

Jônatas Cruz Santos

# **Modelagem Matemática-Computacional das Pregas Vocais Aplicada à Síntese de Vozes Alteradas**

São Cristóvão - SE

Julho de 2021

Jônatas Cruz Santos

# **Modelagem Matemática-Computacional das Pregas Vocais Aplicada à Síntese de Vozes Alteradas**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica - PROEE, da Universidade Federal de Sergipe, como parte dos requisitos necessários para a obtenção do título de Mestre em Engenharia Elétrica

Universidade Federal de Sergipe – UFS

Programa de Pós-Graduação em Engenharia Elétrica - PROEE

Orientador: Prof. Jugurta Rosa Montalvão Filho

São Cristóvão - SE

Julho de 2021



**UNIVERSIDADE FEDERAL DE SERGIPE**  
**PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**  
**COORDENAÇÃO DE PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA-PROEE**

## TERMO DE APROVAÇÃO

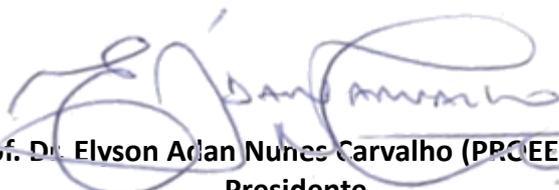
### "Modelagem Matemática-Computacional das Pregas Vocais Aplicada à Síntese de Vozes Alteradas"

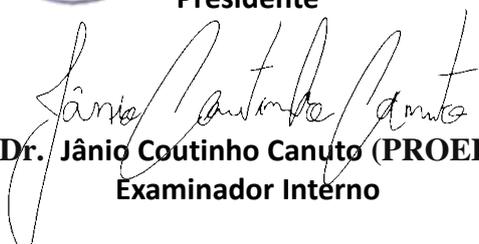
Discente:

**Jônatas Cruz Santos**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Sergipe, como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica.

Aprovada pela banca examinadora composta por:

  
**Prof. Dr. Elvson Adlan Nunes Carvalho (PROEE/UFS)**  
**Presidente**

  
**Prof. Dr. Jânio Coutinho Canuto (PROEE/UFS)**  
**Examinador Interno**

  
**Prof. Dr. Jorge Carlos Lucero (UNB)**  
**Examinador Externo**

  
**Jônatas Cruz Santos**  
**Candidato**

Cidade Universitária "Prof. José Aloísio de Campos", 19 de julho de 2021.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL  
UNIVERSIDADE FEDERAL DE SERGIPE

Santos, Jônatas Cruz  
S237m Modelagem matemática-computacional das pregas vocais aplicada à síntese de vozes alteradas / Jônatas Cruz Santos ; orientador Jugurta Rosa Montalvão Filho. – São Cristóvão, 2021. 96 f.; il.

Dissertação (mestrado em Engenharia Elétrica) – Universidade Federal de Sergipe, 2021.

1. Cordas vocais. 2. Modelos matemáticos. 3. Voz. 4. Distúrbios da voz. I. Montalvão Filho, Jugurta Rosa orient. II. Título.

CDU 621.3

*"A ciência não é um sistema de enunciados certos ou bem estabelecidos, nem é um sistema que avance continuamente em direção a um estado de finalidade.*

*Nossa ciência não é conhecimento (episteme): ela jamais pode proclamar haver atingido a verdade ou um substituto da verdade, como a probabilidade.*

*Não obstante, a ciência tem mais que um simples valor de sobrevivência biológica. Não é tão somente um instrumento útil. Embora não possa alcançar a verdade nem a probabilidade, o esforço por conhecer e a busca da verdade continuam a ser as razões mais fortes da investigação científica."*

*(Karl Popper, A Lógica da Pesquisa Científica)*



---

# Agradecimentos

Agradeço à minha esposa Allana, pelo apoio incondicional, pela companhia, pelo encorajamento, pelas revisões do texto, por todo carinho e por seu amor. Seu incentivo foi essencial para a conclusão dessa dissertação. Eu te amo!

Ao meu orientador, Jugurta, pela confiança depositada, pelo apoio no prosseguimento do trabalho, pelas correções que tornaram meus textos melhores e por tudo que aprendi como aluno e orientando.

Aos membros da banca de avaliação deste trabalho pela disponibilidade, pela boa vontade, pelas observações e contribuições ao trabalho. A Jânio pelas boas ideias passadas debaixo da árvore do DEL e por auxiliar a enxergar a parte pratica de alguns conceitos. A Elyson por cordialmente aceitar o convite de última hora e pela colaboração. Ao professor Lucero pela atenção dispensada e pela prontidão para aceitar o convite para a banca.

À professora Raíssa, que infelizmente não pôde estar na banca, pelas sugestões e contribuições ao trabalho, pela atenção e pelo incentivo.

À professora Ariane Pellicani, pelo disponibilidade ao fornecer amostras, protocolos, avaliações e por ensinar a enxergar melhor o mundo da fonoaudiologia.

Aos meus pais, Jair e Dolores, e à minha irmã, Larissa, pelo apoio ao longo da graduação e do mestrado, pelo companheirismo e amizade. Amo vocês!

A Thiago, Israel, Ítalo, Vitor Magno, João Marcus, Gabriel e aos demais membros do grupo BioChaves pela enriquecedora troca de ideias e colaboração.

A Lucas Molina, Eduardo, Guilherme, Ramírez e aos demais professores do DEL e do PROEE pela contribuição na minha formação e no meu crescimento acadêmico.

A Mayane pela disponibilidade em ajudar a resolver os problemas burocráticos que tive ao longo do mestrado.

A Lucas, Matheus, Iuri (Jury), Vanilson, Neto, Henrique e demais colegas da UFS que colaboraram com essa jornada que está prestes a se encerrar.

Agradeço à CAPES pelo suporte financeiro ao longo do mestrado.

Resumo da Dissertação apresentada ao PROEE/UFS como parte dos requisitos necessários para a obtenção do grau de Mestre (Me.)

