciais perto de 500 Hz tendem a ser mais altos e, acima desta região de frequência, os níveis dos componentes do espectro diminuem com a frequência crescente. O espectro do som da fala normal tem características similares à orquestra sinfônica, portanto, se um cantor emitira sons com a mesma distribuição de energia própria da fala normal, os sons mais altos dos instrumentos da orquestra afogariam a sua voz.

O formante 1 (F_1) de um som vocálico cantado nunca pode ser gerado abaixo da frequência glótica (F_0) que está sendo emitida. No canto feminino, principalmente, quando emitem-se notas agudas, aproximadamente acima do $La4$ ($440 Hz$), deve-se abrir progressivamente a boca em busca da configuração correta com sua respectiva afinação e uma boa qualidade vocal presente graças ao formante do cantor. Isso gera o deslocamento do F_1 para valores mais elevados dos que têm a voz falada. O F_1 aproxima-se notavelmente ao valor F_0 [29, 30]. Em consequência disso, a emissão do fonema vocálico é modificada em muitos casos, porém, o cantor deve treinar a sua voz e conseguir que a dicção da vocal cantada seja sacrificada no menor grau possível.

O tempo de emissão dos fonemas são, em alguns casos, maiores no canto do que na fala e dependem da melodia, porém, a respiração nos cantores deve ser treinada para que os fonemas sejam emitidos de acordo com a melodia. Todas as pessoas têm um timbre natural, próprio e original, que depende do trato vocal e que permite diferenciar cada indivíduo ao executar a mesma frequência. No canto, o timbre é o mesmo que na fala, mas alguns cantores fazem ajustes no timbre para enriquecer a sua voz.

Capítulo 3

Modelagem da fonte glotal

3.1 Introdução

A oscilação das cordas vocais é produzida pela modulação da área glótica, devido ao fluxo da coluna de ar proveniente dos pulmões, o fluxo glotal, como já discutido. Essa oscilação dá origem a uma fonte acústica quase periódica, que apresenta pequena flutuação em torno do período fundamental, denominada *jitter*. Se esta flutuação ultrapassa, em média, o valor de 1%, indica alguma patologia relacionada à voz. A teoria Fonte-Filtro simplifica as características da fonte acústica, a fonte glotal, e facilita que seja modelada através de um trem de pulsos glotais com características similares à produzida na glote. O pulso glotal representa um ciclo glotal completo, composto por fases referidas aos aspectos fisiológicos da fonação (ao movimento vibratório das cordas vocais) e é modelado sob certas suposições do seu comportamento, entre eles o modelo glotal de Rosenberg, o modelo glotal de Fant e o modelo glotal de Liljencrants-Fant (LF). Devido à elaboração de diversos modelos glotais, surgiram parâmetros semelhantes com nomes diferentes e modelos com números diferentes de parâmetros. O objetivo deste capítulo é a unificação de parâmetros dos modelos temporais do pulso glotal. É importante destacar que essa é uma novidade apresentada por essa dissertação.

3.2 Fluxo glotal

O fluxo glotal apresenta as seguintes características:

- A velocidade de deslocamento das pregas vocais e o fluxo glotal não estão sincronizados, isto é, quando o fluxo de ar se aproxima às pregas vocais, o acúmulo e colapso do ar que ocorre na glote, gera um atraso em relação à abertura e fechamento glotal.
- O fluxo glotal é uma função não linear da área glotal e da pressão subglótica.
- O fluxo glotal depende da forma do trato vocal. Como consequência, o trato vocal é considerado como uma carga acústica acima da glote.

3.3 Fonte glotal

Qualquer tecido que possa vibrar no aparelho fonador pode ser considerado como uma possível fonte acústica, contudo, considera-se apenas a vibração das cordas vocais como a fonte dos fonemas vozeados em que a qualidade da voz depende [15]. Essa vibração é quase periódica, dado que o movimento das cordas vocais é mantido em um mecanismo auto-sustentado.

Embora o modelo fonte-filtro seja considerado a base da produção da fala, ele apresenta algumas diferenças importantes que não permitem descrever exatamente o modelo acústico. Uma destas diferenças está relacionada à modelagem da fonte glotal, sendo essa a principal diferença entre o modelo fonte-filtro e o modelo acústico, no entanto, o modelo acústico foi simplificado pelo modelo fonte-filtro.

A fonte glotal foi abordada através do desenvolvimento de modelos, que tentaram explicá-la e reproduzi-la fielmente, portanto, ela é representada pela forma de onda da sua velocidade que, em um ciclo glotal completo, é designado de pulso glotal.

Embora, na implementação sintética da fala, a fonte glotal possa ser gerada por qualquer trem de pulsos, se quisermos obter maior naturalidade da voz é melhor que o trem de pulsos seja o mais próximo possível do que é produzido na glote, como é ilustrado na Fig. 3.1.

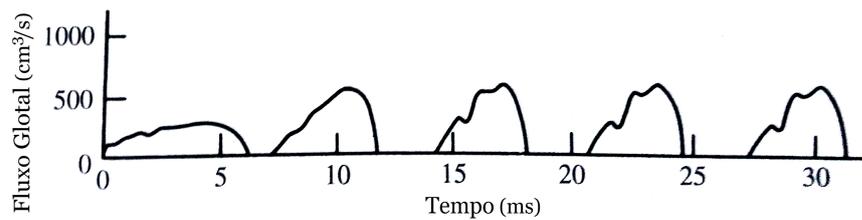


Figura 3.1: Forma de onda da fonte fluxo glotal [27].

Na forma de onda da Fig. 3.1, observa-se que a inclinação do pulso glotal para a direita é causada por uma ascensão lenta e uma queda abrupta do fluxo, isto significa que o acúmulo de fluxo de ar está atrasado em relação ao deslocamento das cordas vocais.

3.3.1 *Jitter*

As oscilações das cordas vocais não são exatamente periódicas e os pulsos de ar, que compõem o sinal glótico, não têm exatamente a mesma duração de tempo. A pequena flutuação aleatória em cada comprimento do ciclo glótico é chamada de *jitter* e seu estudo é particularmente importante em diferentes áreas relacionadas à geração de voz. Os valores típicos do *jitter* estão entre 0.1% e 1% do período fundamental, para as chamadas vozes normais; isto é, sem presença de patologias [17, 18].

3.4 Pulso glotal

O pulso glotal é o fluxo glotal correspondente a um ciclo glotal completo e é descrito através de uma terminologia, que é referida aos aspectos fisiológicos da fonação (ao movimento vibratório das cordas

vocais): instantes de abertura e fechamento glótico, fase de abertura, fase aberta, fase de fechamento, fase fechada e fase de retorno.

Antes de tratar as diferentes fases do movimento glótico, primeiro é necessário compreender os instantes em que ocorre a abertura e o fechamento glótico, assim como o instante quando ocorre o máximo fluxo glotal, de um ponto de vista fisiológico.

Para começar, o instante de abertura glótica se define pelo instante de aumento de fluxo glótico em comparação com seu valor mínimo. Fisiologicamente, esse corresponde ao instante onde as cordas vocais começam a separar-se desde a sua parte superior. O instante de fechamento glótico está associado ao máximo de excitação glótica, isto é, o momento em que a derivada do fluxo glótico atinge seu valor mínimo. Fisiologicamente, esse corresponde ao instante onde as cordas vocais se unem desde a parte inferior.

Usa-se o termo de fechamento abrupto, se a derivada do fluxo glótico é descontínua no instante de fechamento glótico [19].

Fase de abertura A fase de abertura é definida entre o instante de abertura glótica e o instante de máximo fluxo.

Fase de fechamento A fase de fechamento é definida entre o instante de máximo de fluxo e o instante de fechamento glótico.

Fase aberta A fase aberta está composta pelas fases de abertura e fechamento, ou seja, é definida entre o instante de abertura e o instante de fechamento glótico.

Fase fechada A fase fechada é definida entre os instantes de fechamento e de abertura glótica do ciclo seguinte.

As fases glóticas são ilustradas na Fig. 3.2.

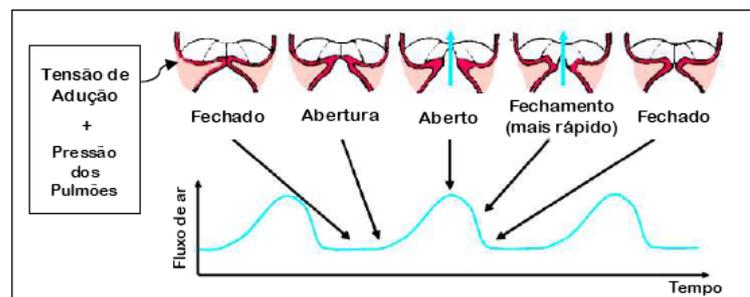


Figura 3.2: Fases do movimento glótico [3].

3.5 Modelos do pulso glotal

Muitos procedimentos médicos usados para o diagnóstico e a análise das oscilações das cordas vocais são do tipo invasivos, que causam desconforto ao falante e interferem com o comportamento de fonação normal, no entanto, vários modelos glóticos têm sido propostos para a análise ou diagnóstico de maneira não invasiva a partir da estimativa do fluxo glótico de um sinal de fala (chamado de filtragem inversa).

Segundo [3], a maioria dos modelos do pulso glotal referidos em inúmeras pesquisas foram projetados a partir das seguintes suposições:

- O fluxo glótico é sempre positivo ou nulo.
- O fluxo glótico é quase-periódico.
- O fluxo glótico é uma função contínua do tempo.
- Em um único período, o fluxo glotal é em forma de sino: ele começa a aumentar, depois diminui e, finalmente, é nulo.
- O fluxo glótico é uma função diferenciável no tempo, exceto em algumas instantes como o instante de fechamento glótico (GCI) e o instante de abertura glótica (GOI).
- A fase de abertura glótica é mais longa que a fase de fechamento glotal, como é ilustrado na Fig. 3.3.

Apesar de a maioria dos modelos do pulso glótico basear-se nas suposições anteriores, é importante indicar que devido a um vazamento glótico durante a fase fechada não é possível considerar o fluxo glótico como nulo durante esta fase.

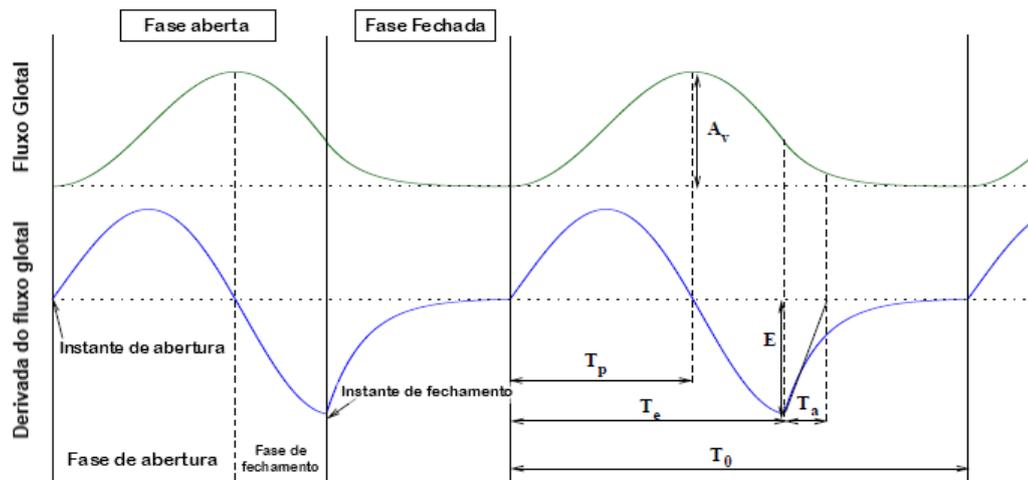


Figura 3.3: Terminologia usada para descrever a fonte glótica [19].

A Fig. 3.3 ilustra a terminologia usada para descrever o pulso glótico. Na parte superior da figura, observa-se a forma de onda do fluxo glotal e, na parte inferior, a forma de onda de sua derivada. T_0 é o período fundamental de um pulso glótico, T_e representa o tempo de duração da fase de aberta, T_p é o tempo de duração da fase de abertura, T_a significa o tempo de duração efetiva da fase de retorno, E corresponde à velocidade de fechamento e A_v é a amplitude de vozeamento.

Todos esses parâmetros são os mais usados entre os modelos glóticos, ainda que, em alguns casos, não têm o mesmo nome e não usam o mesmo número de parâmetros.

3.5.1 Modelo glotal de Rosenberg

Criar um modelo na forma básica do pulso glótico, variando os tempos relativos de abertura e fechamento glotal, foi um dos experimentos feitos por A.E. Rosenberg [20] para determinar o efeito de sua variação na qualidade vocal. Ele usou a técnica de filtragem inversa para extrair da forma de onda glotal três parâmetros, e modelou, com eles, seis formas do pulso glótico através de seis expressões matemáticas diferentes. A forma do pulso glótico mais conhecido de Rosenberg é expressada por:

$$g(t) = \begin{cases} 0.5A_v[1 - \cos(\pi t/(T_P))] & , 0 \leq t \leq T_P \\ A_v \cos(\pi(t - T_P)/(2T_N)) & , T_P < t \leq T_P + T_N \\ 0 & , T_P + T_N < t \leq T_0, \end{cases} \quad (3.1)$$

este modelo trigonométrico possui duas funções para representar as fases de abertura e fechamento do fluxo glótico, onde A_v é uma constante relacionada à amplitude do pulso glotal, T_P (tempo de abertura) e T_N (tempo de fechamento) são parâmetros relacionados à porção do pulso com inclinação positiva e à porção do pulso com inclinação negativa, respectivamente. A forma de onda do pulso glotal e a sua derivada estão ilustradas na Fig. 3.4.

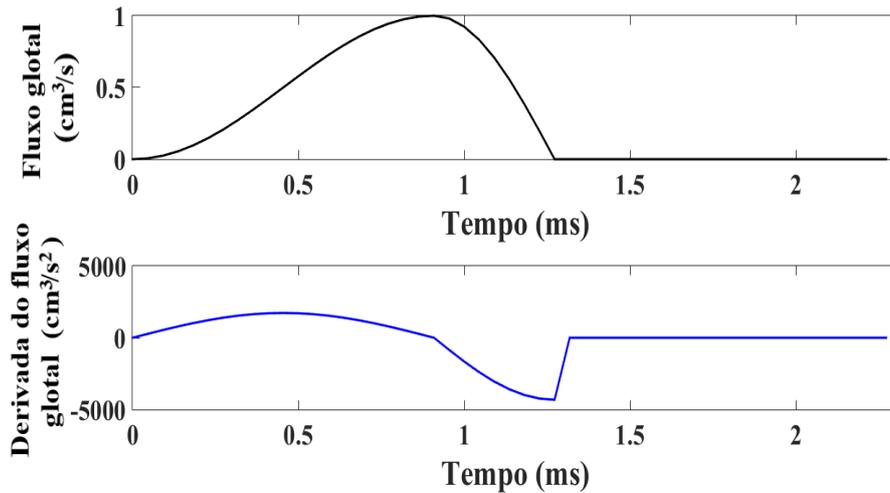


Figura 3.4: Pulso glotal do modelo de Rosenberg e sua derivada: $A_v = 1$, $F_0 = 440 \text{ Hz}$, $\alpha_1 = 40\%$, $\alpha_2 = 16\%$.

O parâmetro $\alpha_1 = T_P/T_0$ corresponde ao tempo relativo de abertura, e $\alpha_2 = T_N/T_0$ ao tempo relativo de fechamento. Eles são mantidos fixos durante a síntese do pulso, e seus valores são dados em porcentagem. Por outro lado, pode-se observar que este modelo é muito flexível, e através de outras expressões matemáticas do mesmo modelo, se pode obter as diferentes formas do pulso glótico que se ilustram na Fig. 3.5.

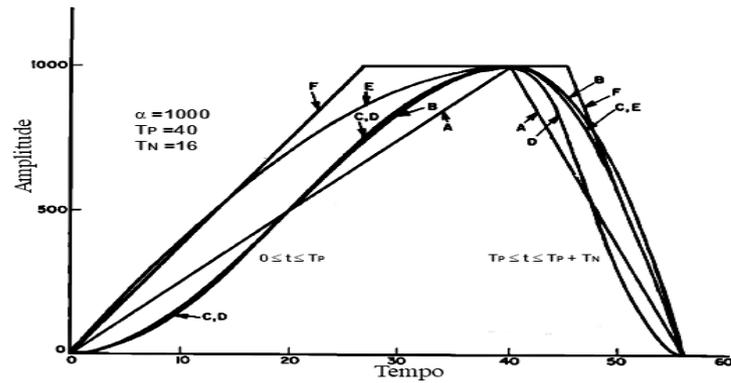


Figura 3.5: Formas de pulso glotal do Modelo de Rosenberg [20].

A forma de pulso glotal influencia algumas características que percebemos na voz do falante; por exemplo, um tom mais grave ou mais agudo é obtido diminuindo ou aumentando a frequência do pulso glótico. As mudanças na forma de onda que representam a velocidade volumétrica glotal produzem alguns tipos de voz bem identificados.

3.5.2 Modelo glotal de Fant

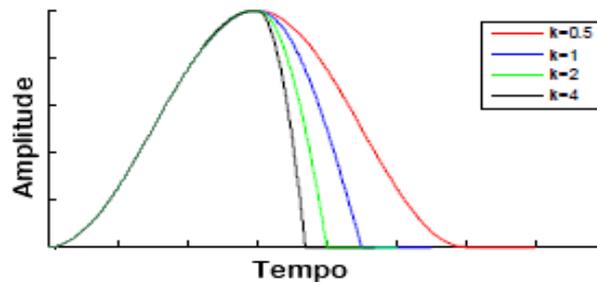


Figura 3.6: Modelo do pulso glótico adotado por Fant [3].

O modelo do pulso glotal empregado pelo G. Fant (1979) [22] é traçado na Fig. 3.6. Além da frequência fundamental da voz F_0 , seus parâmetros básicos são: o fluxo máximo U_0 , a frequência glotal $F_g = \frac{W_g}{2\pi}$ e o fator de assimetria K . O pulso pode ser atribuído a um ponto de partida $t = T_1$. Este pulso é descrito por um ramo ascendente através da seguinte expressão:

$$U = 0.5 U_0 [1 - \cos(W_g (t - T_1))], \quad (3.2)$$

o seu valor máximo U_0 é obtido em $t = T_2$. Temos,

$$T_2 - T_1 = \frac{1}{2F_g}. \quad (3.3)$$

O ramo descendente do pulso é descrito pela seguinte expressão:

$$U = U_0 [K \cos(W_g (t - T_2)) - K + 1], \quad (3.4)$$

onde atinge o zero, após um tempo descrito por:

$$T_3 - T_2 = \frac{1}{W_g} \arccos\left(\frac{K-1}{K}\right), \quad (3.5)$$

para $0.5 < K < 1$, a inclinação máxima ocorre antes do fechamento. Para $K = 0.5$, o ramo descendente é simétrico em relação ao ramo ascendente, e para $K > 1$, tem-se a inclinação máxima durante o curso da queda.

3.5.3 Modelo glotal de Liljencrants-Fant

Liljencrants e Fant sugeriram, em 1985, um modelo para a derivada do fluxo glótico, chamado modelo LF [24]. Este é um modelo de cinco parâmetros no domínio do tempo que, juntamente com o comprimento do ciclo glotal, determina unicamente a forma do pulso.

- E_e : Valor mínimo da derivada do fluxo glótico (máximo de excitação).
- T_0 : Período fundamental.
- T_e : Instante de excitação máxima.
- T_p : Instante do máximo do fluxo glotal.
- T_a : Constante do tempo da fase de retorno.

Esse modelo é construído por uma parte senoidal modulada por uma exponencial crescente (entre 0 e T_e) seguido de uma fase de retorno exponencial decrescente (entre T_e e T_0) [25]:

$$U'_g(t) = \begin{cases} -Ee[e^{a(t-T_e)} \frac{\sin(\frac{\pi t}{Tp})}{\sin(\frac{\pi T_e}{Tp})}] & , 0 \leq t \leq T_e \\ \frac{-Ee}{\varepsilon T_a} (e^{-\varepsilon(t-T_e)} - e^{-\varepsilon(T_0-T_e)}) & , T_e \leq t \leq T_0, \end{cases} \quad (3.6)$$

o parâmetro ε é calculado a partir da condição de continuidade da derivada do fluxo no instante de fechamento glótico.

$$U'_g(t = T_e^-) = U'_g(t = T_e^+), \quad (3.7)$$

que dá como resultado a seguinte equação implícita:

$$\varepsilon T_a = 1 - e^{-\varepsilon(T_0-T_e)}, \quad (3.8)$$

o parâmetro a é calculado a partir da condição de continuidade do fluxo no instante de fechamento glótico.

$$U_g(t = T_e^-) = U_g(t = T_e^+), \quad (3.9)$$

a partir desta condição de continuidade, surge a seguinte equação implícita:

$$\frac{1}{a^2 + (\frac{\pi}{Tp})^2} (e^{-aT_e} \frac{\pi/Tp}{\sin(\pi T_e/Tp)} + a - \frac{\pi}{Tp} \cotg(\pi T_e/Tp)) = \frac{T_0 - T_e}{e^{\varepsilon(T_0-T_e)} - 1} - \frac{1}{\varepsilon} \quad (3.10)$$

3.6 Unificação dos modelos temporais da fonte glotal

O fato de os diferentes modelos de fluxo glótico de domínio do tempo proposto para análise e síntese de fala não usarem o mesmo número de parâmetros, ou o mesmo nome para parâmetros semelhantes, torna difícil comparar as vantagens e desvantagens entre eles. Boris Doval e Christophe d'Alessandro [26] propuseram, entretanto, uma estrutura unificada para estudar as propriedades dos domínios de tempo e frequência dos modelos de fluxo glótico, onde conseguiram mostrar que todos os modelos podem ser representados por um conjunto comum de cinco parâmetros no domínio do tempo. Os parâmetros foram classificados da seguinte maneira: três parâmetros de escala (período fundamental, amplitude de vozeamento, quociente de abertura) um parâmetro de forma (coeficiente de assimetria), e um parâmetro de continuidade de fechamento (Constante de tempo de fase de retorno). Dessa forma, os parâmetros dos modelos do fluxo glotal podem ser reescritos a partir dos cinco parâmetros seguintes:

1. **Amplitude de vozeamento** A_v : O parâmetro é definido como a diferença entre o valor mínimo e o valor máximo do fluxo glótico.
2. **Velocidade de Fechamento** E_e : Este parâmetro é definido como o valor mínimo da derivada do fluxo glótico e corresponde à velocidade do fluxo no instante de fechamento.
3. **Quociente de abertura** O_q : É definido como a relação entre a duração da fase aberta e o período fundamental. Em teoria, esse parâmetro pode ter valores entre 0 (sem abertura) e 1 (sem fechamento ou um fechamento incompleto), mas, na prática, ele pode variar entre 0.3 e 0.98. Este parâmetro pode ser expresso como: $O_q = \frac{T_e}{T_0}$.
4. **Coefficiente de assimetria** α_m : Número sem dimensão que define o instante de máximo da onda de fluxo glótico relativo a O_q e T_0 ; varia entre 0.5 (forma simétrica) e 1 (caso limite de uma forma muito assimétrica). Este parâmetro pode ser expresso como $\alpha_m = \frac{T_p}{O_q T_0}$.
5. **Constante de tempo de fase de retorno** Ta : O parâmetro corresponde à diferença entre o instante do máximo de excitação e o instante onde o fluxo é quase o seu valor mínimo. O valor deste parâmetro é nulo em caso de um fechamento abrupto.

Todos os parâmetros anteriormente descritos são ilustrados na Fig. 3.7.

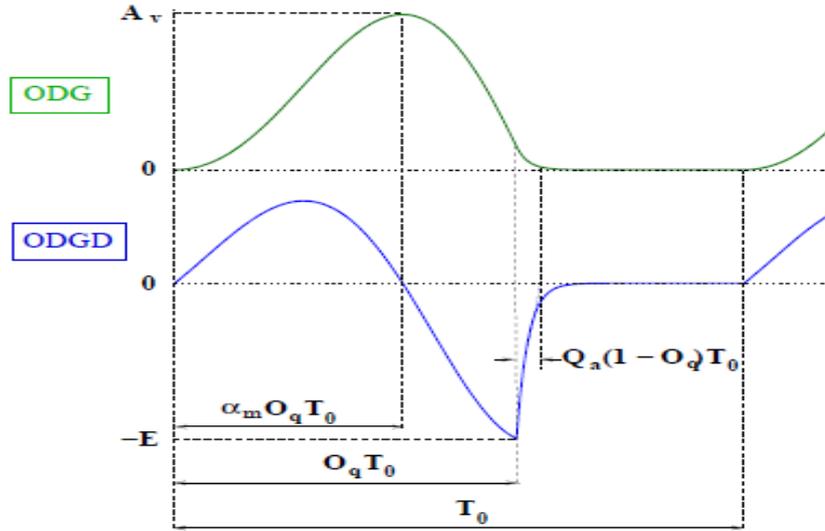


Figura 3.7: Parâmetros unificados da fonte glótica [19].

3.6.1 Modelo glotal de Rosenberg com parâmetros unificados

A expressão da Eq. 3.11 corresponde ao modelo de pulso glotal de Rosenberg com parâmetros unificados equivalente à expressão da Eq. 3.1. Os parâmetros no domínio do tempo são obtidos a partir dos parâmetros originais como: $O_q = (T_P + T_N)/T_0$ e $\alpha_m = T_P/(T_P + T_N)$.

$$g(t) = \begin{cases} 0.5A_v[1 - \cos(\pi \frac{t}{\alpha_m O_q T_0})] & , 0 \leq t \leq \alpha_m O_q T_0 \\ A_v \cos(\frac{\pi(t - \alpha_m O_q T_0)}{2O_q T_0(1 - \alpha_m)}) & , \alpha_m O_q T_0 < t \leq O_q T_0 \\ 0 & , O_q T_0 < t \leq T_0 \end{cases} \quad (3.11)$$

3.6.2 Modelo glotal de Liljencrants-Fant com parâmetros unificados

Este modelo é composto por uma parte senoidal modulada por uma exponencial ascendente (entre 0 e $O_q T_0$), seguida de uma fase de retorno exponencial decrescente (entre $O_q T_0$ e T_0). O modelo é muito versátil para criar a onda do pulso glotal com a maior naturalidade possível como vantagem.

$$U_g(t) = \begin{cases} -\frac{E_e e^{-a O_q T_0}}{\sin(\frac{\pi}{\alpha_m})(a^2 + (\frac{\pi}{\alpha_m O_q T_0})^2)} (\frac{\pi}{\alpha_m O_q T_0} + a e^{at} \sin(\frac{\pi}{\alpha_m O_q T_0} t) - \frac{\pi}{\alpha_m O_q T_0} e^{at} \cos(\frac{\pi}{\alpha_m O_q T_0} t)) & , 0 \leq t \leq O_q T_0 \\ -E_e (\frac{1}{\varepsilon Q_a (1 - O_q) T_0} - 1) (T_0 - t + \frac{1 - e^{\varepsilon(T_0 - t)}}{\varepsilon}) & , O_q T_0 \leq t \leq T_0. \end{cases} \quad (3.12)$$

A expressão temporal da Eq. 3.12 corresponde ao modelo do pulso glotal de Liljencrants-Fant com parâmetros unificados no caso de um fechamento não abrupto ($Q_a > 0$). Este parâmetro tem relação com a duração efetiva da fase de retorno através de $T_a = Q_a(1 - O_q T_0)$ e é conhecido como o coeficiente da fase de retorno. Esta é uma das características que diferencia o modelo de Rosenberg do modelo LF. O parâmetro O_q é o quociente aberto e é definido como instante de fechamento glótico ($t = O_q T_0$) relativo ao T_0 . O período fundamental é T_0 , e α_m é o coeficiente de assimetria definido como o instante de máximo do fluxo glotal ($T_m = \alpha_m O_q T_0$), relativo a T_0 e O_q , a faixa de valores deste coeficiente está entre 0 e 1, mas, devido à fase de abertura glotal ser sempre maior do que a fase de fechamento glotal, a faixa de valores fica restrita entre 0.5 e 1. Por último, o parâmetro E_e representa a máxima amplitude de excitação da derivada do fluxo glótico.

Capítulo 4

Modelagem do trato vocal humano

4.1 Introdução

O som gerado pela fonte acústica, devido às oscilações das cordas vocais, propaga-se no ar contido no trato vocal e é radiado pelos lábios através de uma onda de pressão sonora. Esse som apresenta características que podem ser representadas ao modelar o trato vocal como um tubo que é excitado por uma fonte de som. O tubo com características próprias do trato vocal humano tais como paredes moles, fricção viscosa, condução térmica e radiação nos lábios, pode ser representado, no domínio da frequência, pelas frequências de ressonância e pelas correspondentes larguras de banda. As frequências de ressonância determinam a qualidade da vogal e dão grandes contribuições ao timbre de voz pessoal, enquanto que as larguras de banda representam as perdas no trato vocal. No trato vocal humano há quatro ou cinco frequências de ressonância de interesse: as duas frequências mais baixas determinam em maior parte a vogal, a terceira, quarta e quinta apresentam maior significado no timbre da voz pessoal, que ao ficar uma próxima da outra geram o formante do cantor como característica mais relevante que diferencia a **voz falada** da **voz cantada**.

4.2 Geração e propagação do som

O som pode ser definido, como um distúrbio do equilíbrio de densidade ou pressão de um gás, líquido ou sólido e, de acordo com a psicofísica, como apenas a percepção do som.

O som é audível se possui frequências na faixa de 20 Hz a 20.000 Hz . Se a frequência é menor que 20 Hz , é denominado infrassônico, e se a frequência é maior que 20.000 Hz , é chamando ultra-sônico.

O som é gerado através de ondas de pressão acústica criadas pela vibração de partículas do meio onde se propagam. O som necessita de um meio material de propagação. Estas ondas obedecem às leis da Física, que são a base para descrever a geração e a propagação do som no sistema vocal.

As leis consideram o fluido compressível e de baixa viscosidade (ar), que é o meio para propagação de som na fala. Na medida em que o som avança, sua energia vai sendo atenuada, pois os meios materiais não são perfeitamente elásticos. A velocidade de propagação do som no ar à temperatura ambiente é igual

a 340 m/s ; na água é igual a 1400 m/s e nos sólidos as velocidades são maiores.

A maioria dos sons da natureza são complexos, pois possuem uma frequência fundamental e outras, múltiplas dessa, conhecidas como harmônicos. A frequência fundamental varia desde, aproximadamente, 60 Hz (*Si1*) na voz mais grave do cantor masculino, tendo registrado os 2000 Hz (*Si6*) na voz mais aguda da voz feminina.

Na harmonia musical, uma oitava corresponde a um som (nota) cuja frequência obedece exatamente ao dobro da frequência com a que se compara, ou seja, *igual nota, embora mais aguda*. Por exemplo, se a nota emitida pelo cantor é Ré2 (73 Hz), então sua oitava corresponde a Ré3 (146 Hz).

4.3 Propagação do som no trato vocal

Uma detalhada teoria acústica deve considerar os seguintes efeitos:

1. Variação no tempo e espaço da forma do trato vocal.
2. Perdas devido à condução de calor e fricção viscosa nas paredes do trato vocal.
3. Suavidade das paredes do trato vocal.
4. Radiação do som nos lábios.
5. Acoplamento nasal.
6. Excitação do som no trato vocal

Embora uma teoria que envolve os efeitos mencionados anteriormente não seja geralmente disponível, a propagação do som no trato vocal humano é modelado sob suposições muito simples sobre a forma do trato vocal e as perdas de energia no sistema vocal. Apesar disso, certos filtros digitais têm notável sucesso simulando a transmissão de som no sistema vocal humano.

O trato vocal é modelado como um tubo de área de seção transversal não uniforme, variável no tempo, que corresponde à configuração mais simples que dá uma interpretação importante em relação ao processo de produção da fala, que é o resultado das variações de fluxo de ar nesse sistema. Um esquema do trato vocal é ilustrado na Fig. 4.1.

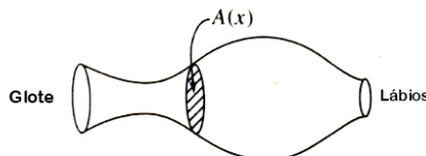


Figura 4.1: Esquema do trato vocal [27].

O trato vocal tem 4 ou 5 frequências de ressonância importantes, denominadas formantes, que dependem da adaptação das partes móveis do sistema da ressonância articulatória. O trato vocal se contrai e se dilata em várias formas mais ou menos complexas, variando as frequências formantes em cada som emitido.

É importante lembrar que o tom glotal pode ser representado através de um espectro de forma triangular onde o seu primeiro harmônico ou tom glotal é acompanhado pelos harmônicos respectivos, que são obtidos multiplicando o valor da referida frequência fundamental por 2, 3, etc, diminuindo a intensidade em 12 dB por oitava.

Todos os harmônicos da fonte glotal passam pelo trato com mais ou menos sucesso. Quanto mais próxima a frequência estiver de uma frequência formante, maior será a amplitude e, também, a qualidade de som da emissão.

As cavidades supraglóticas (faríngea, oral, nasal), além de atuar como cavidades de ressonância, também podem gerar som, por exemplo, na emissão de certas consoantes.

Na articulação de vogais (ou sons vozeados), o trato apresenta uma configuração relativamente aberta e a fonte de excitação é sempre glotal, ou seja, gerada pela vibração das cordas vocais. Nos sons das consoantes, a fonte de energia acústica pode ser uma fonte de ruído turbulento ou transitório, embora alguns também possam possuir intervenção glótica ($m, n, l, b, \text{etc.}$).

Ao falar ou cantar, as frequências próprias das diferentes acomodações do trato vocal se fundem com o som emitido pelas cordas vocais, resultando nos sons de vogais produzidos pela excitação glótica.

Usando as leis da conservação da massa, momento, e energia, e as seguintes suposições simplificadoras da teoria acústica:

1. Para frequências correspondentes a comprimentos de onda que são longos comparados com as dimensões do trato vocal, assume-se a propagação de onda plana ao longo do eixo do tubo.
2. Não há perdas devido à viscosidade ou condução térmica tanto na maior parte do fluido quanto nas paredes do tubo.

O comportamento das ondas do som no trato vocal obedecem a um par de equações diferenciais:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial(u/A)}{\partial t}, \quad (4.1)$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial(\rho A)}{\partial t} + \frac{\partial A}{\partial t}, \quad (4.2)$$

onde:

$p = p(x, t)$ é a variação na pressão do som no tubo na posição x e o tempo t .

$u = u(x, t)$ é a derivada do fluxo glotal na posição x e o tempo t .

ρ = densidade do ar no tubo.

c = a velocidade do som (aproximadamente 35.000 cm/s).

$A = A(x, t)$ = é a função área do tubo, ou seja, o valor da área da secção transversal normal ao eixo do tubo, como uma função da distancia ao longo do tubo e do tempo.

Para que tais equações sejam resolvidas, deve-se encontrar a pressão e, também, o fluxo glotal, para valores de x e t na região limitada pela glote e os lábios. Para conseguir essa solução, as seguintes condições de contorno devem ser estabelecidas em cada extremidade do tubo:

1. No final do lábio, a condição de contorno deve explicar os efeitos da radiação sonora.
2. Na glote, a condição de contorno é imposta pela natureza da excitação.

Adicionalmente às condições de contorno descritas anteriormente, a área do trato vocal $A(x, t)$, deve ser conhecida para todo x , $0 \leq x \leq l$, e para todo t . Ainda que medições detalhadas da área sejam difíceis de obter, consideram-se valores aproximados. Por exemplo, para sons contínuos (sustentados), assume-se que $A(x, t)$ não muda com o tempo.

E, para obter apenas uma visão da natureza do sinal da fala, uma variedade de aproximações e simplificações podem ser tratadas para fazer a solução factível.

4.4 Tubo uniforme sem perdas

Considera-se um modelo muito simples no qual a função área do trato vocal, $A(x, t)$, é assumida constante (com seção transversal uniforme invariante no tempo).

Neste modelo, assume-se que o tubo é excitado por uma fonte ideal de fluxo glotal, representada por um pistão, que pode ser movido de qualquer maneira desejada, independentemente das variações de pressão no tubo.

Outra grande suposição, que não é possível obter na prática, consiste em que não há variações de pressão no ar, somente no fluxo glotal no extremo aberto do tubo.

Este exemplo é muito útil, porque o enfoque da análise e as características da solução têm muito em comum com os modelos mais realistas. Se $A(x, t) = A$ é uma constante, independente de x e t , então as equações diferenciais Eq. 4.1 e Eq. 4.2 ficam reduzidas à forma:

$$-\frac{\partial p}{\partial x} = \frac{\rho}{A} \frac{\partial u}{\partial t}, \quad (4.3)$$

$$-\frac{\partial u}{\partial x} = \frac{A}{\rho c^2} \frac{\partial p}{\partial t}, \quad (4.4)$$

a solução destas equações tem a forma:

$$u(x, t) = [u^+(t - x/c) - u^-(t + x/c)], \quad (4.5)$$

$$p(x, t) = \frac{\rho c}{A} [u^+(t - x/c) + u^-(t + x/c)], \quad (4.6)$$

onde $u^+(t - x/c)$ e $u^-(t + x/c)$ são interpretadas como ondas que viajam na direção positiva e negativa, respectivamente.

Na teoria das linhas de transmissão elétrica, a voltagem $v(x, t)$ e a corrente $i(x, t)$ em uma linha sem perda uniforme satisfaz às seguintes equações:

$$-\frac{\partial v}{\partial x} = L \frac{\partial i}{\partial t}, \quad (4.7)$$

$$-\frac{\partial i}{\partial x} = C \frac{\partial v}{\partial t}, \quad (4.8)$$

onde L e C são a indutância e a capacitância por unidade de comprimento, respectivamente. Portanto, a teoria da linha de transmissão elétrica uniforme, sem perdas, aplicam-se diretamente ao tubo acústico uniforme de acordo com as analogias mostradas na Tab. 4.1 [27].

Quantidade Acústica	Quantidade Elétrica Análoga
p: pressão	v: voltagem
u: fluxo glotal	i: corrente
ρ/A : indutância acústica	L: indutância
$A/(\rho c^2)$: capacitância acústica	C: capacitância

Tabela 4.1: Analogias entre quantidades acústicas e elétricas.

Porém, por analogia, o tubo acústico uniforme terminado como na Fig. 4.2 (a), comporta-se igual à linha de transmissão elétrica uniforme sem perda que termina em um curto circuito em um extremo ($v(l, t) = 0$) e excitado por uma fonte de corrente no outro extremo ($i(0, t) = i_G(t)$), assim como se observa na Fig. 4.2 (b).

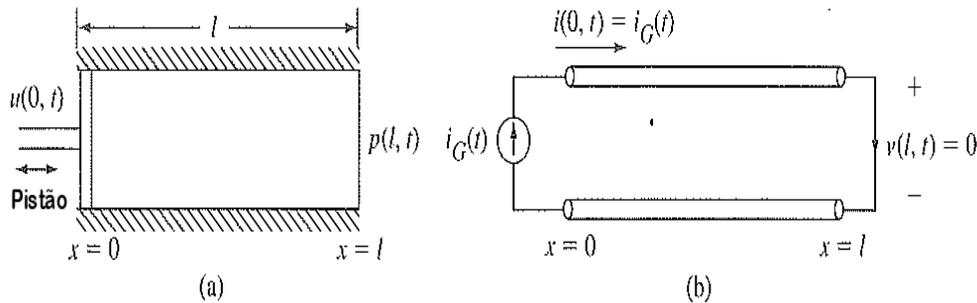


Figura 4.2: (a) Tubo uniforme sem perdas com terminação ideal, (b) Analogia à linha de transmissão elétrica [27].

A representação no domínio da frequência deste modelo é obtida a partir da condição de contorno $x = 0$, então, o fluxo glotal nesse ponto é representada por:

$$u(0, t) = u_G(t) = U_G(\Omega)e^{j\Omega t}, \quad (4.9)$$

onde se observa que o tubo é excitado por uma variação exponencial complexa de fluxo glotal de frequência em radiano, Ω , e amplitude complexa, $u_G(\Omega)$. Visto que as equações Eq. 4.3 e Eq. 4.4 são lineares e invariantes no tempo, as soluções para $u^+(t - x/c)$ e $u^-(t + x/c)$ devem ser da forma:

$$u^+(t - x/c) = K^+ e^{j\Omega(t-x/c)}, \quad (4.10)$$

$$u^-(t + x/c) = K^- e^{j\Omega(t+x/c)}, \quad (4.11)$$

logo, por substituição das expressões anteriores nas equações Eq. 4.5 e Eq. 4.6 e aplicando a condição de

contorno do curto circuito acústico, obtém-se a expressão correspondente para o final do lábio do tubo e a expressão da Eq. 4.9 para o extremo da glote:

$$p(l, t) = 0. \quad (4.12)$$

Os símbolos desconhecidos K^+ e K^- são constantes que podem ser encontradas a partir das seguintes condições de contorno:

$$u(0, t) = U_G(\Omega)e^{j\Omega t} = K^+e^{j\Omega t} - K^-e^{j\Omega t}, \quad (4.13)$$

$$p(l, t) = 0 = \frac{\rho c}{A}[K^+e^{j\Omega(t-l/c)} + K^-e^{j\Omega(t+l/c)}], \quad (4.14)$$

assim, o resultado é:

$$K^+ = U_G(\Omega) \frac{e^{2j\Omega l/c}}{1 + e^{j\Omega 2l/c}}, \quad (4.15)$$

$$K^- = -\frac{U_G(\Omega)}{1 + e^{j\Omega 2l/c}}, \quad (4.16)$$

com consequência, as soluções em estado estacionário senoidal para $u(x, t)$ e $p(x, t)$ são:

$$u(x, t) = \frac{\cos[\Omega(l-x)/c]}{\cos[\Omega l/c]} U_G(\Omega) e^{j\Omega t}, \quad (4.17)$$

$$p(x, t) = jZ_0 \frac{\sin[\Omega(l-x)/c]}{\cos[\Omega l/c]} U_G(\Omega) e^{j\Omega t}, \quad (4.18)$$

onde $Z_0 = \frac{\rho c}{A}$ é chamada impedância acústica característica do tubo, que é uma analogia à linha de transmissão elétrica.

O fluxo glotal nos lábios é encontrada da Eq. 4.17 quando $x = l$ da seguinte forma:

$$u(l, t) = \frac{1}{\cos(\Omega l/c)} U_G(\Omega) e^{j\Omega t} = U(l, \Omega) e^{j\Omega t}, \quad (4.19)$$

onde $U(l, \Omega)$ corresponde à amplitude complexa do fluxo glotal no final dos lábios do tubo.

Ao considerar a relação entre a fonte do fluxo glotal (onde $x = 0$) e o fluxo glotal nos lábios (onde $x = l$), obtém-se a seguinte expressão, que representa a resposta em frequência do sistema:

$$V_a(j\Omega) = \frac{U(l, \Omega)}{U_G(\Omega)} = \frac{1}{\cos(\Omega l/c)}. \quad (4.20)$$

Esta resposta em frequência é apresentada na Fig. 4.3 (a), para $l = 17.5 \text{ cm}$ e $c = 35.000 \text{ cm/s}$. Substituindo $j\Omega$ por s , consegue-se a transformada de Laplace da função sistema:

$$V_a(s) = \frac{1}{\cosh(sl/c)} = \frac{2e^{-sl/c}}{1 + e^{-2sl/c}}. \quad (4.21)$$

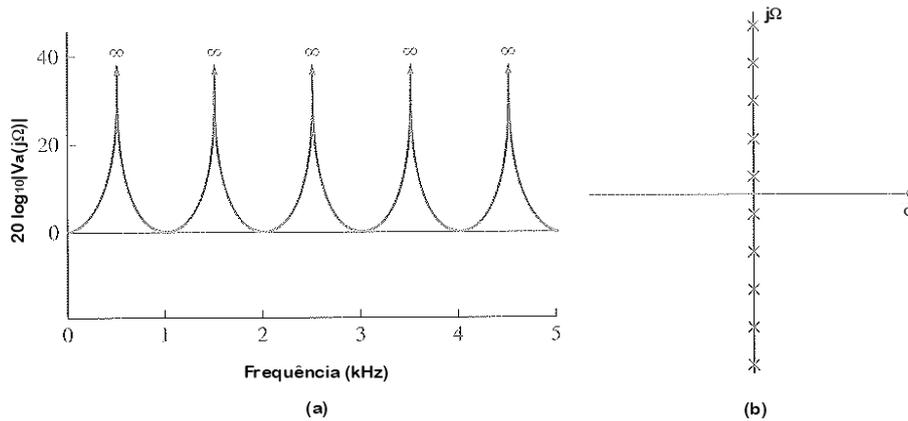


Figura 4.3: (a) Resposta em frequência magnitude Log, (b) Posições dos polos para o tubo uniforme sem perdas [27].

Então, para um tubo acústico sem perdas terminado em um curto circuito acústico e representado pela função do sistema $V_a(s)$, seus polos (valores de s tal que o denominador $(1 + e^{-s2l/c}) = 0$) são um numero infinito, que ficam igualmente espaçados sobre o eixo $j\Omega$ do plano s , como é ilustrado na Fig. 4.3 (b).

Uma interpretação do sistema $V_a(s)$, onde se observa uma rota de transmissão representada pelo fator $e^{sl/c}$ (correspondente ao tempo $\tau = l/c$ segundos, necessários para cruzar uma distancia l a uma velocidade c), é a seguinte:

A fonte do fluxo viaja desde a glote ate os lábios com um atraso de $1/c$ para cada rota do trato vocal; logo o sinal reflete-se nos lábios e move-se no sentido inverso com igual atraso de $1/c$ por cada rota para trás do trato vocal. Esse ciclo de reflexão na glote e nos lábios repete-se infinitamente, portanto, se o trato vocal fosse um tubo uniforme sem perdas de $17,5\text{ cm}$ de comprimento, excitado em um extremo pela fonte glotal $u_G(t)$ e terminado em um curto circuito acústico no final do lábio, então as frequências naturais estariam no conjunto de frequências $F_n = 500(2n + 1)$, $n = 0, 1, \dots$ ou, melhor, as primeiras quatro formantes seriam: $500, 1500, 2500$ e 3500 Hz . Se o trato vocal fosse maior ou menor, essas frequências seriam menores ou superiores, respectivamente. Devido à diferença de sexo no comprimento do trato vocal, uma configuração articulatória similar produz frequências de formantes próximas de 40% , mais altas em crianças do que em homens adultos. As mulheres adultas normalmente têm um trato vocal mais curto do que os homens adultos e, como consequência, elas têm uma frequência de formantes 15% maior do que os homens adultos [28].

Isso nos mostra que o comprimento do trato vocal tem um papel muito importante na determinação dos formantes e, portanto, na função de transferência.