MODELAGEM MATEMÁTICA-COMPUTACIONAL DAS PREGAS VOCAIS APLICADA À  
SÍNTESE DE VOZES ALTERADAS

Jônatas Cruz Santos

Julho/2021

Orientador: Prof. Jugurta Rosa Montalvão Filho

Programa: Engenharia Elétrica

**Resumo**

O uso de modelos matemático-computacionais da fonação vem sendo desenvolvido desde a década de 1960, se aperfeiçoando ao longo dos anos. Esses modelos possuem grande utilidade por permitir a reprodução da dinâmica vocal e do comportamento das pregas vocais durante a emissão da voz e da fala, podendo auxiliar os avanços da compreensão dos mecanismos vocais e de suas atividades relacionadas à produção vocal, permitindo também a geração de sinais sintéticos de voz em diferentes condições. Neste trabalho, os elementos anatômicos e fisiológicos envolvidos na fonação e os principais modelos para as pregas vocais e para o trato vocal são apresentados. Com base nisso, uma adaptação de modelos da literatura é proposta, para permitir a representação de vozes com uma patologia vocal (pólipo). Com base nas amostras sintéticas geradas por esse modelo, são feitas conjecturas analíticas, que mostram amplo potencial para o aperfeiçoamento de detectores de pólipo em amostras de voz humana. Também é realizada a implementação de um sintetizador, com base em um modelo bem estabelecido na literatura, que gera amostras de voz sintéticas, sendo controlado por parâmetros musculares e aerodinâmicos da produção vocal, também permitindo o ajuste de condições alteradas da voz. Tal sintetizador pode se mostrar útil, sendo aplicado, por exemplo, no estudo da dinâmica vocal e no auxílio à avaliação vocal por meio de métodos de análise-pela-síntese.

**Palavras-chave:** pregas vocais. modelagem matemática-computacional. alterações vocais.



Abstract of Dissertation presented to PROEE/UFS as a partial fulfillment of the requirements for the degree of Master

MATHEMATICAL-COMPUTATIONAL MODELLING OF VOCAL FOLDS APPLIED TO THE  
SYNTHESIS OF ABNORMAL VOICES

Jônatas Cruz Santos

July/2021

Advisor: Prof. Jugurta Rosa Montalvão Filho

Department: Electrical Engineering

**Abstract**

The use of mathematical-computational models of phonation has been developed since the 60's, being improved throughout the years. These models are very useful by allowing the reproduction of vocal dynamics and of vocal folds behavior during voice and speech emission, being able to help the advances of vocal mechanisms comprehension and its activities related to voice production, also allowing the generation of synthetic voice signals in different conditions. In this work, the anatomical and physiological elements that take part in phonation and the main models for the vocal folds and vocal tract are presented. Based on that, an adaptation of literature models is proposed, allowing the representation of voices with a vocal pathology (vocal polyp). Based on the synthetic samples generated by this model, analytical conjectures are performed, showing a high potential for improving polyp detectors at human voice samples performance. Also, a synthesizer implementation is done, based on a well established model on literature, able to generate synthetic voice samples, being controlled by muscular and aerodynamic parameters of voice production, also allowing the adjust of voice disturbance conditions. That synthesizer might be useful, being applied, for example, at the vocal dynamics study and helping vocal evaluation by means of analysis-by-synthesis methods.

**Keywords:** vocal folds. mathematical-computational modelling. vocal disturbance.



# Sumário

<b>Lista de ilustrações</b> . . . . .	<b>iii</b>	
<b>Lista de tabelas</b> . . . . .	<b>iv</b>	
<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>1</b>
<b>1.1</b>	<b>Objetivos</b> . . . . .	<b>2</b>
<b>1.2</b>	<b>Organização do Trabalho</b> . . . . .	<b>3</b>
<b>2</b>	<b>ANATOMIA E FISIOLOGIA DA PRODUÇÃO VOCAL</b> . . . . .	<b>4</b>
<b>2.1</b>	<b>Sistema Respiratório</b> . . . . .	<b>4</b>
2.1.1	Respiração Durante a Fonação . . . . .	5
<b>2.2</b>	<b>Sistema Fonatório/Laríngeo</b> . . . . .	<b>6</b>
2.2.1	Esqueleto Laríngeo . . . . .	6
2.2.2	Válvulas da Laringe . . . . .	8
2.2.2.1	Modelo Corpo-Cobertura das Pregas Vocais . . . . .	10
2.2.3	Musculatura Laríngeo . . . . .	12
2.2.3.1	Músculos Extrínsecos da Laringe . . . . .	12
2.2.3.2	Músculos Intrínsecos da Laringe . . . . .	12
<b>2.3</b>	<b>Sistema Articulatório-Ressonatório</b> . . . . .	<b>14</b>
2.3.1	Faringe . . . . .	15
2.3.2	Cavidade Oral . . . . .	16
2.3.3	Cavidade Nasal . . . . .	17
2.3.4	Ressonâncias do Trato Vocal . . . . .	18
<b>3</b>	<b>TEORIA MIOELÁSTICA-AERODINÂMICA DA PRODUÇÃO VOCAL</b> .	<b>19</b>
<b>3.1</b>	<b>Modelos Matemático-Computacionais das Pregas Vocais</b> . . . . .	<b>24</b>
3.1.1	Modelos de Uma Massa . . . . .	24
3.1.1.1	Variações do Modelo . . . . .	26
3.1.2	Modelos de Duas Massas . . . . .	26
3.1.2.1	Variações do Modelo . . . . .	29
3.1.3	Modelos <i>Body-Cover</i> . . . . .	30
3.1.3.1	Equacionamento Atualizado e Regras Musculares . . . . .	35

3.1.3.2	Variações do Modelo . . . . .	40
3.1.4	Outros Modelos: Método dos Elementos Finitos . . . . .	41
<b>3.2</b>	<b>Modelos de Propagação de Ondas no Trato Vocal . . . . .</b>	<b>41</b>
3.2.1	Circuito Acústico . . . . .	42
3.2.2	Linha de Transmissão Analógica por Reflexão de Ondas . . . . .	43
<b>4</b>	<b>QUALIDADE VOCAL, DISFONIAS E PATOLOGIAS VOCAIS . . . . .</b>	<b>46</b>
<b>4.1</b>	<b>Qualidade Vocal . . . . .</b>	<b>46</b>
<b>4.2</b>	<b>Voz Normal, Disfonias e Patologias Vocais . . . . .</b>	<b>46</b>
4.2.1	Classificação Etiológica das Disfonias . . . . .	47
<b>4.3</b>	<b>Avaliação da Qualidade Vocal . . . . .</b>	<b>48</b>
<b>4.4</b>	<b>Síntese de Vozes Alteradas . . . . .</b>	<b>48</b>
4.4.1	Síntese Digital de Vozes Alteradas . . . . .	49
4.4.2	Síntese Mecânica-Acústica de Vozes Alteradas . . . . .	50
<b>5</b>	<b>DIMENSÃO INTRÍNSECA E ENTROPIA DIFERENCIAL . . . . .</b>	<b>54</b>
<b>5.1</b>	<b>Método de Montalvão para Análise Visual de Dimensão Intrínseca e Entropia Diferencial . . . . .</b>	<b>54</b>
<b>6</b>	<b>MODELO DAS PREGAS VOCAIS COM UM PÓLIPO VOCAL E APLICAÇÕES POTENCIAIS . . . . .</b>	<b>58</b>
<b>6.1</b>	<b>Modelo <i>Body-Cover</i> das Pregas Vocais com um Pólipo . . . . .</b>	<b>58</b>
6.1.1	Equações de Movimento . . . . .	59
6.1.2	Aerodinâmica Glotal . . . . .	60
6.1.3	Fluxo Glotal e Acoplamento do Trato Vocal . . . . .	61
6.1.4	Regras Musculares e Demais Parâmetros do Modelo . . . . .	61
<b>6.2</b>	<b>Detecção de Pólipo Vocal a partir da Dimensão Intrínseca e Entropia Diferencial . . . . .</b>	<b>62</b>
<b>7</b>	<b>SINTETIZADOR DE VOZ CONTROLADO PELA ATIVAÇÃO MUSCULAR DAS PREGAS VOCAIS . . . . .</b>	<b>67</b>
<b>8</b>	<b>CONCLUSÃO E TRABALHOS FUTUROS . . . . .</b>	<b>72</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>74</b>
	<b>APÊNDICES . . . . .</b>	<b>85</b>
	<b>APÊNDICE A – MODELO <i>BODY-COVER</i> COM MASSA ROTACIONAL . . . . .</b>	<b>86</b>