4.5 Efeitos de perda no trato vocal

O modelo examinado anteriormente corresponde ao modelo acústico para a produção da fala, onde se assume a propagação do som no trato vocal sem perda de energia, porém, voltando a examinar

modelos mais realistas, é importante destacar a perda de energia no trato vocal. Essa perda é resultado da fricção viscosa entre o ar e as paredes do tubo, a condução de calor através das paredes do tubo e a vibração das paredes do tubo.

4.5.1 Perdas devido a paredes moles

As variações de pressão de ar dentro do trato vocal causam uma vibração em suas paredes. Se as paredes forem elásticas, a área de secção transversal do tubo muda, dependendo da pressão dentro do tubo. Dado que as variações de pressão são muito pequenas, então a variação resultante na área de secção transversal pode ser tratada como uma pequena perturbação da área nominal. Assim, a nova expressão de área do tubo é a seguinte:

$$A(x, t) = A_0(x, t) + \delta A(x, t), \quad (4.22)$$

onde, $A_0(x, t)$ é a área nominal e $\delta A(x, t)$ é uma pequena perturbação.

Devido ao fato que a massa e a elasticidade das paredes do trato vocal relacionam-se com a perturbação da área, $\delta A(x, t)$, e variações de pressão $p(x, t)$, pode-se construir um modelo da forma:

$$m_w \frac{d^2(\delta A)}{dt^2} + b_w \frac{d(\delta A)}{dt} + k_w(\delta A) = p(x, t), \quad (4.23)$$

onde:

$m_w(x)$ é a massa por unidade de comprimento das paredes do trato vocal;

$b_w(x)$ é a amortecimento por unidade de comprimento das paredes do trato vocal;

$k_w(x)$ é a rigidez por unidade de comprimento das paredes do trato vocal.

Substituindo a Eq. 4.23 nas equações Eq. 4.1 e Eq. 4.2 e omitindo os termos de segunda ordem, nas quantidades u/A e ρA , obtêm-se as seguintes equações diferenciais:

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial(u/A_0)}{\partial t}, \quad (4.24)$$

$$-\frac{\partial u}{\partial x} = \frac{1}{\rho c^2} \frac{\partial(pA_0)}{\partial t} + \frac{\partial A_0}{\partial t} + \frac{\partial(\delta A)}{\partial t}, \quad (4.25)$$

por conseguinte, as equações Eq. 4.22, Eq. 4.23, Eq. 4.24 e Eq. 4.25 descrevem a propagação do som no tubo de paredes moles, que devem ser resolvidas simultaneamente para $p(x, t)$ e $u(x, t)$ dada $A_0(x, t)$ e as condições de contorno apropriadas. Considera-se que este modelo seria o apropriado para um som contínuo como uma vogal [27].

Resolvendo numericamente as equações diferenciais e usando as condições de contorno na glote $u(0, t) = U_G(\Omega)e^{j\Omega t}$ e no final dos lábios $p(l, t) = 0$, curto circuito acústico, obtêm-se a seguinte resposta em frequência:

$$V_a(j\Omega) = \frac{U(l, \Omega)}{U_G(\Omega)}. \quad (4.26)$$

Se o trato vocal fosse um tubo uniforme de paredes moles com uma área de seção transversal de 5 cm^2 e 17.5 cm de comprimento, excitado em um extremo pela fonte glotal $u_G(t)$ e terminado em um curto circuito acústico no final do lábio, então suas frequências de ressonância não estão mais exatamente no eixo $j\Omega$ do plano s , como no caso do modelo de tubo uniforme sem perdas. Este é o motivo porque a resposta em frequência não é mais infinita a frequências de 500 Hz , 1500 Hz , 2500 Hz , etc, embora a resposta alcance seu pico na proximidade dessas frequências.

Na Fig. 4.4, observam-se as frequências de ressonância $F(k)$ e as largura de banda $B(k)$, das ressonâncias (em Hz) da tabela associada.

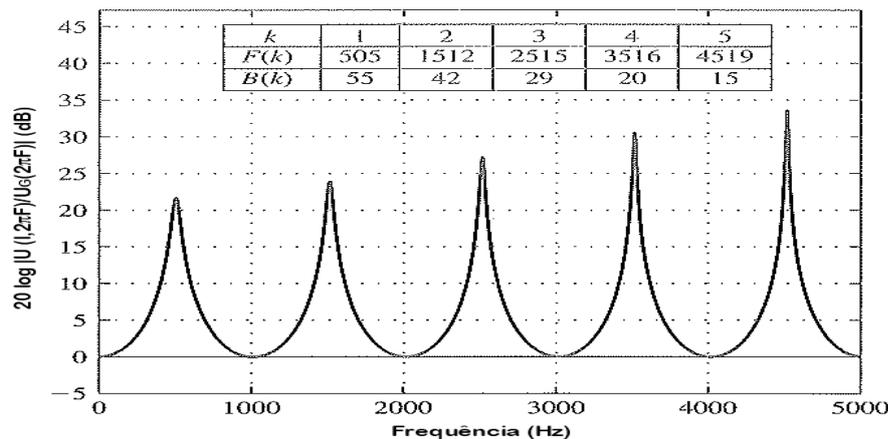


Figura 4.4: Resposta em frequência magnitude Log para o tubo uniforme de paredes moles [27].

Os efeitos mais importantes, no caso do tubo de paredes moles, são as frequências de ressonância ligeiramente superiores às do caso sem perdas, e a largura de banda das ressonâncias não é mais zero como no caso sem perdas, já que o valor máximo não é mais infinito.

A largura de banda de uma ressonância está relacionada com a faixa de frequências em torno da ressonância, na qual a resposta em frequência é maior que $1/\sqrt{2}$ vezes (equivalente a 3 dB) o valor de pico na frequência.

4.5.2 Perdas devido a fricção viscosa e condução térmica

Os efeitos de perda devido à fricção viscosa e de condução de calor através das paredes do trato vocal devem ser considerados na representação no domínio da frequência, incluindo um termo real na expressão da impedância acústica Z e, também, da admitância acústica Y , respectivamente. Estas expressões de admitância e impedância acústica junto com as condições de contorno para uma fonte de fluxo glotal ideal, representada por $u(0, t) = U_G(\Omega)e^{j\Omega t}$, e um curto circuito acústico no final dos lábios, representado por $p(l, t) = 0$, são usadas para resolver as equações diferenciais do modelo.

A resposta em frequência de um tubo uniforme com paredes moles e perdas devido à fricção viscosa e condução térmica é ilustrada na Fig. 4.5, onde as frequências centrais de ressonância e as larguras de banda (em Hz) estão na tabela associada.

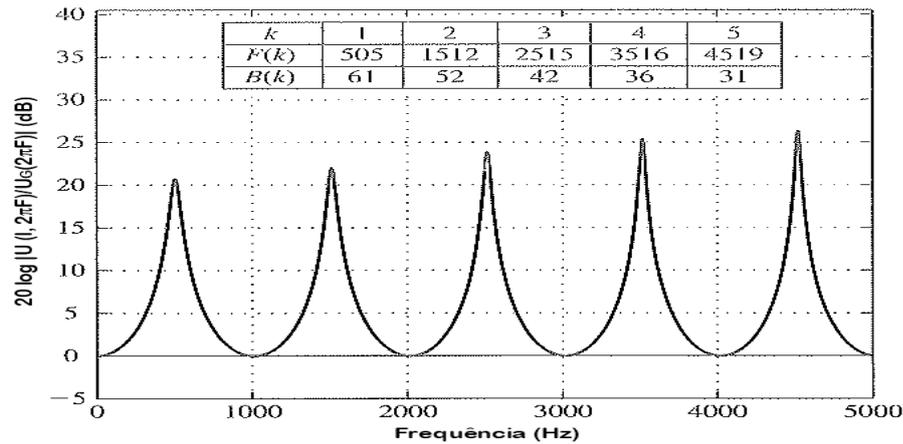


Figura 4.5: Resposta em frequência magnitude Log para o tubo uniforme com paredes moles e perdas por fricção e condução térmica [27].

Comparada com a Fig. 4.4, a Fig. 4.5 mostra que o principal efeito é incrementar as larguras de banda dos formantes pois os efeitos de perda devidos à fricção e condução térmica incrementam em $\Omega^{1/2}$ as frequências de ressonância mais altas, experimentando um maior alargamento do que as ressonâncias mais baixas.

4.5.3 Perdas devido a radiação nos lábios

Os modelos descritos anteriormente apenas consideraram as perdas internas que afetam a transmissão do som no trato vocal, pois assumem a condição de contorno $p(l, t) = 0$ nos lábios, que corresponde a um curto circuito acústico, indicando que não existe abertura nos lábios, porém sem perda devido à radiação. Um modelo mais real, entretanto, considera esta abertura dos lábios (abertura nasal no caso de sons não vozeados) como um orifício em uma esfera (ver Fig. 4.6).

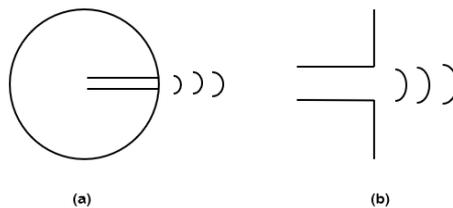


Figura 4.6: (a) Radiação de um defletor esférico, (b) Radiação de um defletor plano infinito [27].

Em baixas frequências, a abertura pode ser considerada uma superfície radiante, com as ondas sonoras irradiadas pelo defletor esférico, que representa a cabeça, embora, os efeitos de difração sejam complexos e difícil de representar. Portanto, se a superfície radiante é modelada por um defletor plano de extensão infinita, a relação de estado estacionário senoidal entre as amplitudes complexas de pressão e o fluxo glotal nos lábios é representada por:

$$P(l, \Omega) = Z_L(\Omega) U(l, \Omega), \quad (4.27)$$

onde a impedância de radiação $Z_L(\Omega)$ nos lábios é, aproximadamente, da forma:

$$Z_L(\Omega) = \frac{j\Omega L_r R_r}{R_r + j\Omega L_r}. \quad (4.28)$$

Os valores R_r e L_r fornecem uma ótima aproximação para o defletor plano infinito

$$R_r = \frac{128}{9\pi^2}, \quad (4.29)$$

$$L_r = \frac{8a}{3\pi c}, \quad (4.30)$$

sendo a o raio da abertura dos lábios $A_0(l) = \pi a^2$ e c é a velocidade do som.

Por outro lado, a partir da expressão Eq. 4.28, pode-se observar que nas frequências muito baixas $Z_L(\Omega) \approx 0$ e o comportamento da impedância de radiação aproxima-se a uma terminação de curto circuito. Em uma faixa intermediária de frequências temos $\Omega L_r \ll R_r$, e $Z_L(\Omega) \approx j\Omega L_r$ e, nas altas frequências $\Omega L_r \gg R_r$, $Z_L(\Omega) \approx R_r$.

Capítulo 5

Resultados

5.1 Introdução

Neste capítulo, apresenta-se, detalhadamente, o algoritmo usado para sintetizar, ou criar de maneira artificial, os fonemas vocálicos do espanhol entoados por um cantor nas diferentes notas da escala musical. O código recebe o fonema vocálico e as frequências de todas as notas que compõem a extensão vocal do cantor. O algoritmo usado para a síntese da voz foi implementado em Matlab, e pode-se dividir de acordo com os componentes do modelo fonte-filtro. Para a modelagem do fluxo glotal (fonte de excitação) o qual representa o sinal glotal, foram usados dois modelos glotais: o modelo de Rosenberg e o modelo de Liljencrants-Fant (LF), segundo a representação dos parâmetros unificados propostos por Boris Doval e Christophe d’Alessandro [25]. Os dois modelos glotais escolhidos apresentam vantagens e desvantagens e permitem gerar diferentes tipos de voz, tentando gerar um sinal mais próximo ao que é produzido no trato vocal.

O trato vocal é modelado por um filtro digital, cuja função de transferência se caracteriza por ter apenas polos que correspondem às frequências de ressonância (formantes) da fala vozeada, no entanto, nos fonemas nasais e fricativos são necessárias ressonâncias e antirressonâncias (polos e zeros). O algoritmo usado para representar os efeitos produzidos pelo sistema do trato vocal baseia-se na expressão da resposta em frequência do filtro digital, que considera as frequências formantes da fala e os efeitos de perdas por paredes moles, fricção e condução térmica e a radiação nos lábios do trato vocal, conhecidas como larguras de banda.

A finalidade é analisar os diferentes tipos de sons produzidos ao sintetizar os fonemas vocálicos entoados nas diferentes notas musicais; perceber a naturalidade das vozes que se tenta gerar com cada modelo glotal utilizado, e os diversos tipos de voz produzidos com as mudanças em seus parâmetros.

5.2 Síntese da voz

Todo modelo de síntese de voz inclui um bloco que representa a fonte de excitação do trato vocal, embora na implementação sintética, a fonte possa ser gerada por qualquer trem de pulsos. Se quisermos

obter maior naturalidade da voz, é melhor que o trem de pulsos seja o mais próximo possível do que é produzido na glote. Para os sons vocalizados, existem diversos modelos que geram o pulso glotal.

Neste trabalho, com os dois modelos usados, as metas a serem alcançadas são: apresentar as características gerais da forma de onda do pulso glotal com modelos diferentes, criar diferentes tipos de voz ao variar os parâmetros do pulso glotal, gerar as diferentes vogais do espanhol cantadas por cantores com diferente classificação vocal e realizar a síntese das diferentes vogais do espanhol cantadas nas diferentes notas musicais com *jitter*.

Os códigos para a síntese foram implementados em Matlab, com a vantagem de poder variar os parâmetros e, portanto, gerar diferentes sons e figuras correspondentes.

5.3 Síntese do pulso glotal com o modelo de Rosenberg

Como foi mencionado anteriormente, um dos modelos usados neste trabalho é conhecido como o modelo de Rosenberg, que considera o pulso glotal chamado de $g(t)$ dado pela Eq. 3.1.

Com o uso do código em Matlab (apêndice A) e variando os parâmetros do modelo glotal de Rosenberg, tais como: amplitude máxima do fluxo glótico (A_v), frequência fundamental (F_0), porcentagem do tempo de abertura relativa (α_1) e porcentagem do tempo de fechamento relativo (α_2), foram obtidos diferentes tipos de voz e, com isso, as características principais do fluxo na forma de onda do pulso e a sua derivada correspondente puderam ser observadas. Em seguida, apresentam-se três diferentes tipos da forma de onda do pulso glotal de Rosenberg correspondentes ao fluxo de um ciclo glotal completo e sua derivada. Cada pulso foi sintetizado considerando diferentes parâmetros.

A Fig. 5.1 mostra um pulso glotal com a fase de abertura maior do que a fase de fechamento, de acordo com os parâmetros $\alpha_1 = 40\%$ e $\alpha_2 = 20\%$. A frequência fundamental $F_0 = 262\text{ Hz}$ corresponde a um tom agudo equivalente ao Do4 da escala musical temperada e uma amplitude $A_v = 3$, que tem relação com o volume de fonação.

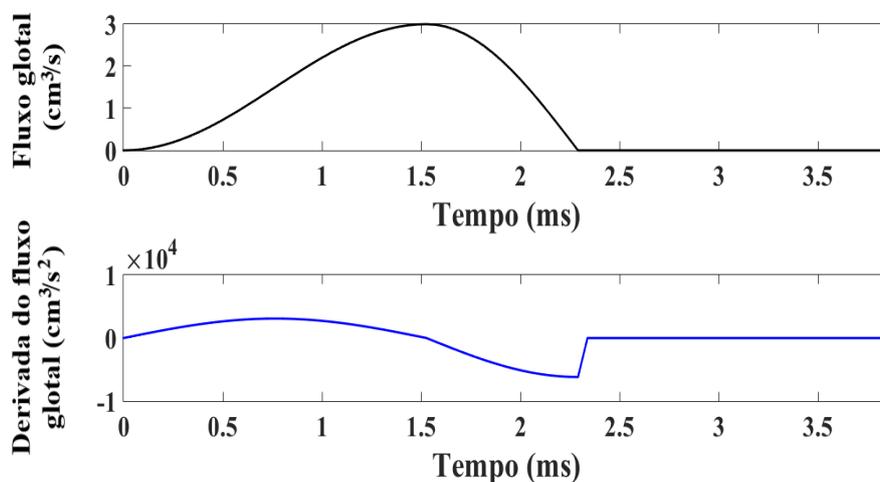


Figura 5.1: Pulso glotal de Rosenberg (em cima) e sua derivada (embaixo) com: $A_v = 3$, $F_0 = 262\text{ Hz}$, $\alpha_1 = 40\%$ e $\alpha_2 = 20\%$.

A Fig. 5.2 apresenta um pulso glotal simétrico com o tempo de abertura igual ao tempo de fechamento, com $\alpha_1 = 50\%$ e $\alpha_2 = 50\%$. A frequência fundamental $F_0 = 262\text{ Hz}$ corresponde a um tom agudo equivalente ao Do4 da escala musical temperada; igual ao tom do pulso anterior e uma amplitude $A_v = 1$, equivalente a um volume de fonação menor que o pulso anterior.

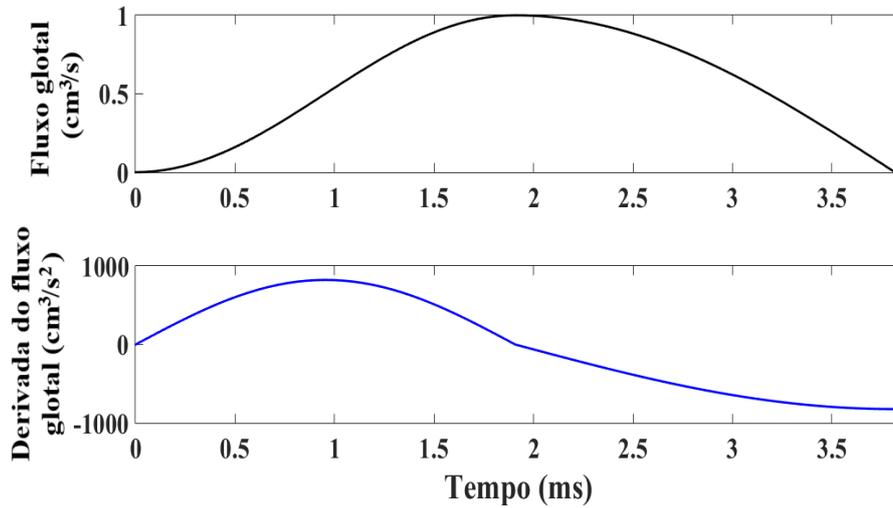


Figura 5.2: Pulso glotal de Rosenberg (em cima) e sua derivada (embaixo) com: $A_v = 1$, $F_0 = 262\text{ Hz}$, $\alpha_1 = 50\%$ e $\alpha_2 = 50\%$.

Na Fig. 5.3, há um pulso glotal de um tipo de voz chamada sussurrante, com uma fase de abertura menor do que a fase de fechamento com $\alpha_1 = 30\%$ e $\alpha_2 = 50\%$. A frequência fundamental $F_0 = 440\text{ Hz}$ corresponde a um tom ainda mais agudo que o anterior equivalente ao La4 da escala musical temperada e uma amplitude $A_v = 1$.

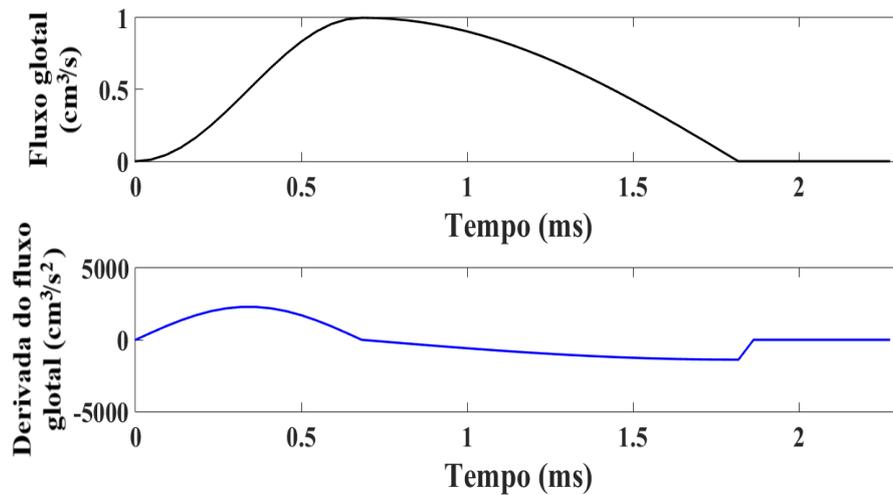


Figura 5.3: Pulso glotal de Rosenberg (em cima) e sua derivada (embaixo) com: $A_v = 1$, $F_0 = 440\text{ Hz}$, $\alpha_1 = 30\%$ e $\alpha_2 = 50\%$.

5.4 Síntese do pulso glotal com o modelo de Liljencrants-Fant

Outro modelo da fonte glotal, usado neste trabalho, é conhecido como modelo de Liljencrants-Fant, com os parâmetros unificados. O pulso glotal é denotado por $U_g(t)$, dado pela Eq. 3.12.

Com o uso do código em Matlab, do apêndice B, e variando os parâmetros do modelo glotal de LF, diferentes formas de onda do pulso para um ciclo glótico completo foram obtidas. Além disso, observam-se as características principais do fluxo na forma de onda do pulso e a sua derivada. Os valores dos parâmetros foram determinados com o propósito de gerar um pulso similar aos casos do modelo de Rosenberg das Figs. 5.1, 5.2 e 5.3. Para isso, usaram-se as expressões de parâmetros unificados que relacionam os parâmetros originais do modelo de Rosenberg (T_P e T_N) com os parâmetros O_q e α_m .

Usando os parâmetros do pulso glotal do modelo de Rosenberg obtido na Fig. 5.1, os parâmetros unificados correspondentes ao modelo LF forma calculados e permitiram gerar um pulso com características semelhantes de modo que a fase de abertura do pulso é maior do que a fase de fechamento, dados pelos parâmetros $O_q = 0.6$, $\alpha_m = \frac{2}{3}$. A amplitude máxima do fluxo glotal é $A_v = 3$, a frequência fundamental é $F_0 = 262 \text{ Hz}$ e representa a nota aguda Do4 ou Do central da escala musical temperada e o coeficiente da fase de retorno está determinado por $Q_a = 0.1$, correspondente a um fechamento não abrupto, assim como se observa na Fig. 5.4.

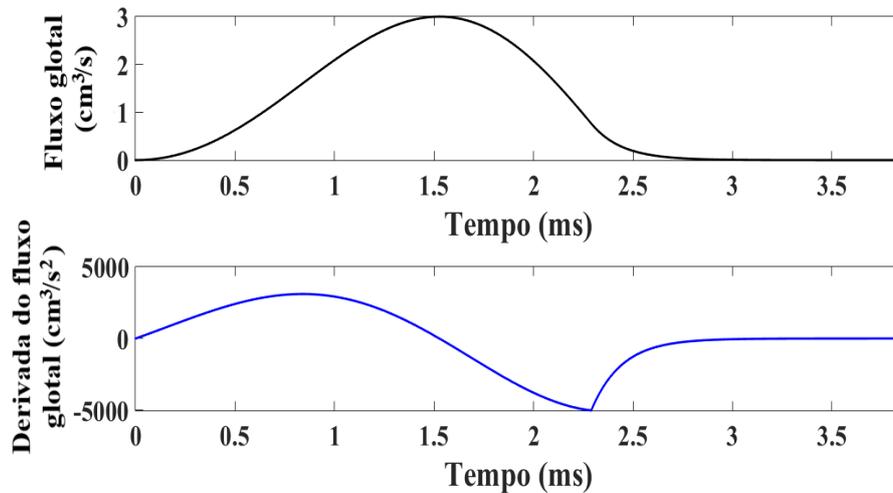


Figura 5.4: Pulso glotal de LF (em cima) e sua derivada (embaixo) com: $A_v = 3$, $F_0 = 262 \text{ Hz}$, $O_q = 0.6$, $\alpha_m = \frac{2}{3}$ e $Q_a = 0.1$.

Para o caso em que a fase de abertura é igual à fase de fechamento, um pulso glotal simétrico foi gerado usando o modelo de LF, a partir dos seguintes parâmetros unificados: $A_v = 1$, $F_0 = 262 \text{ Hz}$, $O_q = 1$, $\alpha_m = 0.5$ e $Q_a = 0.1$. Esta forma do pulso é similar ao pulso produzido através do modelo de Rosenberg da Fig. 5.2 com o mesmo nível de volume de fonação e mesma frequência fundamental. Este ciclo glótico completo apresentado na Fig. 5.5 corresponde apenas à fase aberta do pulso (confinado entre $t = 0$ e $t = O_q T_0$), devido ao valor de $O_q = 1$ (embora no código do apêndice B, o pulso tenha sido gerado com um valor aproximado $O_q = 0.9999$).

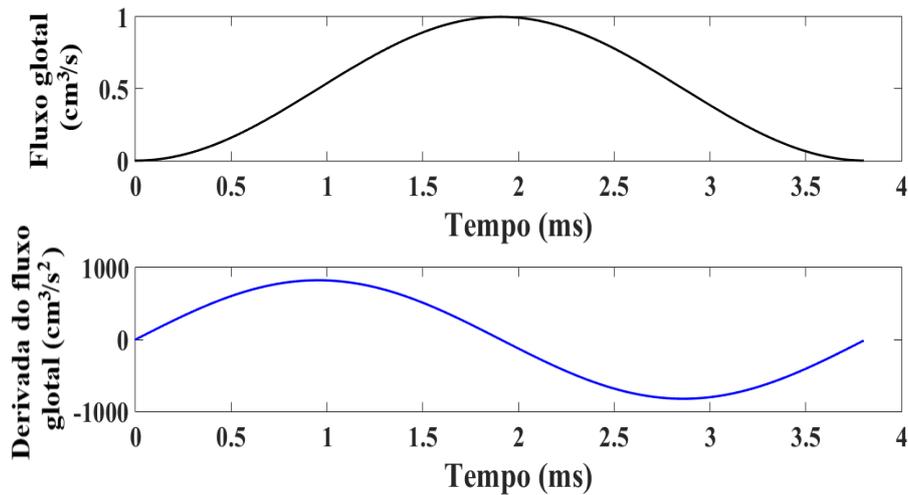


Figura 5.5: Pulso glotal de LF (em cima) e sua derivada (embaixo) com: $A_v = 1$, $F_0 = 262 \text{ Hz}$, $O_q = 1$, $\alpha_m = 0.5$ e $Q_a = 0.1$.

Para o caso em que a fase de abertura do pulso glotal é menor do que a fase de fechamento, os parâmetros unificados do modelo LF correspondentes ao pulso glotal de Rosenberg foram calculados e apresentados na Fig. 5.3. os parâmetros foram os seguintes: $A_v = 1$, $F_0 = 440 \text{ Hz}$, $O_q = 0.8$, $\alpha_m = \frac{3}{8}$ e $Q_a = 0.1$. no entanto, o pulso glotal não foi gerado, dado que o modelo LF considera que a fase de abertura glotal é sempre maior do que a fase de fechamento glotal estabelecendo α_m em um intervalo restrito entre 0.5 e 1.

5.5 Análise de sensibilidade dos parâmetros para o modelo LF

Os parâmetros O_q , α_m e Q_a têm diferentes efeitos no domínio do tempo e da frequência. A Fig. 5.6 e a Fig. 5.7 apresentam o efeito da variação de O_q no domínio do tempo e da frequência respectivamente. Foram mantidos fixos: $A_v = 1$, $F_0 = 262 \text{ Hz}$, $\alpha_m = 2/3$ e $Q_a = 0.1$. Quando o O_q aumenta (0.2, 0.4, 0.6, 0.8), a escala de tempo é expandida porque o O_q está definido como a porção do período fundamental durante o qual as cordas vocais estão abertas. Essa dilatação induz uma compressão no domínio da frequência e, assim, um deslocamento da frequência glotal para as baixas frequências.

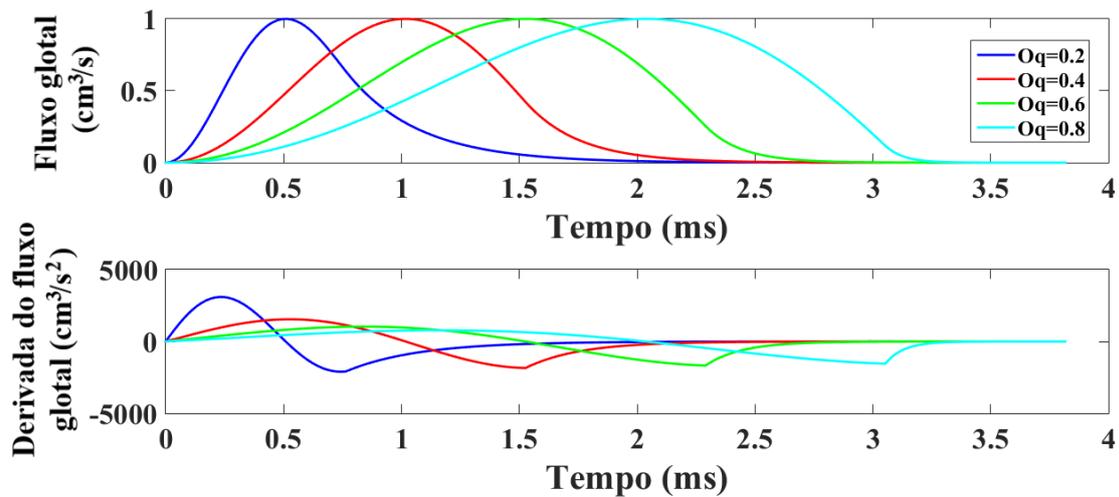


Figura 5.6: Variações temporais do pulso glotal (em cima) e a sua derivada (embaixo) para quatro valores de O_q no caso em que A_v , F_0 , α_m e Q_a sejam mantidos constantes.

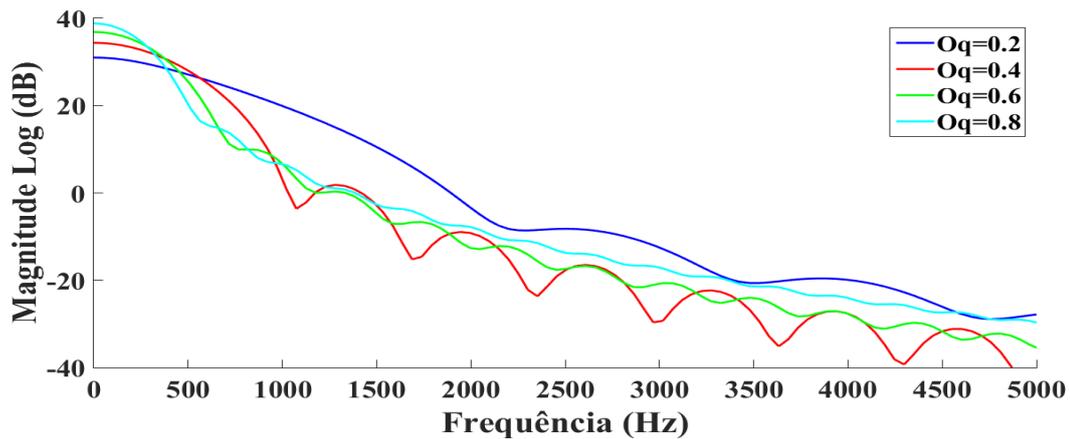


Figura 5.7: Variações espectrais do pulso glotal para quatro valores de O_q no caso em que A_v , F_0 , α_m e Q_a sejam mantidos constantes.

O efeito da variação de α_m está apresentado no domínio do tempo e da frequência na Fig. 5.8 e na Fig. 5.9, respectivamente. Neste caso, os seguintes parâmetros são mantidos constantes: $A_v = 1$, $F_0 = 262 \text{ Hz}$, $O_q = 0.8$ e $Q_a = 0.1$, enquanto que α_m varia. Um incremento no domínio do tempo de α_m induz uma deformação da forma do sino do pulso glótico, enquanto que α_m decrescente induz uma forma mais senoidal da primeira fase da derivada da fonte glótica. No domínio espectral do pulso glotal, α_m crescente induz um deslocamento do formante glotal de uma maneira oposta à da O_q e aumenta a largura de banda do formante glótico.

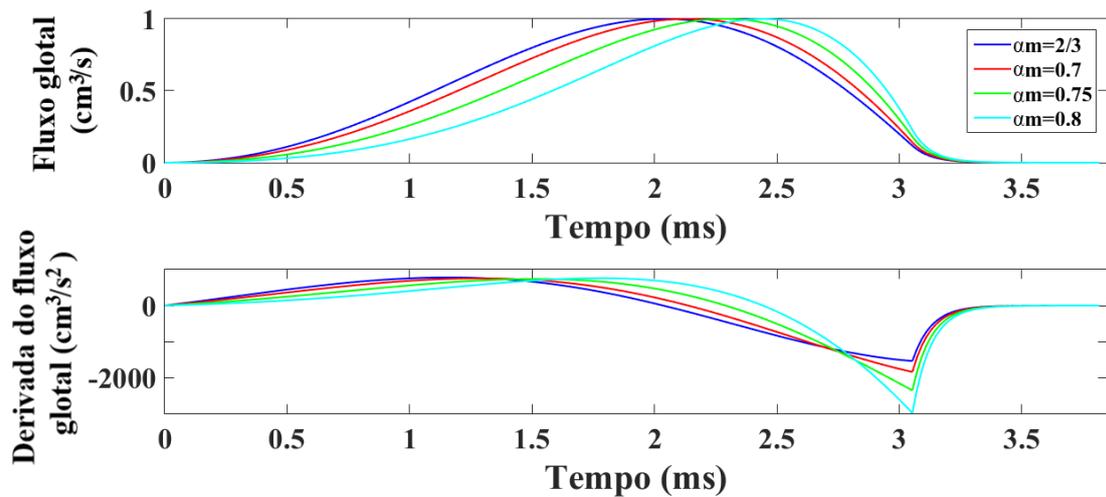


Figura 5.8: Variações temporais do pulso glotal (em cima) e a sua derivada (embaixo) para quatro valores de α_m no caso em que A_v , F_0 , O_q e Q_a sejam mantidos constantes.

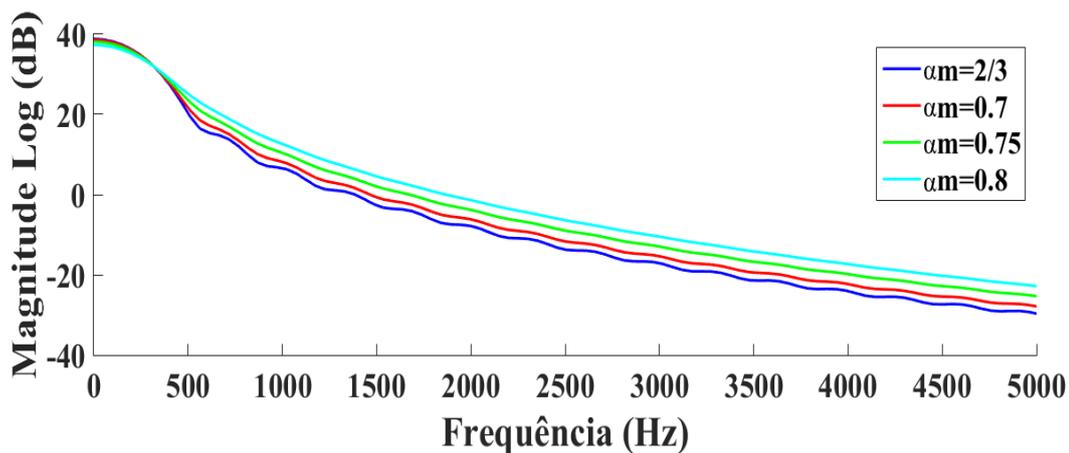


Figura 5.9: Variações espectrais do pulso glotal para quatro valores de α_m no caso em que A_v , F_0 , O_q e Q_a sejam mantidos constantes.

Um pulso glotal do modelo de Rosenberg foi sintetizado, usando os seguintes parâmetros: $A_v = 1$, $F_0 = 262 \text{ Hz}$, $\alpha_1 = 40\%$ e $\alpha_2 = 20\%$, e foi feita a comparação com três pulsos glotais do modelo LF, mantendo fixos os seguintes parâmetros durante a síntese: $A_v = 1$, $F_0 = 262 \text{ Hz}$, $O_q = 0.6$ e $\alpha_m = 2/3$, enquanto que o Q_a teve uma variação crescente para cada pulso nos seguintes valores (0.1, 0.3, 0.5). Esta comparação permitiu observar os efeitos causados pela variação do Q_a no pulso glotal, como mostrado na Fig. 5.10. Da mesma forma, foram comparadas as derivadas dos respectivos pulsos glotais, tal como ilustrado na Fig. 5.11, por conseguinte, observa-se que a derivada do pulso glotal do modelo de Rosenberg corresponde a um fechamento abrupto ($Q_a = 0$), porém, uma descontinuidade no instante de fechamento glotal ($t = O_q T_0$) é observada. Por outro lado, pode-se perceber que à medida em que aumenta o valor de Q_a , o tempo de duração correspondente à fase de retorno no pulso glotal é induzido, com o desaparecimento da descontinuidade e criando uma concavidade voltada para cima na derivada do pulso.

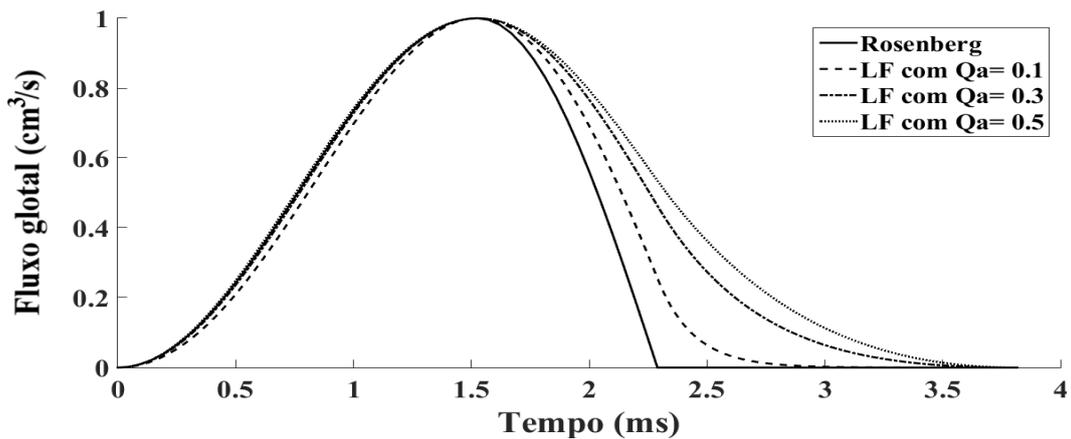


Figura 5.10: Fase de retorno do fluxo glotal dos modelos de Rosenberg e LF para três valores de Q_a .

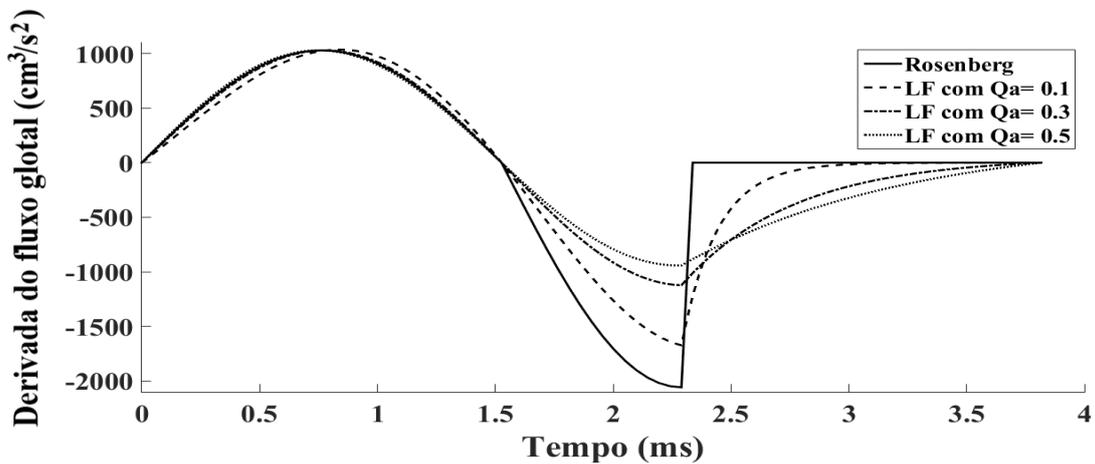


Figura 5.11: Fase de retorno da derivada do fluxo glotal dos modelos de Rosenberg e LF para três valores de Q_a .

5.6 Análise de sensibilidade dos parâmetros do modelo Rosenberg

No pulso glotal de Rosenberg, os parâmetros α_1 e α_2 causam diferentes efeitos da forma de onda no domínio do tempo e da frequência. As Figs. 5.12 e 5.13 apresentam o efeito da variação de α_1 no domínio do tempo e da frequência, respectivamente. Para a síntese, foram mantidos constantes os parâmetros $A_v = 1$, $F_0 = 262 \text{ Hz}$ e $\alpha_2 = 10\%$. Quando o valor de α_1 aumenta (20, 40, 60, 80), o tempo de abertura do pulso glotal também aumenta, pois α_1 está associado com a porção crescente do pulso. Esse aumento da fase de abertura do pulso induz uma compressão no domínio da frequência e, também, um deslocamento da frequência glotal para as baixas frequências, similar ao efeito produzido pela variação de O_q no modelo LF.

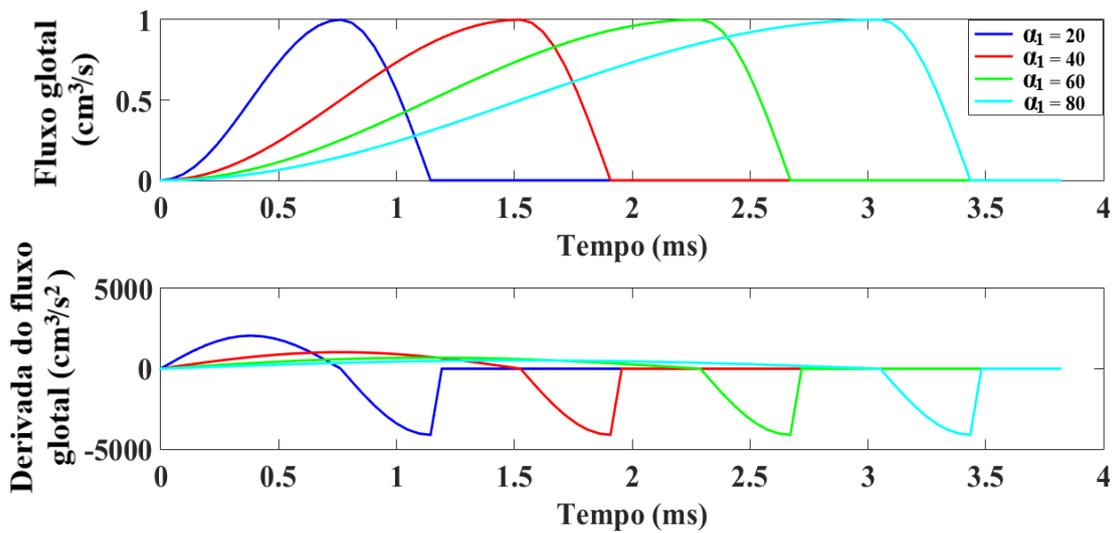


Figura 5.12: Variações temporais do pulso glotal (em cima) e sua derivada (embaixo) para quatro valores de α_1 no caso em que A_v , F_0 e α_2 são mantidos constantes.

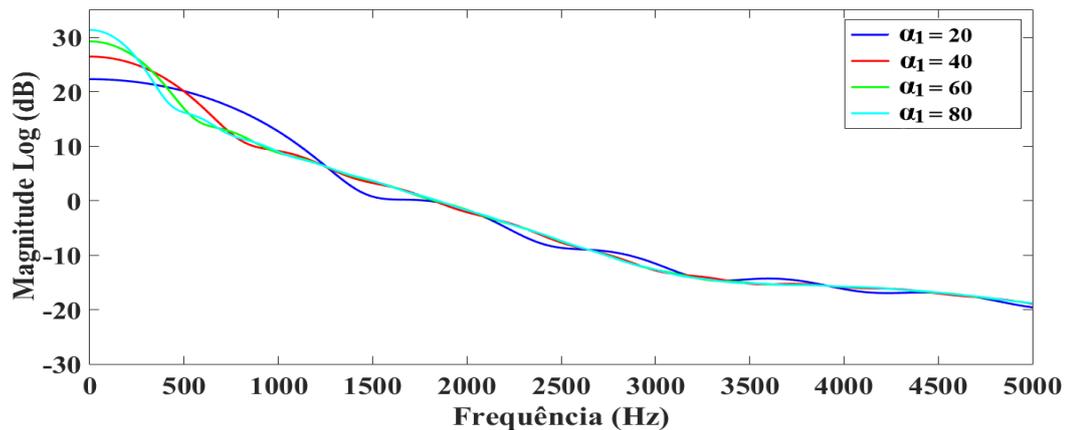


Figura 5.13: Variações espectrais do pulso glotal para quatro valores de α_1 no caso em que A_v , F_0 e α_2 são mantidos constantes.

As Figs. 5.14 e 5.15 apresentam o efeito da variação de α_2 no domínio do tempo e da frequência, respectivamente. Neste caso, os valores mantidos constantes foram: $A_v = 1$, $F_0 = 262 \text{ Hz}$ e $\alpha_1 = 10\%$ e α_2 teve uma variação em (20, 40, 60 e 80). O incremento de α_2 , induz um alongamento da porção decrescente do pulso glotal, equivalente a um maior tempo na fase de fechamento glotal. Esse incremento causa um alongamento da fase aberta do pulso e, também, o fluxo glotal aumenta. No domínio da frequência, um aumento de α_2 induz uma compressão maior da apresentada pelo α_1 e um deslocamento glotal para baixas frequências.

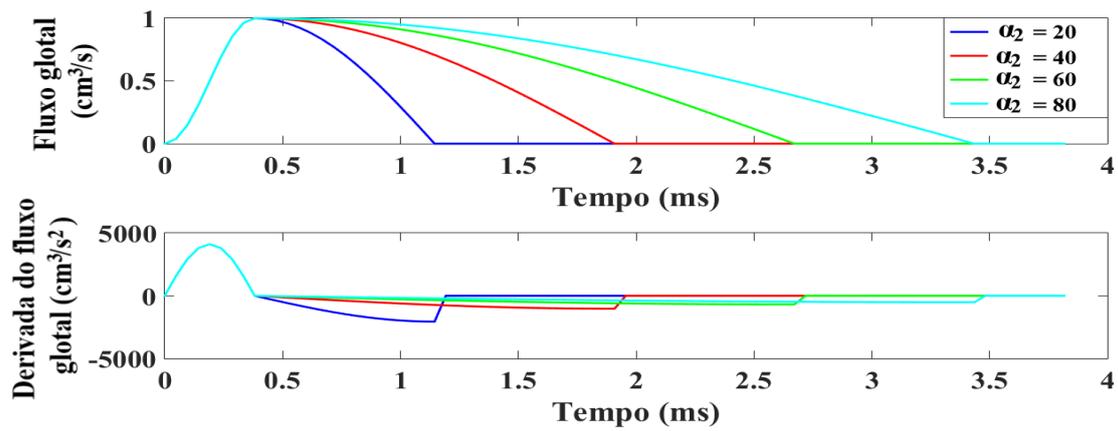


Figura 5.14: Variações temporais do pulso glotal (em cima) e sua derivada (embaixo) para quatro valores de α_2 no caso em que A_v , F_0 e α_1 sejam mantidos constantes.