## Lista de ilustrações

Figura 2.1 – Representação do sistema pulmonar. . . . .	5
Figura 2.2 – Representação do sistema da parede torácica. . . . .	5
Figura 2.3 – Cartilagens do esqueleto laríngeo . . . . .	7
Figura 2.4 – Representação esquemática da estrutura laríngea e da traqueia. . . . .	7
Figura 2.5 – Representação da vista lateral do esqueleto laríngeo. . . . .	7
Figura 2.6 – Cartilagens aritenoideas. . . . .	8
Figura 2.7 – Válvulas da Laringe . . . . .	9
Figura 2.8 – Adução e abdução das pregas vocais . . . . .	9
Figura 2.9 – Camadas das Pregas Vocais . . . . .	10
Figura 2.10–Quatro ajustes das pregas vocais evidenciados por Hirano. $K_b$ representa a rigidez do corpo e $K_c$ representa a rigidez da cobertura. . . . .	11
Figura 2.11–Vista posterior da laringe, incluindo músculos AE, IA e CAP. . . . .	14
Figura 2.12–Vista superior da musculatura intrínseca da laringe. . . . .	14
Figura 2.13–Trato Vocal. . . . .	15
Figura 2.14–Três partes da faringe. . . . .	15
Figura 2.15–Músculos da Faringe. . . . .	15
Figura 2.16–Cavidade oral e seus articuladores. . . . .	16
Figura 3.1 – Diferença de fase vertical na abertura e fechamento glótico. . . . .	20
Figura 3.2 – Ilustração das ressonâncias dos tratos subglotal e supraglotal causados pelo fechamento da glote. As cores mais escuras representam o ar mais comprimido e as mais claras o ar mais diluído. Por efeito de simplificação, os tratos são ilustrados com o mesmo comprimento. . . . .	22
Figura 3.3 – Geometrias convergente e divergente da glote. . . . .	22
Figura 3.4 – Modelo de Uma Massa. . . . .	24
Figura 3.5 – Modelo de Duas Massas. . . . .	27
Figura 3.6 – Circuito Equivalente da Glote . . . . .	29
Figura 3.7 – Modelo Corpo-Cobertura. . . . .	30
Figura 3.8 – Cinco possíveis configurações glotais do modelo corpo-cobertura. . . . .	34
Figura 3.9 – Modelo Corpo-Cobertura com Massa Rotacional de Liljencrants. . . . .	36

Figura 3.10–Discretização do Trato Vocal. . . . .	41
Figura 3.11–Circuito acústico do trato vocal. . . . .	42
Figura 3.12–Ondas de pressão propagadas para frente e para trás. . . . .	44
Figura 5.1 – Ilustração da contagem de coincidências para estimação da entropia quadrática (ou entropia de colisão), que é estimada a partir do logaritmo de $K$ . . . . .	55
Figura 5.2 – Os pontos representam a variável aleatória $X_c$ , que é gerada a partir de uma variável aleatória latente $\mathbf{Z} = [0.1U_1; 2\pi U_2]$ , em que $U_1$ e $U_2$ são variáveis aleatórias uniformes, que assumem valores entre 0 e 1. . . . .	56
Figura 5.3 – Ilustração da estimação conjunta de dimensão intrínseca e entropia diferencial para 3000 instancias da variável $X_c$ . É possível observar que diferentes valores de $r$ podem resultar em estimações diferentes. É possível observar, nesse caso, duas tendências principais. . . . .	56
Figura 6.1 – Modelo <i>Body-Cover</i> das Pregas Vocais com Pólipo Vocal . . . . .	58
Figura 6.2 – Ilustração das áreas $a_i$ e $a_p$ . . . . .	60
Figura 6.3 – Sinais representando instancias de $s(n)$ das amostra com (acima) e sem (abaixo) pólipo. O intervalo de <i>onset</i> da fala e amplitude causada pelo pólipo são indicados. . . . .	62
Figura 6.4 – Os pontos representam o par de coordenadas $(s(n), s(n+\tau f_s))$ para as vogais sintéticas sem (esquerda) e com (direita) a presença de pólipos. . . . .	63
Figura 6.5 – Duas curvas a partir das quais a estimação conjunta da dimensão intrínseca (curves slopes) e da entropia diferencial (distances from corresponding constant slope lines) pode ser obtida para um intervalo de escalas de análise. Para a escala de análise de $2^{-3}$ , a dimensão intrínseca e a entropia diferencial para sinais com e sem pólipo são estimadas como as tangentes dos ângulos $\alpha_0$ e $\alpha_1$ , e as distâncias $h_0$ e $h_1$ , respectivamente. . . . .	65
Figura 6.6 – Ilustração dos pontos correspondentes aos sinais reais de voz de indivíduos com (bolas) e sem (cruzes) a presença de pólipos nas pregas vocais. As coordenadas de cada ponto são dadas pela entropia diferencial e pela dimensão intrínseca estimadas para uma escala de análise de $2^{-3}$ . Dois limiares para a entropia e dimensão intrínseca, $T_h = 1$ e $T_d = 1.75$ respectivamente, são indicados, gerando uma regra simples de detecção de anomalias (pólipos). . . . .	65
Figura 7.1 – Interface do <i>Glottal Generator</i> . . . . .	67
Figura 7.2 – Controles da pressão pulmonar e dos parâmetros musculares, que podem varia ao longo do tempo, de acordo com transição escolhida. . . . .	68
Figura 7.3 – Opções de Síntese. . . . .	68
Figura 7.4 – Controle dos valores para a regra da síntese de vozes alteradas. . . . .	69
Figura 7.5 – Seleção do trato vocal, de acordo com a vogal desejada. . . . .	69
Figura 7.6 – Todas as áreas disponíveis para a síntese de amostras. . . . .	70