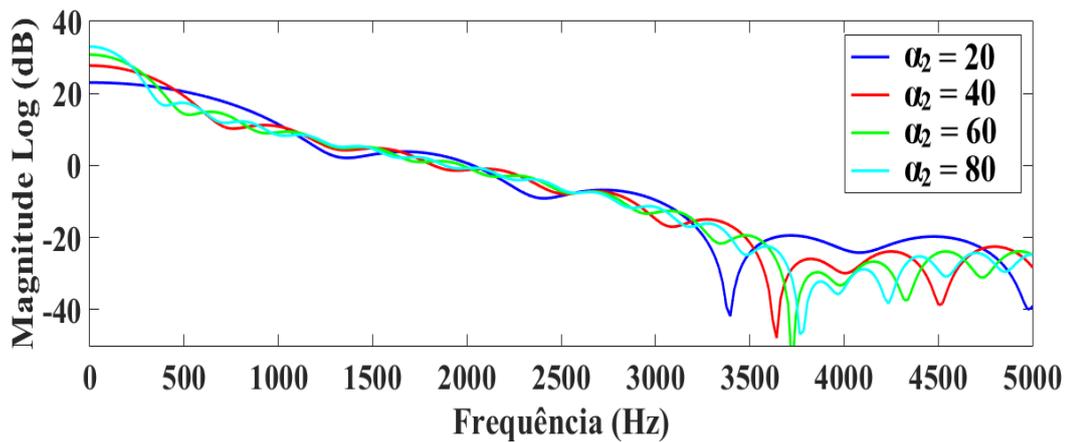


Figura 5.15: Variações espectrais do pulso glotal para quatro valores de α_2 no caso em que A_v , F_0 e α_1 são mantidos constantes.

5.7 Síntese do sinal glotal

A síntese da excitação glotal que entra no trato vocal é gerada com o código em Matlab do apêndice C. O parâmetro T_{fs} é usado para variar o tempo de duração da excitação glotal, que do ponto de vista fisiológico, corresponde ao tempo usado pelo falante para a fonação do som sustentado. No domínio do tempo, um incremento de T_{fs} induz a uma maior duração da excitação glotal. Na Fig. 5.16, apresentam-se, apenas, três pulsos glotais consecutivos que representam a excitação glotal de uma vogal sustentada, com tempo de duração de 3 segundos. Este sinal glotal foi obtido a partir da convolução de um trem de pulsos periódicos, com tempo de duração T_{fs} , e o pulso glotal do modelo LF.

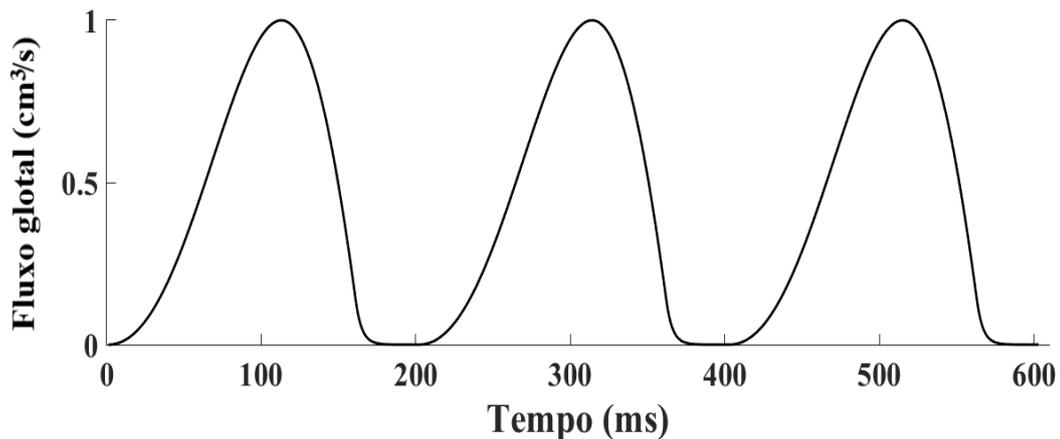


Figura 5.16: Sinal glotal do modelo LF gerado com os parâmetros: $A_v = 1$, $F_0 = 262 \text{ Hz}$, $O_q = 0.8$, $\alpha_m = 0.7$ e $Q_a = 0.1$.

5.8 Síntese das vogais cantadas com fonte glotal pelos modelo de Rosenberg e LF

No domínio da frequência, todas os harmônicos da fonte glotal que passam pelo filtro do trato vocal são atenuados ou amplificados. Quanto mais perto está a frequências formante da frequência de um harmônico, mais aumentará a amplitude e a qualidade sonora da emissão. Neste trabalho, foram consideradas cinco frequências formantes para as vogais do espanhol.

Para sintetizar as vogais do espanhol cantadas, foram desenvolvidos os códigos em Matlab dos apêndices D e E. O apêndice D sintetiza as vogais cantadas, usando a fonte glotal pelo modelo de Rosenberg, e o apêndice E usando a fonte glotal pelo modelo LF.

Para a frequência de fonação F_0 foram escolhidas frequências equivalentes às notas musicais que pertencem a determinada extensão das categorias vocais: baixo, barítono, tenor e altos.

Em relação às frequências formantes do filtro do trato vocal, F_1 e F_2 , foram escolhidas aquelas que dão mais clareza às vogais faladas e, portanto, às cantadas. Por outro lado, as formantes F_3 , F_4 e F_5 fazem parte do agrupamento que geram o formante do cantor, então, a F_4 corresponde à frequência central do formante do cantor e as frequências F_3 e F_5 foram ajustadas às harmônicas de F_0 mais próximas à frequência central pela esquerda e pela direita, respectivamente.

A Fig. 5.17, mostra a resposta em frequência do trato vocal na síntese das cinco vogais do espanhol cantadas no caso de um cantor com categoria vocal de baixo e usando o mecanismo de cobertura. Para síntese, usou-se a fonte glotal do modelo LF com os seguintes parâmetros: $A_v = 7$, $F_0 = 98 \text{ Hz}$, $O_q = 0.8$, $\alpha_m = 0.75$ e $Q_a = 0.3$. A F_0 corresponde à nota musical *Sol2* sustentada por um tempo de fonação $T_{fs} = 3 \text{ s}$. Os áudios correspondentes à síntese das vogais são apresentados em: <https://www.dropbox.com/sh/ba94bvt6d5743k3/AACHXnZUGB3PXsmmd18NH88Ja?dl=0>.

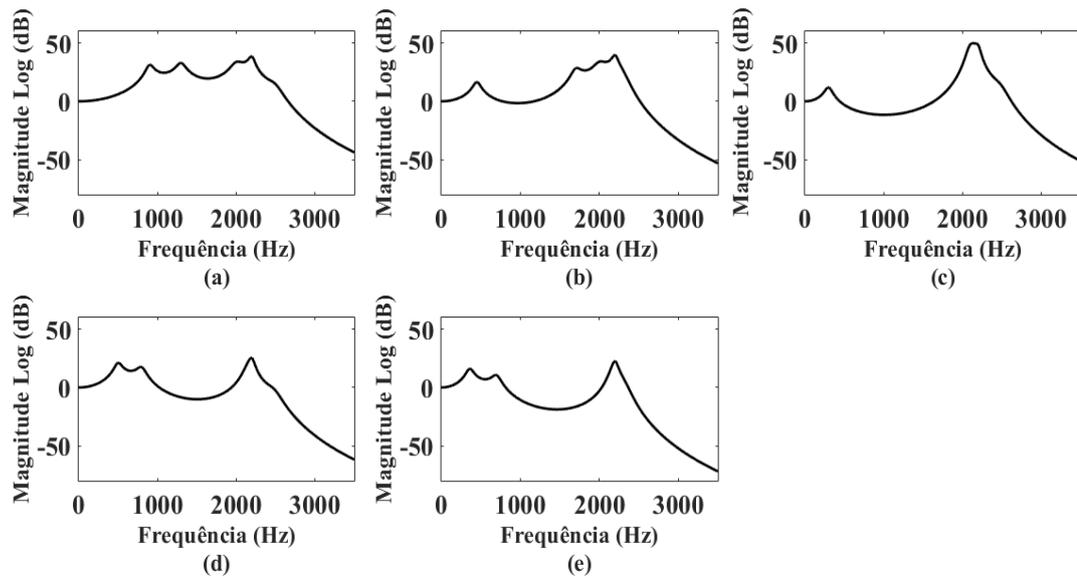


Figura 5.17: Resposta em frequência do trato vocal para as vogais do espanhol cantadas no caso de um baixo e uso do modelo LF. (a) vogal a, (b) vogal e, (c) vogal i, (d) vogal o, (e) vogal u.

A Fig. 5.18 apresenta a resposta em frequência do trato vocal na síntese das cinco vogais do espanhol cantadas com cobertura para o mesmo caso do teste da Fig. 5.17, com diferença na natureza da fonte de excitação, criada a partir do modelo de Rosenberg. Os parâmetros do pulso glotal de Rosenberg permitiram gerar um pulso glotal equivalente do modelo LF. Os parâmetros são os seguintes: $A_v = 7$, $F_0 = 98 \text{ Hz}$, $\alpha_1 = 60\%$ e $\alpha_2 = 20\%$. Os áudios da síntese das vogais do espanhol cantadas correspondentes a este caso estão disponíveis em https://www.dropbox.com/sh/ay2fyig936pbtvz/AAABuGzbSdrk6o_kz4px0-iHa?dl=0. Ao comparar os áudios da síntese das vogais cantadas pelos dois modelos glotais, pode-se perceber que no modelo de Rosenberg o nível da pressão do som é maior do que modelo LF. Isso indica que a fonte glotal do modelo de Rosenberg gera maior energia acústica do que modelo LF. Por outro lado, a qualidade do som é mais metálico e pressionado no modelo de Rosenberg, devido ao fato que a forma de onda do pulso apresenta uma queda abrupta na parte decrescente, enquanto que, no caso da fonte de excitação do modelo LF, uma qualidade aflautada no som é percebida, devido ao fato do fluxo mudar gradualmente e não apresentar descontinuidades, quando $Q_a > 0$.

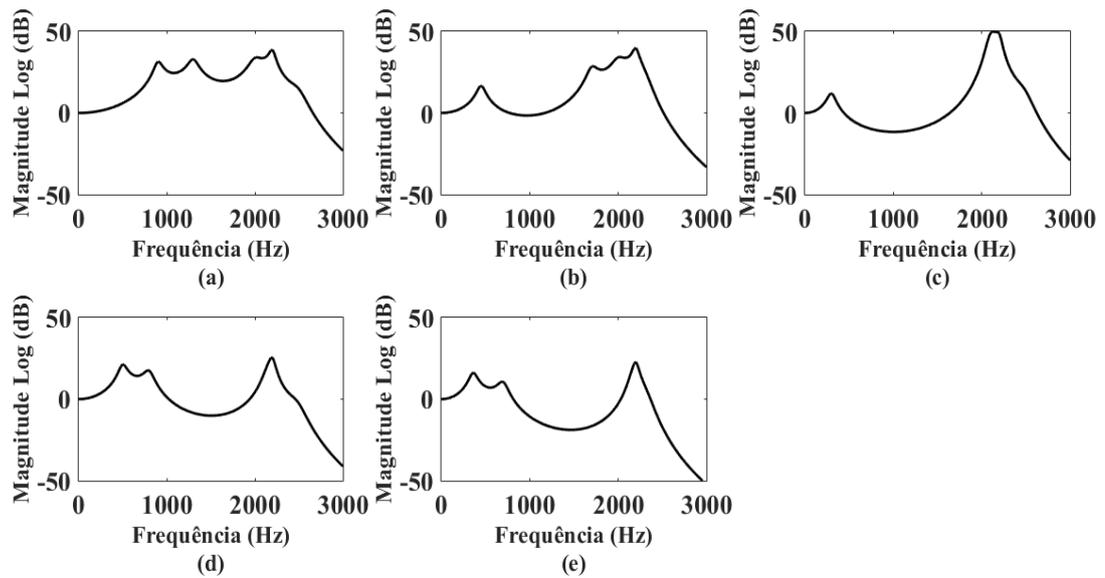


Figura 5.18: Resposta em frequência do trato vocal para as vogais do espanhol cantadas no caso de um baixo e uso do modelo de Rosenberg. (a) vogal a, (b) vogal e, (c) vogal i, (d) vogal o, (e) vogal u.

A Tab. 5.1, apresenta os valores (em Hertz) para as cinco primeiras frequências de ressonância do trato vocal que determinam as vogais do espanhol cantadas da Fig. 5.17 e Fig. 5.18. F_4 foi escolhida como a frequência central do formante do cantor com valor de 2.2 kHz , correspondente à encontrada no espectro de um cantor de categoria vocal de baixo. Por outro lado, o valor de F_3 equivale à harmônica 20 e F_4 à harmônica 25 de F_0 . A F_1 e F_2 foram selecionadas da carta das frequências formantes do espanhol, como aquelas que permitiram perceber maior clareza às vogais.

	F1	F2	F3	F4	F5
a	900	1300	2000	2200	2500
e	450	1700	2000	2200	2310
i	300	1900	2100	2200	2490
o	500	800	2150	2200	2490
u	360	700	2170	2200	2330

Tabela 5.1: Formantes do filtro do trato vocal para o caso de um cantor com classificação vocal de baixo.

As perdas do som no trato vocal são representadas pelas larguras de banda das frequências formantes da Tab. 5.2. As perdas pelas paredes moles do trato vocal são pequenas como se observa nas larguras de banda dos formantes mais baixos (F_1 e F_2), e as perdas por fricção viscosa, condução térmica e radiação nos lábios são um pouco mais altas como se observa nas larguras de banda dos formantes superiores, ainda que a F_4 seja um pouco mais seletiva para gerar a voz cantada.

k	1	2	3	4	5
Bw(k)	41	52	70	32	100

Tabela 5.2: Larguras de Banda para os cinco formantes do trato vocal no caso de um cantor com classificação vocal de baixo.

A Fig. 5.19 corresponde à resposta em frequência do trato vocal para as vogais do espanhol cantadas no caso de um cantor com categoria vocal de tenor. A fonte de excitação foi gerada com o uso do modelo de LF com os seguintes parâmetros: $A_v = 7$, $F_0 = 440 \text{ Hz}$, $O_q = 0.8$, $\alpha_m = 0.75$ e $Q_a = 0.3$. As cinco vogais do espanhol foram sustentadas por um tempo de fonação $T_{fs} = 3 \text{ s}$ sob a nota musical $La4$. Os áudios de todas as vogais cantadas são apresentadas em: <https://www.dropbox.com/sh/za2nhnqcd5ev9zp/AABzS8nB4wr23UYgk0za1A0sa?dl=0>. No mesmo enlace, encontram-se os áudios das vogais sintetizadas com a fonte de excitação pelo modelo de Rosenberg, com a forma de pulso equivalente ao modelo LF, para o mesmo caso de um cantor com categoria vocal de tenor. Neste caso, o som das vogais cantadas, tanto pelo modelo de LF como pelo modelo de Rosenberg, apresenta diferenças na energia gerada por cada fonte, pois o modelo de Rosenberg entrega uma quantidade maior, assim como as características do som. No modelo de LF, é mais flautada, enquanto que, no modelo de Rosenberg, continua sendo mais metalizada. Por outro lado, como se esperava, o tom das vogais cantadas pelo tenor é mais agudo que pelo baixo.

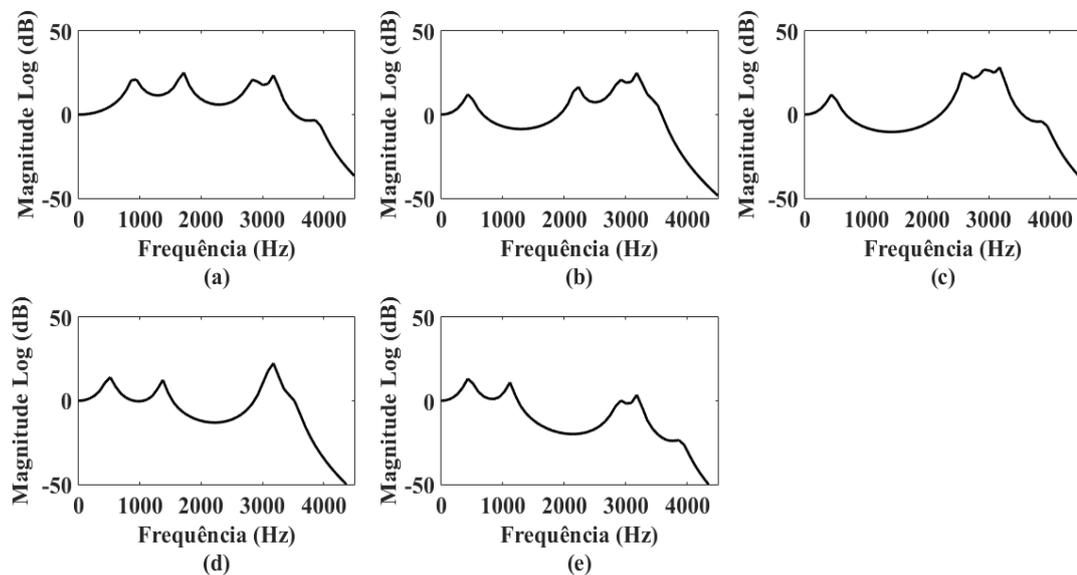


Figura 5.19: Resposta em frequência do trato vocal para as vogais do espanhol cantadas. (a) vogal a, (b) vogal e, (c) vogal i, (d) vogal o, (e) vogal u.

A Tab. 5.3, expõe as formantes usada para gerar as vogais cantadas no caso de um tenor. A F_1 e a F_2 permitiram que as vogais tiveram maior clareza. Como a extensão vocal do tenor compreende notas muito agudas, então a F_0 ultrapassa a F_1 de algumas vogais, porém, para as vogais i e u utilizou-se um deslocamento do valor da frequência da F_1 para F_0 , mudando de certa forma a identidade da vogal.

As frequências F_3 , F_4 e F_5 foram escolhidas para criar a voz cantada, onde F_3 e F_4 correspondem às harmônicas 7 e 8 respectivamente, e F_4 é a frequência central característica da formante do cantor do tenor.

	F1	F2	F3	F4	F5
a	900	1700	2850	3200	3900
e	400	2200	2900	3200	3510
i	300	2600	2950	3200	3920
o	500	1370	3080	3200	3520
u	360	1120	2900	3200	3910

Tabela 5.3: Formantes do filtro do trato vocal para o caso de um cantor com classificação vocal de tenor.

As larguras de banda que representam as perdas do som no trato vocal do cantor com categoria de tenor são apresentadas na Tab. 5.4, onde as perdas por radiação nos lábios, condução térmica e fricção viscosa (B_{w3} , B_{w4} , B_{w5}) são maiores que das parede moles (B_{w1} e B_{w2}).

k	1	2	3	4	5
Bw(k)	60	45	80	40	100

Tabela 5.4: Larguras de Banda dos cinco formantes do trato vocal para o caso de um cantor com classificação vocal de tenor.

Para a síntese das vogais cantadas no caso de uma cantora com categoria vocal de mezzo-soprano, utilizou-se a fonte glotal criada a partir dos modelos LF e Rosenberg e a resposta em frequência do trato vocal para cada vogal do espanhol é apresentada na Fig. 5.20. Como dito anteriormente, com a síntese das vogais no caso das categorias de baixo e tenor, a natureza da fonte glotal influencia as características do som das vogais e na energia gerada pela fonte devido à forma do pulso glotal, assim como se percebe nos áudios apresentados no seguinte enlace: <https://www.dropbox.com/sh/ak36xuxkj9hc5u6/AADzi5e7sJRSEo1iyyB0hp0ta?dl=0>. Os parâmetros foram calculados com a intenção de que a forma dos pulsos glotais fossem similares. Os parâmetros do modelo LF e de Rosenberg são iguais aos casos anteriores. A nota musical escolhida foi o *Re5*, equivalente à frequência de fonação $F_0 = 587 \text{ Hz}$, que pertence à extensão vocal da Mezzo-soprano e também à soprano. O tempo de fonação da vogal é $T_{fs} = 3 \text{ s}$.

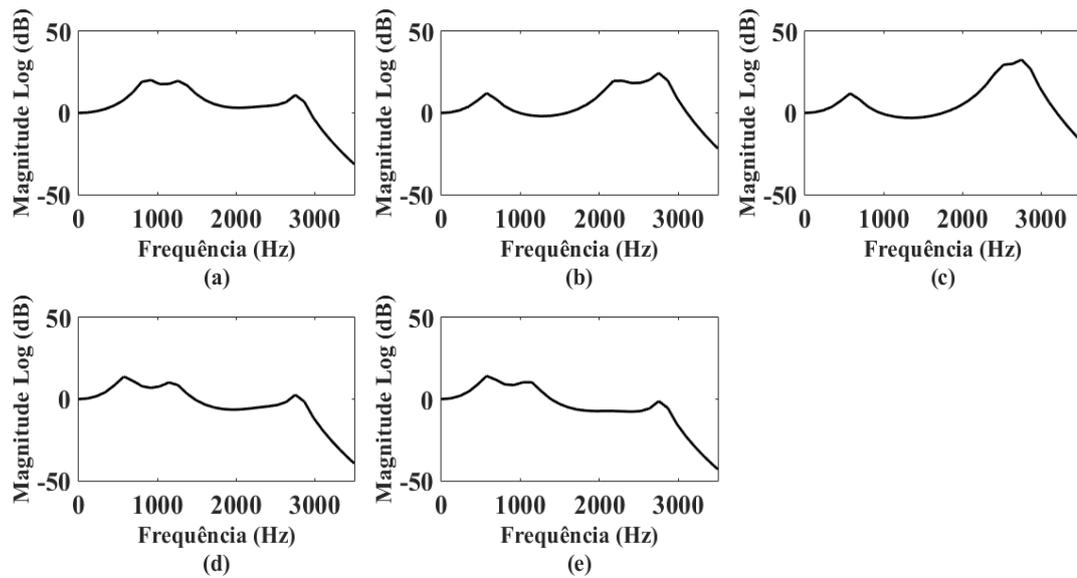


Figura 5.20: Resposta em frequência do trato vocal para as vogais do espanhol cantadas. (a) vogal a, (b) vogal e, (c) vogal i, (d) vogal o, (e) vogal u.

As frequências de ressonância que determinaram a síntese das vogais cantadas, no caso da mezzo-soprano, são mostradas na Tab. 5.5. Como as categorias vocais de mezzo-soprano e soprano são próprias do canto feminino, que ao emitir notas agudas, com harmônicas de valores frequenciais elevados, como neste caso $F_0 = 587 Hz$, então, os primeiros dois formantes de todas as vogais foram definidos com valores altos em frequência. Ainda assim, para as vogais *e*, *i*, *o*, *u*, o valor da frequência da F_1 foi deslocada para o valor de F_0 e, desse jeito, cobrir a voz, o que significa que, fisiologicamente, o som está sendo emitido com a maior abertura bucal possível, chegando em alguns casos a mudar a identidade da vogal [30]. A F_4 apresenta o valor frequencial de $2.8 kHz$, que corresponde ao valor mais comum para a frequência central da formante do cantor de um alto.

	F1	F2	F3	F4	F5
a	850	1300	2340	2800	2930
e	400	2200	2650	2800	2930
i	300	2500	2680	2800	2960
o	500	1200	2350	2800	2930
u	350	1120	2200	2800	2900

Tabela 5.5: Formantes do filtro do trato vocal para o caso de uma cantora com classificação vocal de mezzo-soprano.

A Tab. 5.6 expõe os valores das larguras de banda das frequências de ressonância do trato vocal no caso da mezzo-soprano. Como pode-se observar os valores são maiores comparados com os casos anteriores (baixo e tenor), indicando que as perdas do som no trato vocal são maiores, ainda que para a F_4 a largura de banda seja mais seletiva.

k	1	2	3	4	5
Bw(k)	90	110	380	60	450

Tabela 5.6: Larguras de Banda dos cinco primeiros formantes do trato vocal para o caso de uma cantora com classificação vocal de mezzo-soprano.

5.9 Efeitos produzidos pela forma de onda dos pulsos da fonte glotal na voz cantada

No seguinte link <https://www.dropbox.com/sh/h3dv147gu8nfqog/AAAmzxxk0rhV-9rtIfsdepnaxa?dl=0>, apresentam-se os arquivos de áudio da síntese das vogais do espanhol cantadas com cobertura na nota musical *Sol2* (equivalente a $F_0 = 98 \text{ Hz}$), que pertence à extensão vocal de um cantor com categoria de baixo. A fonte de excitação foi obtida a partir do modelo de Rosenberg e a sua forma varia de acordo com os seguintes casos:

1. A fase de abertura do pulso glotal é maior do que a fase de fechamento.
2. A fase de abertura do pulso glotal é igual do que a fase de fechamento.
3. A fase de abertura é menor do que a fase de fechamento.

Observe-se que as formas de onda do pulso glotal do modelo de Rosenberg para os casos 1, 2 e 3 foram apresentadas nas Figs. 5.1, 5.2 e 5.3, respectivamente, e que além de ter trocado o valor da F_0 dos pulsos (de 262 Hz para 98 Hz) também mudou-se o valor de A_v (de 1 para 7).

O efeito produzido pela mudança da forma do pulso nos três casos permite perceber que: no primeiro caso, o som apresenta maior energia acústica, associada com uma qualidade pressionada, enquanto que nos casos dois e três, percebe-se que o som tem uma qualidade velada e com pouca energia acústica, especialmente neste caso de baixa frequência de fonação.

Por outro lado, as vogais do espanhol cantadas com cobertura foram sintetizadas novamente na nota musical *Sol2* da extensão vocal de um cantor com categoria de baixo, no entanto, a fonte de excitação foi gerada a partir do modelo do pulso glotal de LF. A forma de onda do pulso foi submetida aos três casos anteriormente mencionados para perceber os efeitos causados pelo modelo no som.

Observa-se que as formas de onda do pulso glotal do modelo de LF para os casos 1 e 2 foram apresentadas nas Figs. 5.4 e 5.5, respectivamente, e que a forma de onda para o caso 3 não foi obtida devido ao fato que o modelo LF considera que a fase de abertura glotal é sempre maior do que a fase de fechamento glotal. Os arquivos de áudio da síntese estão disponíveis no seguinte link: <https://www.dropbox.com/sh/32x7oi7uebsjs8m/AACR4w-h1YqJcZ04uu6FGySka?dl=0>.

Em relação ao sons dos fonemas produzidos nos dois casos apresentados anteriormente, pode-se perceber claramente as diferenças em cada caso. Ao mudar a forma de onda do pulso como no caso 1, os fonemas cantados tendem a ser mais naturais com um efeito levemente metálico e o nível de pressão sonora é de tipo regular. Para o caso 2, a forma de onda do pulso é simétrica, porém, os sons dos fonemas

cantados compreendem a qualidade velada e a sua energia acústica é bastante menor do que a apresentada no mesmo caso para o modelo de pulso glotal de Rosenberg.

5.10 Análise de sensibilidade da voz cantada com fonte glotal pelo modelo LF

Devido ao fato que a forma de onda do pulso glotal influencia as qualidades acústicas do som, sintetizou-se o fonema vocálico /a/ cantado com cobertura por um cantor de categoria vocal baixo na nota musical *Sol2* (equivalente a $F_0 = 98 \text{ Hz}$), através do programa no apêndice E, usando a fonte glotal pelo modelo LF. A variação dos parâmetros do modelo de LF induzem diferentes efeitos na qualidade do som. Foram usados os seguintes casos:

1. Foram mantidos constantes os seguintes parâmetros: $A_v = 7$, $\alpha_m = 2/3$ e $Q_a = 0.1$ e o valor de O_q foi variado (0.2, 0.4, 0.6, 0.8).
2. Foram mantidos constantes $A_v = 7$, $O_q = 0.8$ e $Q_a = 0.1$, enquanto que, α_m varia em (2/3, 0.7, 0.75, 0.8).
3. Foram mantidos constantes $A_v = 7$, $O_q = 0.6$ e $\alpha_m = 2/3$, enquanto que o Q_a teve uma variação crescente para cada pulso nos seguintes valores (0.1, 0.3, 0.5).

Os arquivos de áudio que revelam os efeitos produzidos na qualidade acústica da síntese da vogal /a/, cantada para os três casos anteriores, estão disponíveis no seguinte link: <https://www.dropbox.com/sh/tntpwmvtsnwnfhu/AACS4efyMdQ0i3foUWD2PaFDa?dl=0>. De acordo com o caso 1, percebe-se que à medida em que O_q aumenta, o som passa de uma qualidade pressionada a uma qualidade velada, que é percebida facilmente nas frequências baixas. Para o caso 2, percebe-se que à medida em que α_m aumenta, o som passa de uma qualidade aflautada com $\alpha_m = 2/3$ a uma qualidade metálica e pressionada com $\alpha_m = 0.8$. Isso acontece porque o tempo de máximo fluxo está sendo deslocado à direita, causando mudanças abruptas na pendente negativa do pulso, que em altas frequências o som é totalmente metalizado. No caso 3, à medida que incrementa Q_a , o som passa a uma qualidade tênue e com pouca sonoridade.

5.11 Análise de sensibilidade da voz cantada com fonte glotal pelo modelo Rosenberg

Para esta análise, foi sintetizada de forma similar ao caso anterior, o fonema vocálico /a/ cantado com cobertura por um cantor de categoria vocal baixo na nota musical *Sol2* (equivalente a $F_0 = 98 \text{ Hz}$), com o programa no apêndice D, usando a fonte glotal pelo modelo Rosenberg. A variação dos parâmetros do modelo de Rosenberg induzem diferentes efeitos na qualidade do som. Foram considerados os seguintes casos:

1. Foram mantidos constantes $A_v = 7$, $\alpha_2 = 10\%$, enquanto que o valor de α_1 varia (20, 40, 60 e 80).
2. Foram mantidos constantes $A_v = 7$, $\alpha_1 = 10\%$, enquanto que o valor de α_2 varia (20, 40, 60 e 80).

De acordo com esses dois casos de variação de parâmetros, foram evidenciados os efeitos na qualidade acústica da síntese da vogal /a/ cantada na nota musical *Sol2*, através dos arquivos de áudio gerados e apresentados no seguinte enlace: https://www.dropbox.com/sh/2nr1vukvq2ru1xw/AABQrQsyu4IK5disMVTbYi_5a?dl=0. Em relação ao caso 1, percebeu-se que todas as variações de α_1 induzem uma qualidade metalizada ao som. Isto é devido a mudanças abruptas nas pendentes de todas as formas de onda do pulso glotal. Além disso, o som está mais associado com uma qualidade pressionada quando $\alpha_1 = 20\%$. Para o caso 2, identificou-se um som pressionado e metalizado para $\alpha_2 = 20\%$, tornando-se cada vez mais velado na medida em que α_2 aumenta. Os sons estudados com esse modelo de pulso glotal apresentam maior energia acústica que no caso do modelo LF.

5.12 Síntese de uma sequência de vogais em diferentes notas musicais

Uns dos códigos em Matlab mais versáteis para a síntese das vogais cantadas, implementados neste trabalho, correspondem aos apresentados nos apêndices F e G. Eles têm a particularidade de permitir que seja gerada qualquer sequência de vogais do espanhol cantadas, para uma sequência de notas musicais de igual tamanho que a sequência de vogais. Cada vogal cantada dentro da sequência contém um tempo de duração $T_{fs} = 2s$. Os arquivos de áudio estão disponíveis no seguinte link: https://www.dropbox.com/sh/r0oa5465ad0ectu/AABtaCgLTnkvKEQ8SgXmo_a?dl=0 e apresentam a sequência de vogais /a/, /e/, /i/, /o/, /u/, /a/, /e/, /i/ para as notas *Do2*, *Re2*, *Mi2*, *Fa2*, *Sol2*, *La2*, *Si2*, *Do3*. Nota-se que esta sequência de notas musicais equivale a uma oitava da extensão vocal de um cantor com categoria de baixo, porém, as frequências formantes usadas coincidem com as apresentadas na Tab. 5.1 e as suas larguras de banda pela Tab. 5.2. Os pulsos glotais da fonte de excitação usada no código do apêndice F foram criados a partir do modelo de Rosenberg, e do apêndice G, pelo modelo de LF. Cada vogal cantada na sequência, para o caso do modelo de Rosenberg, apresenta a qualidade metalizada e pressionada devido à forma do pulso glotal, enquanto que no caso do modelo de LF, o som de cada vogal tem uma qualidade metalizada embora mais tênue do que a apresentada pelo modelo de Rosenberg.

5.13 Síntese das vogais cantadas com *jitter* na fonte glotal

A síntese das vogais cantadas até aqui estão relacionadas com modelos determinísticos do sinal glotal, pois considera as oscilações das cordas vogais como periódicas, no entanto, nesta secção pretende-se incluir uma pequena flutuação aleatória chamada de *jitter* [16, 17, 18] em cada comprimento do ciclo glotal. Através do código mostrado no anexo H, sintetizaram-se as cinco vogais do espanhol cantadas com *jitter* no caso do cantor com categoria vocal de baixo na nota musical *Sol2*. A sua fonte de excitação, apresentada na Fig. 5.21, foi criada a partir do modelo de Rosenberg para um tempo de fonação $T_{fs} = 3s$. Cada período do pulso glotal foi modificado pelo *jitter*, por exemplo, o *jitter* para o primeiro pulso glotal da Fig. 5.21, modifica o período fundamental em 14.65%, para o segundo pulso em 5.36% e para o terceiro, o *jitter* é 15.38%. Todas estas flutuações foram obtidas aleatoriamente.

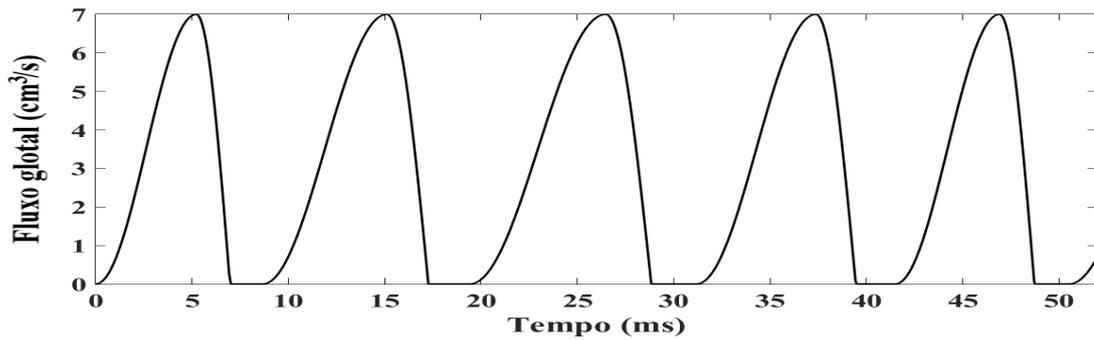


Figura 5.21: Simulação da fonte glotal com *jitter*, usando o modelo de Rosenberg com parâmetros: $A_v = 7$, $F_0 = 98 \text{ Hz}$, $\alpha_1 = 60\%$, $\alpha_2 = 20\%$.

Neste caso, os efeitos produzidos pelo *jitter* na síntese das vogais cantadas são evidenciados nos arquivos de áudio disponíveis no seguinte enlace: <https://www.dropbox.com/sh/nt7lfkvluwxwrf2/AAAIpdeX5yvtUtWzG6ia2ljaa?dl=0>. Neles, percebe-se uma qualidade adicional que torna rouco e áspero o som de todas as vogais sintetizadas, isto, é devido ao fato que o valor do *jitter* ultrapassa o valor de 1% e provoca um distúrbio na voz.

Do mesmo modo, com o código do anexo I, sinterizaram-se as vogais do espanhol cantadas por um cantor de categoria vocal de baixo na nota musical *Sol2*; com especial diferença na natureza da fonte glotal, pois os pulsos glotais foram gerados a partir do modelo LF. Como se observa na Fig. 5.22, cada comprimento do ciclo glótico, na fonte de excitação, apresenta uma flutuação aleatória correspondente ao *jitter*, que nesse caso, tem a variação de 4.81% para o primeiro pulso, 0.31% para o segundo pulso e 21.13% para o terceiro pulso. Os arquivos de áudio obtidos por efeito de essa fonte de excitação são exibidos no seguinte enlace: <https://www.dropbox.com/sh/i9ouk116ma19qhm/AAAVUr2Z36kRaGGI8dquWTj4a?dl=0>.

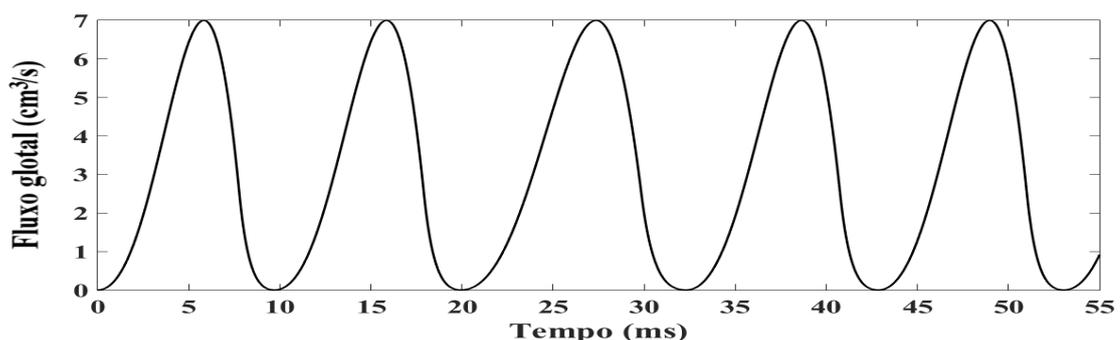


Figura 5.22: Simulação da fonte glotal com *jitter*, usando o modelo de LF com parâmetros: $A_v = 7$, $F_0 = 98 \text{ Hz}$, $O_q = 0.8$, $\alpha_m = 0.75$, $Q_a = 0.3$.

Nos arquivos de áudio, estão disponíveis as cinco vogais do espanhol sintetizadas, nas que é possível perceber o efeito causado pelo *jitter*, no entanto, a qualidade rouca e áspera entregue ao som é mais fraca que no caso da fonte glotal com *jitter* usando o modelo de Rosenberg. A vista disso, tanto o baixo *jitter* gerado neste caso, como o uso do modelo LF na fonte glotal, deram à voz uma qualidade mais natural e sem presença de patologia [35].

Capítulo 6

Conclusões e sugestões para trabalhos futuros

6.1 Conclusões

1. No estudo comparativo entre os modelos de pulso glotal, de Rosenberg e de LF, foi evidenciado que o modelo de Rosenberg permite gerar qualquer tipo de forma de onda enquanto que o modelo de LF apenas considera que a fase de abertura é sempre maior que a fase de fechamento glotal.
2. Em relação à variação de parâmetros, no modelo de LF, mostrou-se que o aumento de O_q induz a uma expansão da fase aberta do pulso glotal, e essa expansão do pulso induz uma compressão no domínio da frequência, causando um deslocamento da frequência glotal para as baixas frequências. Por outro lado, o aumento do parâmetro α_m , causa uma deformação da forma do seno do pulso glótico e um deslocamento da frequência glotal de maneira oposta a de O_q .
3. Como resultado da comparação entre a derivada do pulso glotal do modelo de Rosenberg e a derivada do pulso glotal do modelo de LF, sob a variação do parâmetro Q_a , evidenciou-se que o pulso de Rosenberg apresenta uma descontinuidade no instante de fechamento glotal, o que corresponde ao pulso glotal pelo modelo de LF com um fechamento abrupto. E, à medida em que o valor Q_a aumenta, induz-se uma fase de retorno do pulso glotal, sumindo cada vez mais a descontinuidade e criando uma concavidade voltada para cima na derivada do pulso glotal.
4. O aumento da fase de abertura do pulso glotal do modelo de Rosenberg, obtido através do incremento de α_1 , induz uma compressão no domínio da frequência e, portanto, um deslocamento da frequência glotal para as baixas frequências, similar ao efeito produzido pela variação de O_q no modelo de LF. Por outro lado, o aumento de α_2 induz um alongamento da fase de fechamento do pulso glotal, acrescentando igualmente a fase aberta do pulso e causando uma compressão no domínio da frequência ainda maior, deslocando a frequência glotal para as baixas frequências.
5. Através da síntese das vogais cantadas nas diferentes categorias vocais dos cantores líricos, mostrou-se

que as duas primeiras frequências formantes do filtro do trato vocal são as responsáveis de fornecer clareza à síntese de vogais e as formantes três, quatro e cinco, de gerar a envoltória do formante do cantor.

6. Graças à comparação dos áudios da síntese das vogais cantadas com diferente natureza de fonte glotal, percebeu-se que a fonte glotal do modelo de Rosenberg gera maior energia acústica do que modelo LF, em relação à qualidade do som. O modelo de Rosenberg causa uma qualidade metalizada e pressionada, enquanto que o modelo LF fornece uma qualidade aflautada ao som. Permitiu-se concluir que a conformação da forma de onda do pulso glotal não só ajuda a determinar a quantidade de energia acústica, mas também a qualidade do som.
7. As larguras de banda permitem modelar as perdas do trato vocal. No entanto, para gerar a envoltória da formante do cantor é importante que a largura de banda, correspondente à frequência de ressonância central, seja baixa. Isso permite inferir que o trato vocal do cantor deve apresentar a menor perda possível, em relação à fricção viscosa, condução térmica e radiação no lábios.
8. Em relação à seleção das formantes F_3 e F_4 perto das harmônicas de F_0 , observou-se que o som ficará com mais harmonia musical do que se estivesse longe delas.
9. Para síntese de vogais no caso de cantores com categoria vocal de mezzo-soprano e soprano, observou-se que ao deslocar F_1 à direita, devido à frequência de fonação F_0 ultrapassar F_1 , foi obtida maior cobertura do som emitido, ainda que, piore um pouco a inteligibilidade das vogais $/e/$, $/i/$, $/o/$ e $/u/$.
10. Em relação à síntese com fonte glotal pelo modelo de Rosenberg com o efeito produzido pela mudança da forma do pulso glotal no caso em que a fase de abertura é maior que a fase de fechamento glotal, percebeu-se que o som apresenta maior energia acústica e se associa com uma qualidade pressionada, enquanto que no caso em que a fase de abertura seja menor ou igual à fase de fechamento glotal, o som percebido apresenta uma qualidade velada e com pouca energia acústica.
11. Para a síntese de vogais cantadas com fonte glotal pelo modelo de LF, obteve-se a forma do pulso glotal influenciada no som gerado, no caso em que a fase de abertura é maior do que a fase de fechamento. Percebeu-se um som levemente metalizado e com o nível de pressão sonora de tipo regular, enquanto que quando a fase de abertura é igual à fase de fechamento glotal, o som percebido é velado e com pouca energia acústica.
12. Devido à análise de sensibilidade da voz cantada com fonte glotal pelo modelo de LF, concluiu-se que à medida que O_q aumenta, o som passa de uma qualidade pressionada a uma qualidade velada que é facilmente percebida a baixas frequências de fonação. Quando α_m aumenta, o som passa de uma qualidade aflautada a uma qualidade metálica e pressionada e, à medida que Q_a aumenta, o som se torna tênue e com pouca energia acústica.
13. Todas as variações de α_1 na fonte glotal pelo modelo de Rosenberg induzem à síntese da voz cantada uma qualidade metalizada e uma qualidade pressionada com o menor valor do mesmo. A qualidade

metalizada do som é percebida devido a mudanças abruptas da parte crescente do pulso glotal, e sua percepção incrementa à medida em que a frequência de fonação se torna mais alta.

14. O aumento de α_2 na fonte glotal pelo modelo de Rosenberg induz à síntese da voz cantada uma qualidade velada, ainda que os testes tenham apresentado maior energia acústica do que os testes equivalentes com a fonte glotal pelo modelo de LF.
15. Na síntese de voz cantada com *jitter* na fonte glotal, pelo modelo de Rosenberg, percebeu-se uma qualidade adicional que torna rouco e áspero o som de todas as vogais sintetizadas.
16. Na síntese das vogais cantadas com *jitter* na fonte glotal, pelo modelo de LF, percebeu-se o efeito gerado pelo *jitter*, não obstante, a qualidade que torna rouca e áspera o som é mais fraca que a percebida com a fonte glotal pelo modelo de Rosenberg, indicando a presença de um *jitter* baixo e, portanto, a voz percebida é muito mais natural e sem presença de patologia.

6.2 Sugestões para trabalhos futuros

Como sugestão para trabalhos futuros, está a consideração de outros modelos de pulsos glotais e uma consideração mais detalhada da geração do *jitter* em conjunto com esse estudo realizado, além de produzir cantos com esses modelos. É importante destacar a variação de frequência para o canto e para o *jitter*, uma variação é maior e a outra são de pequenas variações. Um estudo unindo essas duas considerações e os modelos de pulso glotal seria um caminho natural a seguir.

Referências Bibliográficas

- [1] González, S. F., De la Iglesia, F. V., Girbau, M. M., Urrutia, R. G. T., "La historia de la voz". *Rev Med Univ Navarra*, vol. 50, pp. 9-13, 2006.
- [2] Brinkman, R. J., Hage, J. J., "Andreas Vesalius' 500th Anniversary: Initial Integral Understanding of Voice Production". *Journal of Voice*, vol. 31, n. 1, pp. 124.e11-124.e19, 2017.
- [3] Dias, S. D. O., *Estimation of the glottal pulse from speech or singing voice*. Apresentada como dissertação de mestrado em engenharia biomédica, Universidade do Porto, 2012.
- [4] Díaz, C., "La producción de la voz: estructuras anatómicas y biomecánica laríngea". *X Congreso Argentino y V Latinoamericano de Educación Física y Ciencias*, La Plata, 2013.
- [5] Titze, I. R., *Principles of voice production*. New Jersey, Prentice-Hall, 1994. 354 p.
- [6] Dosal González, R., *Producción de la voz y el habla: la fonación*. Apresentada como trabalho de conclusão de curso em enfermagem, Escuela Universitaria de Enfermería casa de salud Valdecilla, 2014.
- [7] Ventura, J. A. P. D., *Biofeedback da voz cantada*. Apresentada como dissertação de mestrado em engenharia electrotécnica e de computadores major telecomunicações, Universidade do Porto, 2011.
- [8] Sundberg, J., *The science of singing voice*. 1990. 216 p.
- [9] Marsola, M., Bae, T., *Canto Uma Expressão*. Irmãos Vitale, 2000. 112 p.
- [10] Juliano, S. M., *Um estudo comparativo entre o sinal electroglotográfico e o sinal de voz*. Apresentada como tese de doutorado. Dissertação de mestrado em engenharia de telecomunicações, Universidade Federal Fluminense (UFF), 2008.
- [11] Santos, G. B. D., *Análise fonético-acústica das vogais orais e nasais do português: Brasil e Portugal*. Apresentada como tese de doutorado em letras e linguística, Universidade Federal de Goiás (UFG), 2013.
- [12] Gurlekian, J. A., Elisei, N., Eleta, M., "Caracterización articulatoria de los sonidos vocálicos del espanol de Buenos Aires mediante técnicas de resonancia magnética". *Revista fonoaudiológica*, vol. 50, pp. 7-14, 2004.