Figura 7.7 – Botões utilizados para gerar o sinal, reproduzir seu áudio e salvar os arquivos disponíveis. . . . .	71
Figura 7.8 – Exemplo de sinal gerado pelo simulador. De cima para baixo: voz; fluxo glotal; área glotal; deslocamento das massas. . . . .	71
Figura A.1 – Esboço das configurações utilizadas para o cálculo das pressões na glote: (a) glote convergente; (b) glote divergente; (c) glote com contato na porção superior; (d) glote com contato na porção inferior. . . . .	87



---

## Lista de tabelas

Tabela 2.1 – Músculos Extrínsecos da Laringe. . . . .	12
Tabela 3.1 – Valores para as 4 configurações propostas por Hirano. . . . .	35
Tabela 3.2 – Constantes usadas para o cálculo das Equações (3.77) e (3.78). . . . .	40

**CAPÍTULO****1****Introdução**

A voz e a fala são os principais meios de comunicação utilizados pelo ser humano. O sistema para a produção da voz é um sistema complexo que envolve o trabalho conjunto de diversos órgãos e músculos do corpo humano para permitir a estruturação e a transmissão sonora de informações [1].

Dentre os elementos envolvidos na produção vocal, é possível argumentar que as pregas vocais são os mais importantes [2]. Isso se reforça pelo fato de que a atividade das pregas vocais durante a fonação é a fonte responsável por gerar toda a produção vocal. Dessa maneira, o estudo desse mecanismo e de suas atividades é relevante para a compreensão da formação da voz e para a aplicação em diferentes áreas, como a síntese de voz, a avaliação clínica da saúde vocal e a análise de qualidade vocal [3].

O processo de produção vocal envolve elementos mecânicos das estruturas musculares que trabalham no processo de fonação [4]. Esses elementos mecânicos interagem com as forças aerodinâmicas decorrentes dos fluxos de ar necessários à produção vocal. A teoria mioelástica-aerodinâmica [5, 6], a mais aceita dentre as teorias da produção vocal [7], versa sobre a interação entre os elementos mecânicos e aerodinâmicos que participam da produção vocal, destacando a importância das atividades das pregas vocais. Essa teoria também é a principal base para diversos modelos matemático-computacionais das pregas vocais [6, 8]. Esses modelos se baseiam na teoria mioelástica-aerodinâmica e na mecânica oscilatória das pregas vocais para o estudo e reprodução da dinâmica vocal.

Apesar do benefício do uso de modelos mecânicos estruturados para reproduzir o comportamento das pregas vocais durante a fonação, ainda foram realizados poucos trabalhos empregando esses métodos para a modelagem e síntese de alterações vocais. Entretanto, o uso desses modelos pode constituir uma ferramenta importante para a compreensão dos mecanismos vocais diante de uma alteração de sua normalidade<sup>1</sup> e para a síntese realista de vozes alteradas como ferramenta auxiliar à avaliação da qualidade vocal [10, 11, 12, 13, 14, 15]. Para tais fins, a formulação de modelos que busquem se adaptar à representação de alte-

---

<sup>1</sup>Não existe uma definição amplamente aceita de voz normal [9]. Para os fins deste trabalho, a normalidade de uma voz é definida pela ausência de problemas ou alterações de interesse fonoaudiológico [7].

rações vocais podem ser tidos como de grande utilidade, auxiliando na compreensão do comportamento vocal mediante certas patologias e no seu diagnóstico [16, 17, 18].

A presença de patologias nas pregas vocais pode afetar a regularidade e modificar suas características vocais. A presença de pólipos vocais, por exemplo, modifica a dinâmica vibratória das pregas vocais ao impedir o fechamento total da glote e modificar a massa de uma das pregas, afetando sua vibração [19, 20]. Desta maneira, patologias vocais podem contaminar significativamente os sinais de voz com ruídos aleatórios, que tendem a ocupar espaços de dimensão elevada [21]. Isto sugere a possibilidade da aplicação de métodos de estimação de dimensão intrínseca e entropia diferencial [22] como ferramentas úteis para a análise de patologias vocais [18].

Esta dissertação apresenta conceitos da dinâmica vocal e da teoria mioelástica-aerodinâmica da produção vocal. Com base nesses conceitos, são apresentados diversos modelos propostos para a representação das pregas vocais. Uma adaptação de modelos das pregas vocais, que permite a representação de um pólipo vocal é apresentada e utilizada para o estudo do efeito do pólipo na dimensão intrínseca e na entropia diferencial dos sinais de voz. Utilizando um modelo da literatura como base, um sintetizador de amostras de voz controlado por parâmetros musculares e aerodinâmicos é desenvolvido e apresentado, sendo este uma ferramenta útil para a geração de sinais que auxiliem o estudo da dinâmica vocal e, potencialmente, possa auxiliar no diagnóstico de alterações vocais.

## 1.1 Objetivos

Este trabalho teve como objetivo estudar e implementar modelos matemático-computacionais das pregas vocais e da produção vocal para a síntese de voz. Isso levou à geração de amostras de vogais em condições normais e em condições alteradas, patológicas ou não, tendo em vista a utilidade desses modelos em diversos campos de aplicação.

Num âmbito mais específico, esse trabalho teve como objetivos:

1. Estudar e apresentar os modelos clássicos propostos na literatura para a representação das pregas vocais e do trato vocal;
2. Apresentar um modelo para a representação das pregas vocais com a presença de um pólipo vocal e estudar os efeitos da presença de um pólipo no sinal de voz produzido;
3. Estudar a aplicabilidade da dimensão intrínseca e da entropia diferencial na discriminação de pólipos vocais;
4. Desenvolver um sintetizador de voz com base em modelos da literatura para as pregas vocais e para o trato vocal, com potencial de aplicação no estudo e na avaliação da qualidade vocal.

## 1.2 Organização do Trabalho

Esse trabalho encontra-se organizado de maneira que no **Capítulo 2** são apresentados os fundamentos anatômicos e fisiológicos da produção vocal, sendo apresentados os sistemas respiratório, fonatório e articulatório-ressonatório. Em seguida, no **Capítulo 3**, a Teoria Mioelástica-Aerodinâmica da Produção Vocal [5] é explanada, sendo descrito o processo de produção vocal com base nessa teoria. Também são apresentados, nesse capítulo, os modelos matemático-computacionais clássicos propostos, com base na teoria, para o processo oscilatório das pregas vocais durante a produção vocal [23, 24, 25] e as duas principais abordagens para modelar a propagação de ondas no trato vocal é descrita [26, 27]. Os conceitos de qualidade vocal, disfonias e patologias vocais são apresentados no **Capítulo 4**, sendo nele feita também uma revisão bibliográfica da síntese de vozes alteradas, em que os principais trabalhos implementados, os métodos aplicados e seus resultados são brevemente descritos. O **Capítulo 5** explana brevemente os conceitos de dimensão intrínseca e entropia diferencial, apresentando o método visual de Montalvão [22] para estimar as duas medidas conjuntamente. Um modelo das pregas vocais com a presença de um pólipó é descrito no **Capítulo 6**, sendo as amostras utilizadas para a propor conjectura analíticas que levam ao emprego da dimensão intrínseca e da entropia diferencial para a detecção de pólipos em amostras de voz humana. No **Capítulo 7** é apresentado um sintetizador, desenvolvido com base em um modelo difundido na literatura, que permite gerar amostras de acordo com a atividade muscular da laringe [28]. Os recursos e a implementação desse sintetizador são descritos também nesse capítulo. Por fim, no **Capítulo 8**, é feita a conclusão do trabalho e são apresentados possíveis trabalhos futuros.

## CAPÍTULO

## 2

---

# Anatomia e Fisiologia da Produção Vocal

A produção vocal é um processo complexo que envolve vários elementos do corpo humano em várias etapas. Apesar de se tratar de um processo único, em que as etapas são interdependentes, para um melhor entendimento de seus elementos, esse processo pode ser analisado a partir da divisão em três sistemas: respiratório (ou subglotal), fonatório/laríngeo (ou glotal) e articulatório-ressonatório (ou supraglotal) [7]. É importante ressaltar, entretanto, a natureza didática dessa divisão sistemática, tendo em vista que alguns órgãos, cavidades e músculos fazem parte de mais de um dos sistemas.

## 2.1 Sistema Respiratório

O sistema respiratório é o sistema responsável por obter, por meio da inspiração, e distribuir, por meio da expiração, o ar necessário para que ocorra a fonação. Durante a produção vocal, esse sistema é capaz de interferir no controle da frequência fundamental, da intensidade e da ênfase dos sons emitidos.

Comumente, o sistema respiratório é separado em dois: o sistema pulmonar e o sistema da parede torácica [29]. O sistema pulmonar, ilustrado na Figura 2.1, engloba os órgãos respiratórios e é composto pelo pulmão e pelas vias aéreas (brônquios, traqueia, laringe, faringe, cavidades orais e cavidades nasais), enquanto o sistema da parede torácica, ilustrado na Figura 2.2, se relaciona à estrutura óssea e à musculatura, sendo composto pela caixa torácica, pelo diafragma e pelo abdome [7].

Os pulmões são estruturas de alta elasticidade, armazenadas no interior da cavidade torácica, que contêm em seu interior os alvéolos, bronquíolos e brônquios, que são responsáveis pelas trocas gasosas com o organismo [31]. Conectada aos pulmões encontra-se a traqueia, um tubo retilíneo, composto normalmente por de 16 a 20 anéis cartilagosos, medindo cerca de 11 cm de comprimento e 2,5 cm de diâmetro [7]. A traqueia faz a ligação entre o pulmão e a laringe, transportando o fluxo de ar pulmonar necessário para que ocorra a fonação.

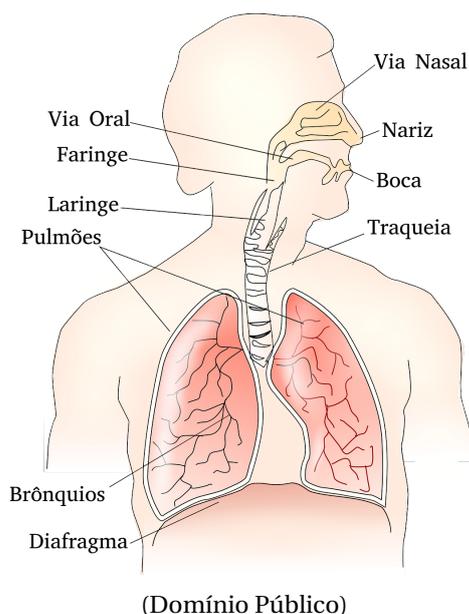


Figura 2.1 – Representação do sistema pulmonar.

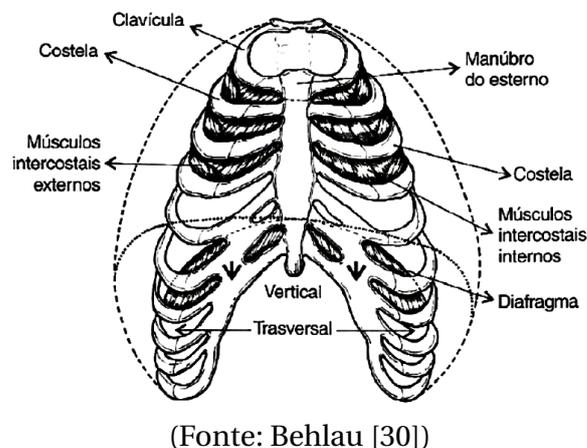


Figura 2.2 – Representação do sistema da parede torácica.

No sistema da parede torácica encontra-se a musculatura essencial para que a respiração ocorra. Os principais músculos que participam do processo da respiração são o diafragma, os onze pares de músculos intercostais externos, os onze pares de músculos intercostais internos e os quatro músculos abdominais, a saber, o reto abdominal, o oblíquo externo, o oblíquo interno e o abdominal transversal [31].

Os intercostais externos e o diafragma aumentam o volume da caixa torácica e dos pulmões quando se contraem e, quando relaxam, a elasticidade do pulmão, a gravidade e o torque das costelas fazem com que o volume se reduza ao seu tamanho original [32]. Em contraste, os intercostais internos e os abdominais realizam o processo oposto, forçando a redução do volume pulmonar durante sua contração.

Para a respiração vital é necessária apenas a contração e o relaxamento do diafragma e dos intercostais externos. Entretanto, na respiração para a produção vocal, na etapa da expiração, é de suma importância a participação dos abdominais e dos intercostais internos. A respiração para a vida e a respiração para a fala também diferem significativamente em outros aspectos: a respiração para a vida costuma ocorrer a partir do nariz e a para a fala ocorre pela boca; a taxa de inspiração por expiração é muito menor na respiração para a fala; o volume de ar nos pulmões é significativamente maior na respiração para a fala.