- [13] Salas, M., "Aplicaciones del análisis acústico en los estudios de la voz humana". *Seminario Internacional de Acústica*, Santiago de Chile, 2003.
- [14] Aguirre, F., "Manual de fonética y fonología". *Universidad Técnica Particular de Loja (UTPL)*, Loja-Ecuador, 2000.
- [15] Degottex, G., *Glottal source and vocal-tract separation*. Apresentada como tese de doutorado em computadores, telecomunicações e eletrônica, Université Pierre et Marie Curie-Paris VI (UPMC), 2010.
- [16] Cataldo, E., Soize, C.. "Jitter generation in voice signals produced by a two-mass stochastic mechanical model." *Biomedical Signal Processing and Control (Print)*, 27, 87–95, 2016.
- [17] Cataldo, E., Soize, C. "Voice signals produced with jitter through a stochastic one-mass mechanical model." *Journal of Voice*, 31 (1), 111e9 – –111e18, 2017.
- [18] Cataldo, E., Soize, C., "Stochastic mechanical model of vocal folds for producing jitter and for identifying pathologies through real voices". *Journal of biomechanics*, vol. 74, pp. 126-133, 2018.
- [19] Henrich, N., *Etude de la source glottique en voix parlée et chantée: modélisation et estimation, mesures acoustiques et électroglottographiques, perception*. Apresentada como tese de doutorado, Université Pierre et Marie Curie-Paris VI (UPMC), 2001.
- [20] Rosenberg, A. E., "Effect of glottal pulse shape on the quality of natural vowels". *The Journal of the Acoustical Society of America*, vol. 49, n. 2B, pp. 583-590, 1971.
- [21] Fant, G., "The acoustic theory of speech production". *S'Gravenhage, Mouton*, 1960.
- [22] Fant, G., "Vocal source analysis-a progress report". *STL-QPSR*, vol. 20, n. 3-4, pp. 31-53, 1979.
- [23] Fant, G., "The source filter concept in voice production". *STL-QPSR*, vol. 22, n. 1, pp. 21-37, 1981.
- [24] Fant, G., Liljencrants, J., Lin, Q. G., "A four-parameter model of glottal flow". *STL-QPSR*, vol. 26, n. 4, pp. 1-13, 1985.
- [25] Doval, B., d'Alessandro, C., Henrich, N., "The spectrum of glottal flow models". *Acta acustica united with acustica*, vol. 92, n. 6, pp. 1026-1046, 2006.
- [26] Henrich, N., d'Alessandro, C., Doval, B., "Glottal flow models: waveforms, spectra and physical measurements". *Forum Acusticum*, pp. 1-6, 2002.
- [27] Rabiner, L. R., Schafer, R. W., *Theory and applications of digital speech processing*. New Jersey, Pearson Higher Education, 2011. 1042 p.
- [28] Guzmán, M., "Acústica del tracto vocal". *Académico Unidad de Voz de la Escuela de Fonoaudiología, Universidad de Chile*, 2010.

- [29] Facal, M. L., *La voz del cantante: estudio comparativo del análisis objetivo y subjetivo de la voz hablada y cantada*. Buenos Aires-Argentina, Librería Akadia Editorial, 2005. Originalmente apresentada como tese de doutorado, Universidad del Museo Social Argentino, 1998.
- [30] Adachi, S., Yu, J., "Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing". *The Journal of the Acoustical Society of America*, vol. 117, n. 5, pp. 3213-3224, 2005.
- [31] Sundberg, J., "Vocal tract resonance in singing". *The NATS Journal*, vol. 44, n. 4, pp. 11-20, 1988.
- [32] Kayte, S., Waghmare, K., Gawali, B., "Marathi Speech Synthesis: A review ". *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 3, n. 6, pp. 3708-3711, 2015.
- [33] Prasad, K. S., Ramaiah, G. K., Manjunatha, M. B., "Backend Tools for Speech Synthesis in Speech Processing". *Indian Journal of Science and Technology*, vol. 10, n. 1, pp. 1-8, 2017.
- [34] Fleischer, M., Pinkert, S., Mattheus, W., Mainka, A., Mürbe, D., "Formant frequencies and bandwidths of the vocal tract transfer function are affected by the mechanical impedance of the vocal tract wall ". *Biomechanics and modeling in mechanobiology* , vol. 14, n. 4, pp. 719-733, 2015.
- [35] Yamasaki, R., Montagnoli, A., Murano, E. Z., Gebrim, E., Hachiya, A., da Silva, J. V. L., Behlau, M., Tsuji, D., "Perturbation Measurements on the Degree of Naturalness of Synthesized Vowels". *Journal of Voice*, vol. 31, pp. 389.e1-389.e8, 2017.

Apêndice A

Algoritmo do pulso glotal com o modelo de Rosenberg

```
clear;
clc;
Av=1;    %Máxima amplitude do fluxo glotal
F0=262;  %Frequência fundamental
T0=1/F0; %Periodo fundamental
Tfs=3;   %Tempo de fonação sustentada
fm=22000; %Máxima frequência de amostragem
alpha1=40; %Porcentagem do tempo de abertura relativa
alpha2=20; %Porcentagem do tempo de fechamento relativa
Tp=(alpha1/100)*T0; %Tempo da abertura glotal
TN=(alpha2/100)*T0; %Tempo da fechamento glotal
n=round(fm/F0,1,'significant');
%-----
%Rosenberg's Glottal Pulse Approximation
%-----
Ts=T0/n;
fs=1/Ts;
t1=0:Ts:Tp;
t2=(Tp+Ts):Ts:(Tp+TN);
t3=(Tp+TN+Ts):Ts:T0;
t=[t1 t2 t3];
%-----
x1=cos(pi*t1/Tp);
gn1=(Av/2)*(1-x1);
%-----
```

```

x2=pi*(t2-Tp);
gn2=Av*cos(x2/(2*TN));
%-----
gn3=zeros(1,length(t3));
gn=[gn1 gn2 gn3];
%-----
%Rosenberg's Glottal Pulse Derivate
%-----
udg=diff(gn)/Ts;
%-----
x3=sin(pi*t1/Tp);
udgn1=(Av*pi/(2*Tp))*x3;
%-----
x4=sin(pi*(t2-Tp)/(2*TN));
udgn2=(-Av*pi/(2*TN))*x4;
%-----
udgn3=zeros(1,length(t3));
%-----
udgn=[udgn1 udgn2 udgn3];
%-----
%Comb Dirac
%-----
bin=fix((Tfs-T0)/(T0));
LD=(length(gn)*bin)+2;
Dirac=zeros(LD,1);
Dirac(1:length(gn):end)=1;
%-----
%Convolution of Glottal Flow and Dirac
%-----
yk=conv(gn,Dirac);
%-----
%Plot of Glottal Pulse
%-----
figure(1)
subplot(2,1,1);
plot(t*1000,gn,'k');
title(['Forma de onda do pulso glotal de Rosenberg']);
xlabel({'Tempo (ms)'})
ylabel('Amplitude')

```

```
%-----  
%plot Rosenberg's Glottal Pulse Derivate  
%-----  
subplot(2,1,2);  
plot(t*1000,udgn,'b');  
title(['Forma de onda da derivada do pulso glotal de Rosenberg']);  
xlabel({'Tempo (ms)'})  
ylabel('Amplitude')
```

Apêndice B

Algoritmo do pulso glotal com o modelo de LF

```
clear;
clc;
Av=1; %Amplitude máxima do Fluxo Glotal.
Tfs=3; %Tempo de fonação sustentado
F0=262; %Frequência fundamental
T0=1/F0; %Período Fundamental
Oq=0.65; %Quociente aberto
m=(0.77); %Coeficiente de assimetria
Qa=0.1; %Coeficiente da fase de retorno (variando de 0.1 <Qa <0.7)

%-----
%The parameter Epsilon (e) is defined by an implicit equation
syms e
pe= solve ( e*Qa*(1-Oq)*T0 == 1-exp(-e*(1-Oq)*T0),e);
E=double(pe);
%-----
%The parameter (a) is defined by an implicit equation
syms a
x0=pi/(m*Oq*T0);
x1=((a^2)+(x0^2));
x2=exp(-a*Oq*T0);
x3=cos(pi/m);
x4=sin(pi/m);
x5=T0*(1-Oq);
x6=exp(E*T0*(1-Oq));
```

```

x7=1/E;
pa=solve((1/x1)*(a+x0*((x2-x3)/x4))==(x5/(x6-1))-x7,a);
A=double(pa);
%-----
%The parameter (Ee) is obtained by to solve the next equation:
x8= exp(-A*0q*T0);
x9=sin(pi/m);
x10=pi/(m*0q*T0);
x11=(A^2)+(x10^2);
x12=x10*exp(A*m*0q*T0);
Ee=(-Av*x9*x11)/(x8*(x10+x12));
%-----
%Glottal Flow LF Model--First part
%-----
Dt=T0/200;
t1=0:Dt:0q*T0;

y1= Ee*exp(-A*0q*T0);
y2=sin(pi/m)*(A^2+(pi/(m*0q*T0))^2);
y3=pi/(m*0q*T0);
y4=A.*exp(A.*t1).*sin(y3.*t1);
y5=y3.*exp(A.*t1).*cos(y3.*t1);
Ug1=-(y1/y2).*(y3+y4-y5);
%-----
%Glottal Flow LF Model--Second part
%-----
t2=(0q*T0)+Dt:Dt:T0;

y6=1/(E*Qa*(1-0q)*T0);
y7=(1-exp(E*(T0-t2)))/E;
Ug2=-Ee*(y6-1)*(T0-t2+y7);
%-----
%Derivate by Nathalie--First part
%-----
z1=exp(A.*(t1-0q*T0));
z2=sin(y3.*t1);
z3=sin(pi/m);
UGD1=-Ee.*z1.*(z2./z3);
%-----

```

```

%Derivate by Nathalie--Second part
%-----
z4=1/(E*Qa*(1-Q)*T0);
z5=exp(E*Q*T0);
z6=(exp(-E.*t2)-exp(-E*T0));
UGD2=-(Ee*z4)*z5.*z6;
%-----
%Glottal Flow LF Model and Derivate
%-----
ug=[Ug1 Ug2];
UGD=[UGD1 UGD2];
ugd=diff(ug)/Dt;
t=[t1 t2];
%-----
%Figure of Glottal Pulse
%-----
figure (1)
subplot (2,1,1);
plot(t*1000,ug,'k');
title(['Forma de onda do pulso glotal de LF']);
xlabel({'Tempo (ms)'})
ylabel('Amplitude')
%-----
%plot Nathalie's Glottal Pulse Derivate
%-----
subplot (2,1,2);
plot(t*1000,UGD,'b');
title(['Forma de onda da derivada do pulso glotal de LF']);
xlabel({'Tempo (ms)'})
ylabel('Amplitude')

```

Apêndice C

Algoritmo para gerar a excitação glotal

```
clear;
clc;
Av=1; %Amplitude máxima do Fluxo Glotal.
Tfs=3; %Tempo de fonação sustentado
F0=262; %Frequência fundamental
T0=1/F0; %Período Fundamental
Oq=0.65; %Quociente aberto
m=(0.77); %Coeficiente de assimetria
Qa=0.1; %Coeficiente da fase de retorno (variando de 0.1 <Qa <0.7)

%-----
%The parameter Epsilon (e) is defined by an implicit equation
syms e
pe= solve ( e*Qa*(1-Oq)*T0 == 1-exp(-e*(1-Oq)*T0),e);
E=double(pe);
%-----
%The parameter (a) is defined by an implicit equation
syms a
x0=pi/(m*Oq*T0);
x1=((a^2)+(x0^2));
x2=exp(-a*Oq*T0);
x3=cos(pi/m);
x4=sin(pi/m);
x5=T0*(1-Oq);
x6=exp(E*T0*(1-Oq));
```

```

x7=1/E;
pa=solve((1/x1)*(a+x0*((x2-x3)/x4))==(x5/(x6-1))-x7,a);
A=double(pa);
%-----
%The parameter (Ee) is obtained by to solve the next equation:
x8= exp(-A*0q*T0);
x9=sin(pi/m);
x10=pi/(m*0q*T0);
x11=(A^2)+(x10^2);
x12=x10*exp(A*m*0q*T0);
Ee=(-Av*x9*x11)/(x8*(x10+x12));
%-----
%Glottal Flow LF Model--First part
%-----
Dt=T0/200;
t1=0:Dt:0q*T0;

y1= Ee*exp(-A*0q*T0);
y2=sin(pi/m)*(A^2+(pi/(m*0q*T0))^2);
y3=pi/(m*0q*T0);
y4=A.*exp(A.*t1).*sin(y3.*t1);
y5=y3.*exp(A.*t1).*cos(y3.*t1);
Ug1=-(y1/y2).*(y3+y4-y5);
%-----
%Glottal Flow LF Model--Second part
%-----
t2=(0q*T0)+Dt:Dt:T0;

y6=1/(E*Qa*(1-0q)*T0);
y7=(1-exp(E*(T0-t2)))/E;
Ug2=-Ee*(y6-1)*(T0-t2+y7);
%-----
%Glottal Flow LF Model and Derivate
%-----
ug=[Ug1 Ug2];
t=[t1 t2];
%-----
%Comb Dirac
%-----

```

```

bin=fix((Tfs-T0)/(T0));
LD=(length(ug)*bin)+2;
Dirac=zeros(LD,1);
Dirac(1:length(ug):end)=1;
%-----
%Convolution of Glottal Flow and Dirac
%-----
pulse_train=conv(ug,Dirac);
%-----
%Figure of Glottal Pulse
%-----
figure (1)
subplot (2,1,1);
plot(t*1000,ug,'k');
title(['Forma de onda do pulso glotal de LF']);
xlabel({'Tempo (ms)'})
ylabel('Amplitude')
%-----
%plot Excitação glotal
%-----
subplot (2,1,2);
plot((1:3*length(t)), pulse_train(1:3*length(t)),'b');
title(['Excitação glotal de LF']);
xlabel({'Tempo (ms)'})
ylabel('Amplitude')

```

Apêndice D

Algoritmo para a síntese das vogais cantadas com fonte glotal pelo modelo de Rosenberg

```
clear;
clc;
Av=7;    %Máxima amplitude do fluxo glotal
F0=98;   %Frequência fundamental
T0=1/F0; %Periodo fundamental
Tfs=3;   %Tempo de fonação sustentada
nfft=1024;
fm=20000; %Máxima frequência de amostragem
alpha1=60; %Porcentagem do tempo de abertura relativa
alpha2=20; %Porcentagem do tempo de fechamento relativa
Tp=(alpha1/100)*T0; %Tempo da abertura glotal
TN=(alpha2/100)*T0; %Tempo da fechamento glotal
n=round(fm/F0,1,'significant');
%-----
%Rosenberg's Glottal Pulse Approximation
%-----
Ts=T0/n;
fs=1/Ts;
t1=0:Ts:Tp;
t2=(Tp+Ts):Ts:(Tp+TN);
t3=(Tp+TN+Ts):Ts:T0;
t=[t1 t2 t3];
```

```

%-----
x1=cos(pi*t1/Tp);
gn1=(Av/2)*(1-x1);
%-----
x2=pi*(t2-Tp);
gn2=Av*cos(x2/(2*TN));
%-----
gn3=zeros(1,length(t3));
gn=[gn1 gn2 gn3];
%-----
%Rosenberg's Glottal Pulse Derivate. "Forma 1"
%-----
udg=diff(gn)/Ts;
%-----
%Rosenberg's Glottal Pulse Derivate. "Forma 2"
%-----
x3=sin(pi*t1/Tp);
udgn1=(Av*pi/(2*Tp))*x3;
%-----
x4=sin(pi*(t2-Tp)/(2*TN));
udgn2=(-Av*pi/(2*TN))*x4;
%-----
udgn3=zeros(1,length(t3));
%-----
udgn=[udgn1 udgn2 udgn3];
%-----
%Comb Dirac
%-----
bin=fix((Tfs-T0)/(T0));
LD=(length(gn)*bin)+2;
Dirac=zeros(LD,1);
Dirac(1:length(gn):end)=1;
%-----
%Convolution of Glottal Flow and Dirac
%-----
yk=conv(gn,Dirac);
%-----
%Vocal Tract
%-----

```

```

H= input('Enter the vowel 1(/a/),2(/e/),3(/i/),4(/o/),5(/u/):');

formants=[900, 1300, 2000, 2200, 2500;
          450, 1700, 2000, 2200, 2310;
          300, 2100, 2100, 2200, 2490;
          500, 800, 2150, 2200, 2490;
          360, 700, 2170, 2200, 2330];

bandwidths=[41, 52, 70, 32, 100];

%-----
% generate vowel impulse response
%-----

fmts=formants(H,:);

if F0 > fmts(1)
    fmts(1)=F0+10;
end

xin=[1 zeros(1,499)];

    for resonance=1:5
        f=fmts(resonance);
        bw=bandwidths(resonance);
        num=1-2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs)+exp(-4*bw*pi/fs);
        den=[1 -2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs) exp(-4*bw*pi/fs)];
        yout=filter(num, den, xin);
        xin=yout;
    end

%-----
% Filtros que modelam a radiação nos lábios
%-----
%1% Filtro que modela os efeitos de radiação nos lábios
%-----

%Rr=128/(9*pi^2);
%Lr=(8*4)/(3*pi*35000);
%numRz=[2*Rr*Lr/(Rr*(1/fs)+2*Lr) -2*Rr*Lr/(Rr*(1/fs)+2*Lr)];
%denRz=[1 (Rr*(1/fs)-2*Lr)/(Rr*(1/fs)+2*Lr)];
%yout=filter(numRz, denRz, yout);
%-----

```

```

%2% Aproximação aos efeitos de radiação nos lábios
%-----
r=0.95;
B1=[1 -r];
A1=[1];
yout=filter(B1, A1, yout);
%-----
zout=conv(yout,yk);
ys=zout;
sound(ys,fs)
%-----
% convert from frequency and bandwidth to all-pole digital transfer
% function coefficients
%-----
zk=exp(-2*pi*bandwidths/fs);
fkn=cos(2*pi*fmts/fs);
%-----
% evaluate vocal tract log magnitude response; save in HT
HR=[];
HT=zeros(1,nfft);
HP=(complex(ones(1,nfft),zeros(1,nfft)))';
M=5;
for k=1:M
    B=1-2*zk(k)*fkn(k)+zk(k)*zk(k);
    A=[1 -2*zk(k)*fkn(k) zk(k)*zk(k)];
    [Tr,W]=freqz(B,A,nfft,'whole',fs);
    HP=HP.*Tr;
    HL=20*log10(abs(Tr));
    HT=HT+HL';
end
%-----
%plot vocal tract log magnitude response, HT.
%-----
plot(W(1:nfft/2+1),HT(1:nfft/2+1),'r--','LineWidth',2);
title(['Resposta em frequência do trato vocal']);
xlabel('Frequência (Hz)'),ylabel('Magnitudo Log (dB)');
%-----
%Gravando a vogal cantada
%-----

```

```
wavwrite(ys,fs,'baixo98Hz-Rosemb_u_detn.wav');
```

Apêndice E

Algoritmo para a síntese das vogais cantadas com fonte glotal pelo modelo LF

```
clear;
clc;
Av=7; %Amplitude máxima do Fluxo Glotal.
Tfs=3; %Tempo de fonação sustentado
nfft=1024; %The n-point
F0=587; %Frequência fundamental
T0=1/F0; %Período Fundamental
Oq=0.8; %Quociente aberto
m=(0.75); %Coeficiente de assimetria
Qa=0.3; %Coeficiente da fase de retorno (variando de 0.1 <Qa <0.7)
%-----
%The parameter Epsilon (e) is defined by an implicit equation
syms e
pe= solve ( e*Qa*(1-Oq)*T0 == 1-exp(-e*(1-Oq)*T0),e);
E=double(pe);
%-----
%The parameter (a) is defined by an implicit equation
syms a
x0=pi/(m*Oq*T0);
x1=((a^2)+(x0^2));
x2=exp(-a*Oq*T0);
x3=cos(pi/m);
```

```

x4=sin(pi/m);
x5=T0*(1-0q);
x6=exp(E*T0*(1-0q));
x7=1/E;
pa=solve((1/x1)*(a+x0*((x2-x3)/x4))==(x5/(x6-1))-x7,a);
A=double(pa);
%-----
%The parameter (Ee) is obtained by to solve the next equation:
x8= exp(-A*0q*T0);
x9=sin(pi/m);
x10=pi/(m*0q*T0);
x11=(A^2)+(x10^2);
x12=x10*exp(A*m*0q*T0);
Ee=(-Av*x9*x11)/(x8*(x10+x12));
%-----
%Glottal Flow LF Model--First part
%-----
Dt=T0/200;
t1=0:Dt:0q*T0;

y1= Ee*exp(-A*0q*T0);
y2=sin(pi/m)*(A^2+(pi/(m*0q*T0))^2);
y3=pi/(m*0q*T0);
y4=A.*exp(A.*t1).*sin(y3.*t1);
y5=y3.*exp(A.*t1).*cos(y3.*t1);
Ug1=-(y1/y2).*(y3+y4-y5);
%-----
%Glottal Flow LF Model--Second part
%-----
t2=(0q*T0)+Dt:Dt:T0;

y6=1/(E*Qa*(1-0q)*T0);
y7=(1-exp(E*(T0-t2)))./E;
Ug2=-Ee*(y6-1)*(T0-t2+y7);
%-----
%Derivate by Nathalie--First part
%-----
z1=exp(A.*(t1-0q*T0));
z2=sin(y3.*t1);

```

```

z3=sin(pi/m);
UGD1=-Ee.*z1.*(z2./z3);
%-----
%Derivate by Nathalie--Second part
%-----
z4=1/(E*Qa*(1-0q)*T0);
z5=exp(E*0q*T0);
z6=(exp(-E.*t2)-exp(-E*T0));
UGD2=-(Ee*z4)*z5.*z6;
%-----
%Glottal Flow LF Model and Derivate
%-----
ug=[Ug1 Ug2];
UGD=[UGD1 UGD2];
ugd=diff(ug)/Dt;
t=[t1 t2];
%-----
%Comb Dirac
%-----
bin=fix((Tfs-T0)/(T0));
LD=(length(ug)*bin)+2;
Dirac=zeros(LD,1);
Dirac(1:length(ug):end)=1;
%-----
%Convolution of Glottal Flow and Dirac
%-----
pulse_train=conv(ug,Dirac);
%-----
%Vocal Tract
%-----
H= input('Digite 1(/a/),2(/e/),3(/i/),4(/o/),5(/u/):');
fs=1/Dt;

formants=[850, 1300, 2340, 2800, 2930;%%
          400, 2200, 2650, 2800, 2930;
          300, 2500, 2680, 2800, 2960;
          500, 1200, 2350, 2800, 2930;
          350, 1120, 2200, 2800, 2900];