### 2.1.1 Respiração Durante a Fonação

O processo de respiração para a produção vocal ocorre, nas etapas de inspiração e expiração, da seguinte maneira:

- **Inspiração** - o processo de inspiração é iniciado com a contração do diafragma e dos músculos intercostais externos. A contração desse músculo provoca o aumento do volume dos pulmões e, por consequência, ocasiona a redução de sua pressão interna. A pressão é reduzida para valores inferiores à pressão atmosférica (cerca de 1 a 2 cmH<sub>2</sub>O abaixo [7]), proporcionando a entrada de ar nos pulmões, até que a pressão se equilibre com a pressão atmosférica. Nos casos em que é necessário um volume maior de ar nos pulmões, outros músculos do sistema da parede torácica também se contraem para forçar uma maior entrada de ar [31].
- **Expiração** - na expiração para a fala, o relaxamento dos músculos da inspiração e a redução natural do volume pulmonar por meio do torque das costelas, da gravidade e da elasticidade dos pulmões não são suficientes [32]. É desejável um maior aproveitamento do volume de ar contido dentro dos pulmões. Para isso, o diafragma e os músculos intercostais externos relaxam controladamente, para que o ar seja expelido paulatinamente. A redução do volume pulmonar eleva a pressão pulmonar para valores acima da pressão atmosférica, fazendo com que o ar seja expelido dos pulmões. É necessário que a pressão do ar na laringe, abaixo da glote, esteja na faixa de 5 a 10 cmH<sub>2</sub>O para que ocorra a fonação [7]. Ao atingir um volume de ar de cerca de 38% da capacidade vital<sup>1</sup> dos pulmões, um estado de equilíbrio é atingido nos pulmões. Assim, é necessária a ativação dos músculos abdominais e dos músculos intercostais internos para continuar a compressão pulmonar, reduzindo sua pressão interna e mantendo a produção vocal.

## 2.2 Sistema Fonatório/Laríngeo

O ar expelido dos pulmões durante a expiração, após passar pela traqueia, chega à laringe, órgão onde ocorre a fonação e que constitui o sistema fonatório. Localizada acima da traqueia (conectada por uma membrana denominada cricotraqueal) e abaixo da faringe, a laringe exerce, primordialmente<sup>1</sup>, três funções no corpo humano: proteção das vias aéreas, respiração e fonação. Para o entendimento do funcionamento da laringe para a fonação, é importante a compreensão de suas estrutura básicas: osso, cartilagens, músculos e válvulas.

### 2.2.1 Esqueleto Laríngeo

A laringe possui um esqueleto ossocartilaginoso, composto essencialmente pelo osso hioide, por cartilagens únicas (tireóidea, cricóidea e epiglote) e por pares de cartilagens (aritenóideas, cuneiformes, corniculadas, sesamoides e tritíceas). O osso hioide tem a função de dar a sustentação para as cartilagens da laringe, se conectando diretamente à cartilagem tireóidea

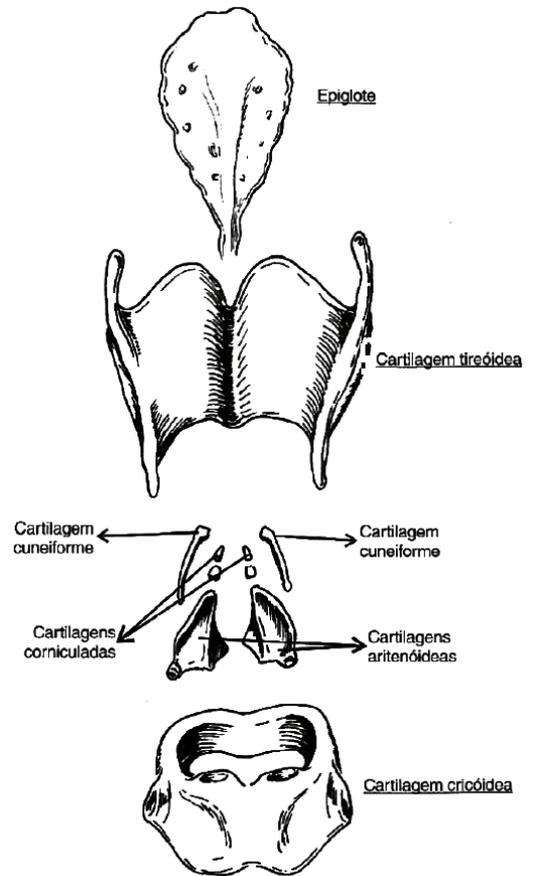
---

<sup>1</sup> Maior quantidade de ar possível de ser inspirada ou expirada pelos pulmões

por meio de uma membrana fina denominada hiotireóidea, e formar um ponto de conexão para a língua [7]. As principais cartilagens laríngeas encontram-se ilustradas na Figura 2.3 e nas Figuras 2.4 e 2.5 são dispostas representações conexas do esqueleto laríngeo.

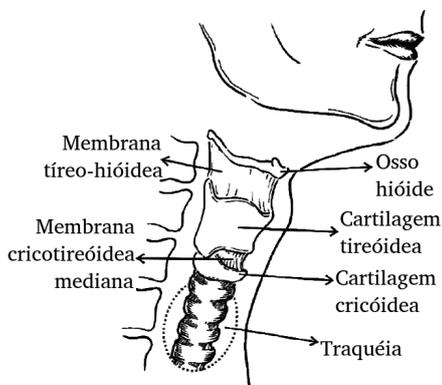
A cartilagem tireóidea é maior cartilagem da laringe, sendo composta por duas lâminas laterais, dois cornos longos projetados para cima e dois cornos curtos projetados para baixo. As lâminas laterais se unem na porção frontal, formando uma saliência denominada pomo de adão, de maneira que as pregas vocais se conectam à superfície interna dessa região saliente. O ângulo da união entre as lâminas varia entre a população masculina e feminina, determinando a diferença entre os sexos da protuberância externa do pomo de adão e o tamanho das pregas vocais [31] (quanto mais protuberante o pomo de adão, mais longas tendem a ser as pregas vocais). Na porção posterior a cartilagem tireóidea é aberta, não havendo contato entre as lâminas laterais.

A cartilagem cricóidea conecta-se com os cornos inferiores da cartilagem tireóidea e com a traqueia (através da membrana cricotraqueal). Essa cartilagem possui um formato de anel fechado, sendo mais estreito em sua porção anterior e alargado e elevado em sua porção posterior. O sexo também interfere no formato da cartilagem cricóidea, de forma que esta tende a ser um anel mais circular em mulheres e



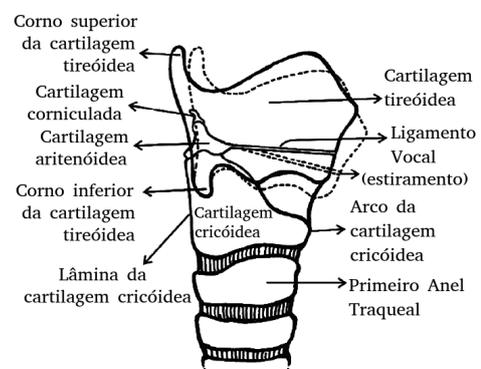
(Fonte: Behlau [30])

Figura 2.3 – Cartilagens do esqueleto laríngeo



(Fonte: Behlau [30])

Figura 2.4 – Representação esquemática da estrutura laríngea e da traqueia.



(Fonte: Behlau [30])

Figura 2.5 – Representação da vista lateral do esqueleto laríngeo.

mais oval em homens [31].

A cartilagem epiglote possui o formato similar ao de uma folha, partindo da região posterior da cartilagem tireóidea até a porção superior da laringe, acima do osso hioide, com o qual a epiglote também se conecta. A cartilagem epiglote exerce o papel de fechamento da glote durante o processo de deglutição, de modo que a epiglote fecha a laringe e direciona o alimento para o esôfago.

As cartilagens aritenoideas são as cartilagens que exercem um maior impacto no processo de fonação. Essas cartilagens são posicionadas sobre a região posterior da cartilagem cricóidea e possuem duas projeções partindo de sua base: o processo vocal e o processo muscular, como ilustrado na Figura 2.6. O processo vocal é projetado na direção anterior e forma o ponto em que a prega vocal é conectada. O processo muscular, por sua vez, é projetado lateralmente e posteriormente, sendo esse processo um ponto de contato com alguns músculos que interferem no processo da fonação.

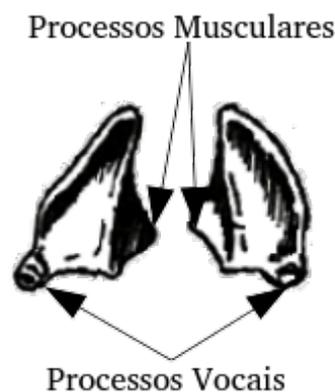


Figura 2.6 – Cartilagens aritenoideas.

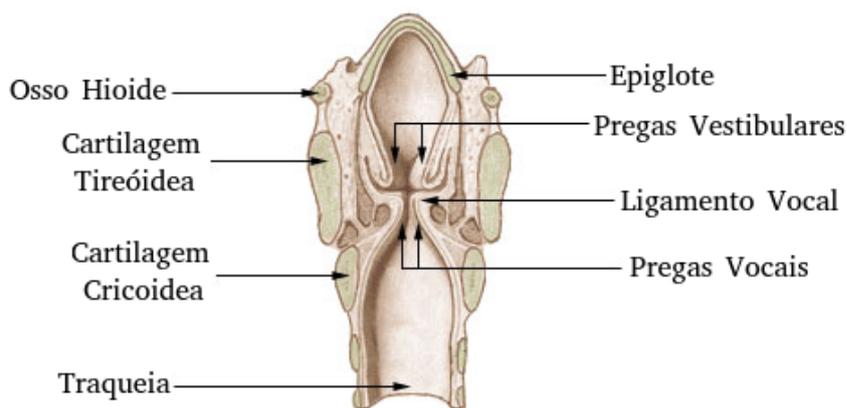
A laringe possui outras cartilagens menores e que apresentam menor impacto no funcionamento da laringe. Essas cartilagens são denominadas cartilagens acessórias [31], sendo elas as cuneiformes, corniculadas, sesamoides e tritíceas.

### 2.2.2 Válvulas da Laringe

A estrutura interna da laringe possui a forma de um tubo contendo três válvulas internas que se abrem ou se fecham para desempenhar suas funções essenciais. Essas válvulas são compostas por tecidos conectivos e músculos nomeadas como pregas, sendo elas as pregas ariepiglóticas, pregas vestibulares e as pregas vocais. As válvulas são ilustradas na Figura 2.7.

As pregas ariepiglóticas são válvulas localizadas na parte superior da laringe, próxima a epiglote. Essas pregas atuam com os músculos que fazem parte de sua composição, se contraindo para puxar a epiglote para trás e fechar a laringe durante a deglutição e para a proteção das vias respiratórias. Essas pregas também se contraem, parcialmente, na fala de alguns idiomas e em alguns estilos de canto [31]. Na Figura 2.7 as pregas ariepiglóticas não são visíveis, porém localizam-se na porção lateral da epiglote.

As pregas vestibulares, também conhecidas como falsas pregas vocais ou como pregas ventriculares [33], se assemelham às pregas vocais, entretanto possuem uma musculatura menos rica. Essas pregas se fecham durante a deglutição, como um proteção extra para as vias

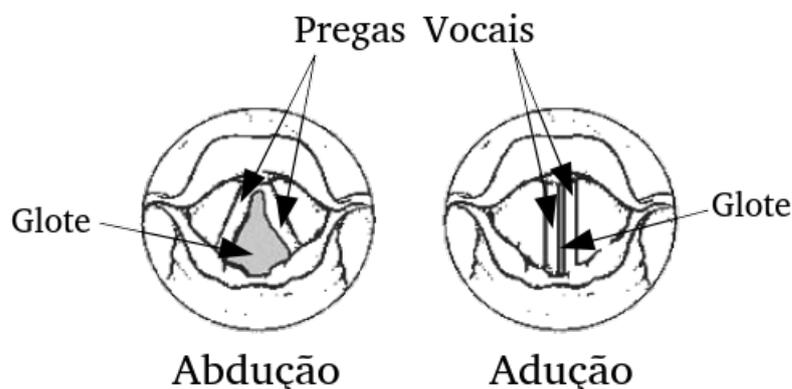


(Domínio Público)

Figura 2.7 – Válvulas da Laringe

respiratórias, e durante atividades de esforço. Sua participação na fonação ocorre em casos patológicos [7], porém há estudos que sugerem uma ação importante das pregas ventriculares na produção da voz sussurrada [34]. As pregas vestibulares se localizam pouco acima das pregas vocais, como ilustrado na Figura 2.7, separadas por um pequeno espaço denominado ventrículo laríngeo. No ventrículo e nas pregas há uma grande quantidade de glândulas que produzem secreção mucosa, mantendo as pregas vocais lubrificadas.

As pregas vocais são as válvulas mais complexas da laringe. Elas são acopladas aos processos vocais das cartilagens aritenoides na região posterior e à cartilagem tireóidea na região anterior. A abertura formada entre as pregas vocais é denominada glote, sendo ela dividida em duas porções: porção membranosa e a porção cartilaginosa [7]. A porção cartilaginosa da glote compreende os primeiros dois quintos da porção posterior da glote. Os três quintos restantes da glote, mais próximos à porção anterior, formam a porção membranosa da glote. Os processos de abertura da glote e de fechamento da glote, por controle muscular, são denominados respectivamente como abdução e adução, como ilustrado na Figura 2.8.