```

```

bandwidths=[90, 110, 380, 60, 450];

%-----
% generate vowel impulse response
%-----

fmts=formants(H,:);

if F0 > fmts(1)
    fmts(1)=F0;
end

xin=[1 zeros(1,499)];

    for resonance=1:5
        f=fmts(resonance);
        bw=bandwidths(resonance);
        num=1-2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs)+exp(-4*bw*pi/fs);
        den=[1 -2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs) exp(-4*bw*pi/fs)];
        yout=filter(num, den, xin);
        xin=yout;
    end

%-----
%Radiation Model
%-----
%Metodo1%
%-----

%Rr=128/(9*pi^2);
%Lr=(8*4)/(3*pi*35000);
%numRz=[2*Rr*Lr/(Rr*(1/fs)+2*Lr) -2*Rr*Lr/(Rr*(1/fs)+2*Lr)];
%denRz=[1 (Rr*(1/fs)-2*Lr)/(Rr*(1/fs)+2*Lr)];
%yout=filter(numRz, denRz, yout);

%-----
%Metodo2% Aproximação aos efeitos de radiação nos lábios
%-----

r=0.95;
B1=[1 -r];
A1=[1];
yout=filter(B1, A1, yout);

%-----

```

```

% convolve vowel impulse response with excitation
%-----
zout=conv(yout,pulse_train);
sound(zout,fs);
%-----
%plot vocal tract log magnitude response, HT.
%-----
%-----
% convert from frequency and bandwidth to all-pole digital transfer
% function coefficients
%-----
zk=exp(-2*pi*bandwidths/fs);
fkn=cos(2*pi*fmts/fs);
%-----
% evaluate vocal tract log magnitude response; save in HT
    HR=[];
    HT=zeros(1,nfft);
    HP=(complex(ones(1,nfft),zeros(1,nfft)))';
    M=5;
    for k=1:M
        B1=1-2*zk(k)*fkn(k)+zk(k)*zk(k);
        A1=[1 -2*zk(k)*fkn(k) zk(k)*zk(k)];
        [Tr,W]=freqz(B1,A1,nfft,'whole',fs);
        HP=HP.*Tr;
        HL=20*log10(abs(Tr));
        HT=HT+HL';
    end

plot(W(1:nfft/2+1),HT(1:nfft/2+1),'r--','LineWidth',2);
xlim([0 6e3])
xlabel('Frequência (Hz)'),ylabel('Magnitude Log (dB)');
%-----
%Gravando a vogal cantada
%-----
%wavwrite(zout, fs, 'mezzo587Hz-LF_u_detn.wav');

```

Apêndice F

Algoritmo para a síntese de uma sequência de vogais em diferentes notas musicais pelo modelo de Rosenberg

```
clc;
clear;
close all;
Tfs=2; %Tempo de fonação sustentada
fs=25000; %Máxima frequência de amostragem
q=1; %q=1 gera um som determinístico
%-----
tons=input('Digite o numero de tons: ');
vogais=tons;
F=zeros(1,tons);
H=zeros(1,vogais);
zout=[];
tt=[0];
g=[];
for k=1:tons
    F(k)=input(['Digite o tom_',num2str(k),' em Hz: ']);
    H(k)=input(['Digite a vogal_',num2str(k),' 1(/a/),2(/e/),3(/i/),4(/o/),5(/u/): ']);
    if F(k)==0
        gn=[];
        pulse=[];
```

```

else
    T=1/F(k);
    [gn,t,udgn]= rosenberg_main (F(k),q,fs);
    bin=fix((Tfs-T)/(T));
    LD=(length(gn)*bin)+2;
    imp_train=zeros(1,LD);
    imp_train(1:length(gn):end)=1;
    pulse=conv(gn,imp_train);
    %-----
    g=[g gn];
    t=tt(end)+t;
    if k==1
        tt=[];
    end
    tt=[tt t];
    %-----
end
%-----
%Vocal Tract
%-----
formants=[900, 1300, 2000, 2200, 2500;%%%%
          450, 1700, 2000, 2200, 2310;
          300, 1900, 2100, 2200, 2490;
          500, 800,  2150, 2200, 2490;
          360, 700,  2170, 2200, 2330];

bandwidths=[41, 52, 70, 32, 100];
%-----
% generate vowel impulse response
%-----
fmts=formants(H(k),:);

if F(k) > fmts(1)
    fmts(1)=F0+10;
end

xin=[1 zeros(1,499)];

for resonance=1:5

```

```

    f=fmts(resonance);
    bw=bandwidths(resonance);
    num=1-2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs)+exp(-4*bw*pi/fs);
    den=[1 -2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs) exp(-4*bw*pi/fs)];
    yout=filter(num, den, xin);
    xin=yout;
end
%-----
%Rr=128/(9*pi^2);
%Lr=(8)/(3*pi*35000);
%numRz=[2*Rr*Lr/(Rr*(1/fs)+2*Lr) -2*Rr*Lr/(Rr*(1/fs)+2*Lr)];
%denRz=[1 (Rr*(1/fs)-2*Lr)/(Rr*(1/fs)+2*Lr)];
%yout=filter(numRz, denRz, yout);
%-----
r=0.95;
B1=[1 -r];
A1=[1];
yout=filter(B1, A1, yout);
zout_v=conv(yout,pulse);
zout=[zout zout_v];

end

sound(zout,fs)
%-----
%Gravando a vogal cantada
%-----
%wavwrite(zout, fs, 'baixo_oitava-Rosenb_aeiou_detn.wav');

Onde rosenberg_main(), corresponde a uma função que gera o pulso glotal
pelo modelo de Rosenberg dado pelo seguinte código:

function [gn,t,udgn]= rosenberg_main (F,q,fs)
Av=7;    %Máxima amplitude do fluxo glotal
alpha1=60; %Porcentagem do tempo de abertura relativa
alpha2=20; %Porcentagem do tempo de fechamento relativo
T0=(1/F)*q; %Periodo fundamental
Tp=(alpha1/100)*T0; %Tempo da abertura glotal
TN=(alpha2/100)*T0; %Tempo da fechamento glotal

```

```

%-----
%Rosenberg's Glottal Pulse Approximation
%-----
Ts=1/fs;
t1=0:Ts:Tp;
t2=(Tp+Ts):Ts:(Tp+TN);
t3=(Tp+TN+Ts):Ts:T0;
t=[t1 t2 t3];
%-----
x1=cos(pi*t1/Tp);
gn1=(Av/2)*(1-x1);
%-----
x2=pi*(t2-Tp);
gn2=Av*cos(x2/(2*TN));
%-----
gn3=zeros(1,length(t3));
gn=[gn1 gn2 gn3];
%-----
%Rosenberg's Glottal Pulse Derivate
%-----
udg=diff(gn)/Ts;
%-----
x3=sin(pi*t1/Tp);
udgn1=(Av*pi/(2*Tp))*x3;
%-----
x4=sin(pi*(t2-Tp)/(2*TN));
udgn2=(-Av*pi/(2*TN))*x4;
%-----
udgn3=zeros(1,length(t3));
%-----
udgn=[udgn1 udgn2 udgn3];
end

```

Apêndice G

Algoritmo para a síntese de uma sequência de vogais em diferentes notas musicais pelo modelo de LF

```
clc;
clear;
close all;
Tfs=2; %Tempo de fonação sustentada
q=1; %q=1 gera um som determinístico
fs=25000; %Máxima frequência de amostragem
%-----
tons=input('Digite o numero de tons: ');
vogais=tons;
F=zeros(1,tons);
H=zeros(1,vogais);
zout=[];
tt=[0];
g=[];
for k=1:tons
    F(k)=input(['Digite o tom_',num2str(k),' em Hz: ']);
    H(k)=input(['Digite a vocal_',num2str(k),' 1(/a/),2(/e/),3(/i/),4(/o/),5(/u/): ']);
    if F(k)==0
        gn=[];
        pulse=[];
    else
        T=1/F(k);
```

```

[gn,t,udgn,Dt]= LF_main_2 (F(k),q,fs);
bin=fix((Tfs-T)/(T));
LD=(length(gn)*bin)+2;
imp_train=zeros(1,LD);
imp_train(1:length(gn):end)=1;
pulse=conv(gn,imp_train);
%-----
g=[g gn];
t=tt(end)+t;
if k==1
    tt=[];
end
tt=[tt t];
%-----

end

%-----
%Vocal Tract
%-----

formants=[900, 1300, 2000, 2200, 2500;%%%%
          450, 1700, 2000, 2200, 2310;
          300, 1900, 2100, 2200, 2490;
          500, 800,  2150, 2200, 2490;
          360, 700,  2170, 2200, 2330];

bandwidths=[41, 52, 70, 32, 100];
%-----
% generate vowel impulse response
%-----
fmts=formants(H(k),:);

if F(k) > fmts(1)
fmts(1)=F0+10;
end

xin=[1 zeros(1,499)];

```

```

for resonance=1:5
    f=fmts(resonance);
    bw=bandwidths(resonance);
    num=1-2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs)+exp(-4*bw*pi/fs);
    den=[1 -2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs) exp(-4*bw*pi/fs)];
    yout=filter(num, den, xin);
    xin=yout;
end
%-----
%Rr=128/(9*pi^2);
%Lr=(8)/(3*pi*35000);
%numRz=[2*Rr*Lr/(Rr*(1/fs)+2*Lr) -2*Rr*Lr/(Rr*(1/fs)+2*Lr)];
%denRz=[1 (Rr*(1/fs)-2*Lr)/(Rr*(1/fs)+2*Lr)];
%yout=filter(numRz, denRz, yout);
%-----
r=0.95;
B1=[1 -r];
A1=[1];
yout=filter(B1, A1, yout);

zout_v=conv(yout,pulse);
zout=[zout zout_v];

end

sound(zout,fs)
%-----
%Gravando a vogal cantada
%-----
wavwrite(zout, fs, 'baixo_oitava-LF_aeiou_detn.wav');

Onde LF_main_2(), corresponde a uma função que gera o pulso glotal pelo
modelo de LF dado pelo seguinte código:

function [gn,t,udgn,Dt]= LF_main_2 (F,q,fs)

clc;
Av=7;      %Amplitude máxima do Fluxo Glotal.

```

```

Oq=0.8; %Quociente aberto
m=(0.75); %Coeficiente de assimetria
Qa=0.3; %Coeficiente da fase de retorno (variando de 0.1 <Qa <0.7)

T0=(1/F)*q; %Período Fundamental

%-----
%The parameter Epsilon (e) is defined by an implicit equation
syms e
pe= solve ( e*Qa*(1-Oq)*T0 == 1-exp(-e*(1-Oq)*T0),e);
E=double(pe);
%-----
%The parameter (a) is defined by an implicit equation
syms a
x0=pi/(m*Oq*T0);
x1=((a^2)+(x0^2));
x2=exp(-a*Oq*T0);
x3=cos(pi/m);
x4=sin(pi/m);
x5=T0*(1-Oq);
x6=exp(E*T0*(1-Oq));
x7=1/E;
pa=solve((1/x1)*(a+x0*((x2-x3)/x4))==(x5/(x6-1))-x7,a);
A=double(pa);
%-----
%The parameter (Ee) is obtained by to solve the next equation:
x8= exp(-A*Oq*T0);
x9=sin(pi/m);
x10=pi/(m*Oq*T0);
x11=(A^2)+(x10^2);
x12=x10*exp(A*m*Oq*T0);
Ee=(-Av*x9*x11)/(x8*(x10+x12));
%-----
%Glottal Flow LF Model--First part
%-----
%Dt=T0/200;
Dt=1/fs;
t1=0:Dt:Oq*T0;

```

```

y1= Ee*exp(-A*0q*T0);
y2=sin(pi/m)*(A^2+(pi/(m*0q*T0))^2);
y3=pi/(m*0q*T0);
y4=A.*exp(A.*t1).*sin(y3.*t1);
y5=y3.*exp(A.*t1).*cos(y3.*t1);
Ug1=-(y1/y2).*(y3+y4-y5);
%-----
%Glottal Flow LF Model--Second part
%-----
t2=(0q*T0)+Dt:Dt:T0;

y6=1/(E*Qa*(1-0q)*T0);
y7=(1-exp(E*(T0-t2)))/E;
Ug2=-Ee*(y6-1)*(T0-t2+y7);
%-----
%Derivate by Nathalie--First part
%-----
z1=exp(A.*(t1-0q*T0));
z2=sin(y3.*t1);
z3=sin(pi/m);
UGD1=-Ee.*z1.*(z2./z3);
%-----
%Derivate by Nathalie--Second part
%-----
z4=1/(E*Qa*(1-0q)*T0);
z5=exp(E*0q*T0);
z6=(exp(-E.*t2)-exp(-E*T0));
UGD2=-(Ee*z4)*z5.*z6;
%-----
%Glottal Flow LF Model and Derivate
%-----
gn=[Ug1 Ug2];
udgn=[UGD1 UGD2];
ugd=diff(gn)/Dt;
t=[t1 t2];
end

```

Apêndice H

Algoritmo para a síntese das vogais cantadas com *jitter* na fonte glotal pelo modelo de Rosenberg

```
clc;
clear;
close all;
q=0.1*randn(1,200)+1; %Variável aleatória
F0=98; %Frequência Fundamental
fs=20000; %Frequência de amostragem
T0=1/F0;
Tfs=3;
Ug=[];
tt=[0];
t0=[];
Udgn=[];
for i=1:200
    [gn,t,udgn]= rosenberg_main (F0,q(i),fs);
    Ug=[Ug gn];
    t0=[t0 t];
    %-----
    t=tt(end)+t;
    if i==1
        tt=[];
    end
    tt=[tt t];
end
```

```

%-----

    Udgn=[Udgn udgn];
end
%-----
%Comb Dirac
%-----
bin=fix((Tfs-tt(end))/(tt(end)));
LD=(length(Ug)*bin)+2;
Dirac=zeros(LD,1);
Dirac(1:length(Ug):end)=1;
tx=[tt (tt(end)+tt)];
%-----
%Convolution of Glottal Flow and Dirac
%-----
yk=conv(Ug,Dirac);
%-----
%Vocal Tract
%-----
H= input('Enter the vowel 1(/a/),2(/e/),3(/i/),4(/o/),5(/u/):');

formants=[900, 1300, 2000, 2200, 2500;
          450, 1700, 2000, 2200, 2310;
          300, 1900, 2100, 2200, 2490;
          500, 800, 2150, 2200, 2490;
          360, 700, 2170, 2200, 2330];

bandwidths=[41, 52, 70, 32, 100];
%-----
% generate vowel impulse response
%-----
fmts=formants(H,:);

if F0 > fmts(1)
    fmts(1)=F0+10;
end

xin=[1 zeros(1,499)];

```

```

    for resonance=1:5
        f=fmts(resonance);
        bw=bandwidths(resonance);
        num=1-2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs)+exp(-4*bw*pi/fs);
        den=[1 -2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs) exp(-4*bw*pi/fs)];
        yout=filter(num, den, xin);
        xin=yout;
    end
%-----
r=0.95;
B1=[1 -r];
A1=[1];
yout=filter(B1, A1, yout);
%-----
zout=conv(yout,yk);
ys=zout;
sound(ys,fs)
%-----
%Plot of Excitation Glottal Pulse
%-----
plot(tx*1000,yk(1:2*length(Ug)), 'r');
title(['Excitation Glottal Pulse In Discrete Time']);
xlabel({'Time (in msec)'})
ylabel('Amplitude')
ylim([0 3])
%-----
%Gravando a vogal cantada
%-----
%wavwrite(ys,fs,'baixo98Hz-Rosemb_a_alet.wav');

```

Onde `rosenberg_main()`, corresponde a uma função que gera o pulso glotal pelo modelo de Rosenberg dado pelo seguinte código:

```

function [gn,t,udgn]= rosenberg_main (F,q,fs)
Av=7;    %Máxima amplitude do fluxo glotal
alpha1=60; %Porcentagem do tempo de abertura relativa
alpha2=20; %Porcentagem do tempo de fechamento relativo
T0=(1/F)*q; %Período fundamental
Tp=(alpha1/100)*T0; %Tempo da abertura glotal

```

```

TN=(alpha2/100)*T0; %Tempo da fechamento glotal
%-----
%Rosenberg's Glottal Pulse Approximation
%-----
Ts=1/fs;
t1=0:Ts:Tp;
t2=(Tp+Ts):Ts:(Tp+TN);
t3=(Tp+TN+Ts):Ts:T0;
t=[t1 t2 t3];
%-----
x1=cos(pi*t1/Tp);
gn1=(Av/2)*(1-x1);
%-----
x2=pi*(t2-Tp);
gn2=Av*cos(x2/(2*TN));
%-----
gn3=zeros(1,length(t3));
gn=[gn1 gn2 gn3];
%-----
%Rosenberg's Glottal Pulse Derivate
%-----
udg=diff(gn)/Ts;
%-----
x3=sin(pi*t1/Tp);
udgn1=(Av*pi/(2*Tp))*x3;
%-----
x4=sin(pi*(t2-Tp)/(2*TN));
udgn2=(-Av*pi/(2*TN))*x4;
%-----
udgn3=zeros(1,length(t3));
%-----
udgn=[udgn1 udgn2 udgn3];
end

```

Apêndice I

Algoritmo para a síntese das vogais cantadas com *jitter* na fonte glotal pelo modelo de LF

```
clear;
clc;
Tfs=3; %Tempo de fonação sustentado
F0=98; %Frequência fundamental
T0=1/F0; %Período Fundamental
q=0.1*randn(1,60)+1; %Variável aleatória
Ug=[];
tt=[0];
t0=[];
Udgn=[];
for i=1:200
    [gn,t,udgn,Dt]= LF_main (F0,q(i));
    Ug=[Ug gn];
    t0=[t0 t];
    %-----
    t=tt(end)+t;
    if i==1
        tt=[];
    end
    tt=[tt t];
    %-----
    Udgn=[Udgn udgn];
end
```

```

end
%-----
%Comb Dirac
%-----
bin=fix((Tfs-tt(end))/(tt(end)));
LD=(length(Ug)*bin)+2;
Dirac=zeros(LD,1);
Dirac(1:length(Ug):end)=1;
tx=[tt (tt(end)+tt)];
%-----
%Convolution of Glottal Flow and Dirac
%-----
pulse_train=conv(Ug,Dirac);
%-----
%Vocal Tract
%-----
H= input('Digite 1(/a/),2(/e/),3(/i/),4(/o/),5(/u/):');
fs=1/Dt;
formants=[900, 1300, 2000, 2200, 2500;
          450, 1700, 2000, 2200, 2310;
          300, 1900, 2100, 2200, 2490;
          500, 800, 2150, 2200, 2490;
          360, 700, 2170, 2200, 2330];

bandwidths=[41, 52, 70, 32, 100];

%-----
% generate vowel impulse response
%-----
fmts=formants(H,:);

if F0 > fmts(1)
    fmts(1)=F0+10;
end

xin=[1 zeros(1,499)];

for resonance=1:5
    f=fmts(resonance);

```

```

    bw=bandwidths(resonance);
    num=1-2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs)+exp(-4*bw*pi/fs);
    den=[1 -2*exp(-bw*2*pi/fs)*cos(2*pi*f/fs) exp(-4*bw*pi/fs)];
    yout=filter(num, den, xin);
    xin=yout;
end
%-----
%Radiation Model
%-----
r=0.95;
B1=[1 -r];
A1=[1];
yout=filter(B1, A1, yout);
%-----
% convolve vowel impulse response with excitation
%-----
zout=conv(yout,pulse_train);
sound(zout,fs);
%-----
%Plot of Glottal Pulse
%-----
figure(1)
plot(tt*1000,Ug,'r');
xlabel({'Tempo (ms)'})
ylabel('Amplitude')
%-----
%Gravando a vogal cantada
%-----
%wavwrite(zout, fs, 'baixo98Hz-LF_u_alet.wav');

```

Onde LF_main(), corresponde a uma função que gera o pulso glotal pelo modelo de LF dado pelo seguinte código:

```

function [gn,t,udgn,Dt]= LF_main (F,q)
clc;
Av=7; %Amplitude máxima do Fluxo Glotal.
Oq=0.8; %Quociente aberto
m=(0.75);%Coeficiente de assimetria
Qa=0.3; %Coeficiente da fase de retorno (variando de 0.1 <Qa <0.7)

```

```

T0=(1/F)*q; %Periodo fundamental com jitter
%-----
%The parameter Epsilon (e) is defined by an implicit equation
syms e
pe= solve ( e*Qa*(1-0q)*T0 == 1-exp(-e*(1-0q)*T0),e);
E=double(pe);
%-----
%The parameter (a) is defined by an implicit equation
syms a
x0=pi/(m*0q*T0);
x1=((a^2)+(x0^2));
x2=exp(-a*0q*T0);
x3=cos(pi/m);
x4=sin(pi/m);
x5=T0*(1-0q);
x6=exp(E*T0*(1-0q));
x7=1/E;
pa=solve((1/x1)*(a+x0*((x2-x3)/x4))== (x5/(x6-1))-x7,a);
A=double(pa);
%-----
%The parameter (Ee) is obtained by to solve the next equation:
x8= exp(-A*0q*T0);
x9=sin(pi/m);
x10=pi/(m*0q*T0);
x11=(A^2)+(x10^2);
x12=x10*exp(A*m*0q*T0);
Ee=(-Av*x9*x11)/(x8*(x10+x12));
%-----
%Glottal Flow LF Model--First part
%-----
Dt=T0/200;
t1=0:Dt:0q*T0;
y1= Ee*exp(-A*0q*T0);
y2=sin(pi/m)*(A^2+(pi/(m*0q*T0))^2);
y3=pi/(m*0q*T0);
y4=A.*exp(A.*t1).*sin(y3.*t1);
y5=y3.*exp(A.*t1).*cos(y3.*t1);
Ug1=-(y1/y2).*(y3+y4-y5);
%-----

```

```

%Glottal Flow LF Model--Second part
%-----
t2=(0q*T0)+Dt:Dt:T0;
y6=1/(E*Qa*(1-0q)*T0);
y7=(1-exp(E*(T0-t2)))/E;
Ug2=-Ee*(y6-1)*(T0-t2+y7);
%-----

%Derivate by Nathalie--First part
%-----
z1=exp(A.*(t1-0q*T0));
z2=sin(y3.*t1);
z3=sin(pi/m);
UGD1=-Ee.*z1.*(z2./z3);
%-----

%Derivate by Nathalie--Second part
%-----
z4=1/(E*Qa*(1-0q)*T0);
z5=exp(E*0q*T0);
z6=(exp(-E.*t2)-exp(-E*T0));
UGD2=-(Ee*z4)*z5.*z6;
%-----

%Glottal Flow LF Model and Derivate
%-----
gn=[Ug1 Ug2];
udgn=[UGD1 UGD2];
ugd=diff(gn)/Dt;
t=[t1 t2];
end

```

Ficha catalográfica automática - SDC/BEE

B275c Barrientos Sandoval, Luis Eduardo
Comparação de modelos de sinal glotal na síntese de vogais, nos casos de vogal sustentada e de voz cantada, considerando sons na língua espanhola / Luis Eduardo Barrientos Sandoval ; Edson Luiz Cataldo Ferreira, orientador. Niterói, 2018.
114 f. : il.

Dissertação (mestrado)-Universidade Federal Fluminense, Niterói, 2018.

1. Voz. 2. Sinal glotal . 3. Processamento de sinais . 4. Canto. 5. Produção intelectual. I. Título II. Cataldo Ferreira, Edson Luiz, orientador. III. Universidade Federal Fluminense. Escola de Engenharia.

CDD -