(Domínio Público)

Figura 2.8 – Adução e abdução das pregas vocais

Cada prega vocal possui uma estrutura de seis camadas, sendo elas:

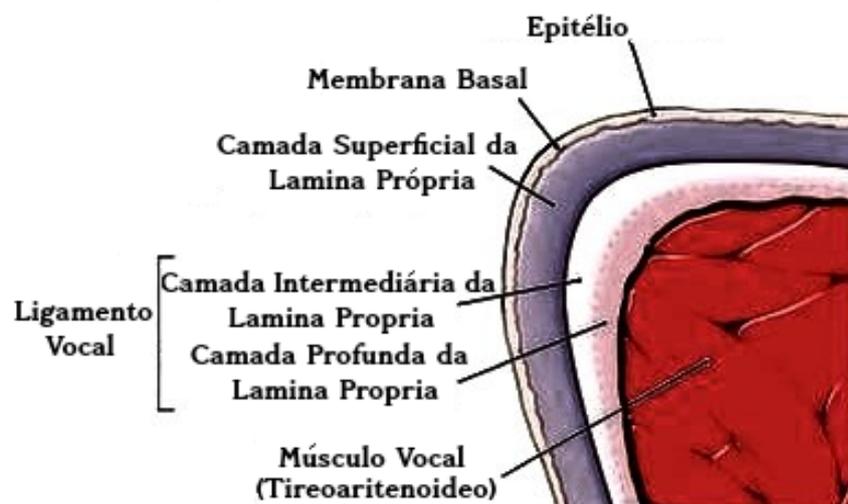
- O músculo vocal (ou tireoaritenoideo), localizado em sua região mais interna;
- Três camadas de uma membrana mucosa denominada Lâmina Própria (camadas profunda, intermediária e superficial), que se encontram cobrindo o músculo vocal;
- Uma zona membranosa, a membrana basal, responsável por proporcionar a aderência entre a lâmina própria e a camada mais externa da prega vocal;
- A camada mais externa, composta de epitélio.

A estrutura em camadas das pregas vocais encontra-se ilustrada na Figura 2.9.

### 2.2.2.1 Modelo Corpo-Cobertura das Pregas Vocais

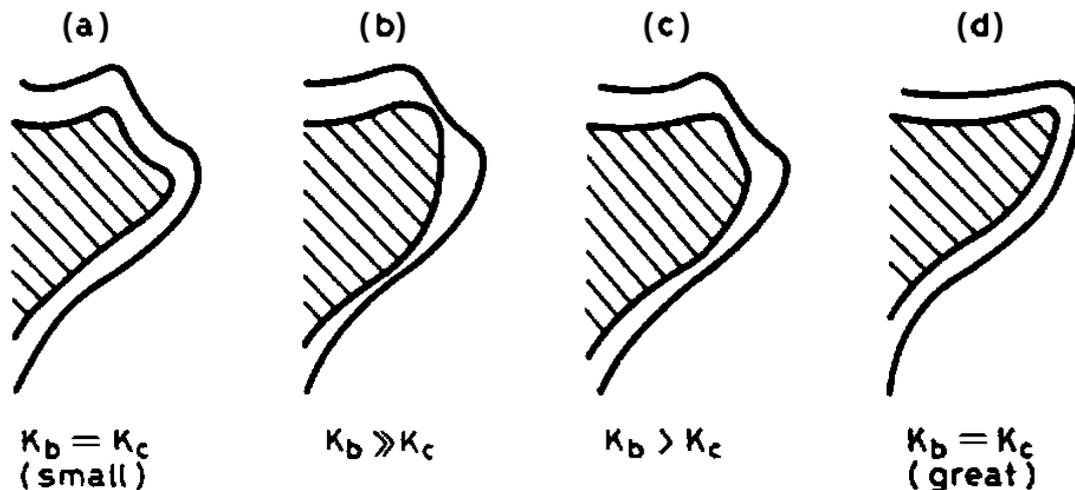
Utilizando dados de eletromiografias e vídeos em alta velocidade de imagens da laringe, M. Hirano [35] observou uma diferença na vibração das diferentes camadas das pregas vocais durante a fonação. Nessas observações, constatou-se que havia uma vibração muito mais significativa do epitélio e da camada superficial da lâmina própria (também conhecida como Espaço de Reinke). Com base nesse fato, Hirano propôs o modelo *Cover-Body* (ou Corpo-Cobertura) das pregas vocais.

Com esse modelo, as camadas das pregas vocais seriam separadas, de acordo com a sua rigidez, em duas estruturas. A **Cobertura** seria a porção mais flexível das pregas e mais suscetível a vibrações, sendo composta pelo epitélio e pela camada superficial.



(Domínio Público)

Figura 2.9 – Camadas das Pregas Vocais



(Adaptado de Hirano [35])

Figura 2.10 – Quatro ajustes das pregas vocais evidenciados por Hirano.  $K_b$  representa a rigidez do corpo e  $K_c$  representa a rigidez da cobertura.

vocal e as camadas intermediária e profunda da lamina própria compoariam o **Corpo**, a porção mais rígida e mais resistente a vibrações das pregas vocais.

Hirano descreveu quatro ajustes típicos da laringe observados em seu estudo, que encontram-se ilustrados na Figura 2.10. Nela, o caso (a) representa o ajuste das pregas em uma fonação suave de frequência baixa, caracterizada por uma baixa rigidez equilibrada entre corpo e cobertura. No caso (b), o ajuste proporciona a produção de uma voz pesada de média frequência, sendo observada uma rigidez muito maior no corpo que na cobertura. O caso (c), representa o ajuste para uma emissão padrão da voz, sendo caracterizada por uma rigidez moderadamente mais alta no corpo que na cobertura. Já o caso (d) é observado na emissão de falsetes e em registros mais agudos, sendo marcante a presença de uma rigidez no corpo e na cobertura igualmente elevadas.

Esse modelo foi atualizado posteriormente com duas principais alterações: a criação de uma camada de transição, tendo em vista que mais observações permitiram constatar que as camadas intermediária e profunda das pregas vocais possuíam uma significativa menor rigidez que os músculos vocais, e a inclusão da recém descoberta membrana basal [36, 37]. Dessa forma, o modelo corpo-cobertura estruturou-se da seguinte maneira [37]:

- **Cobertura** - Epitélio, Membrana Basal e Camada Superficial da Lâmina Própria;
- **Transição** - Camadas Intermediária e Profunda da Lâmina Própria (a junção dessas camadas também é denominada como Ligamento Vocal [30]);
- **Corpo** - Músculo Vocal (ou porção interna do músculo Tireoaritenoideo).

Tabela 2.1 – Músculos Extrínsecos da Laringe.

(Adaptado de Behlau [30])

Músculo		Ação Principal
Infra-hioides	Esterno-hioideo	Abaixa o osso hioide
	Esternotireóideo	Abaixa a cartilagem tireóidea
	Tíreo-hioideo	Aproxima a cartilagem tireóidea e o osso hioideo
	Omo-hioideo	Abaixa e retrai o osso hioideo
Supra-hioides	Estilo-hioideo	Eleva e retrai o osso hioideo
	Digástrico	Eleva o hioideo e deprime a mandíbula
	Milo-hioideo	Eleva e projeta o hioideo e a língua
	Gênio-hioideo	Puxa a língua e o hioideo para a frente

### 2.2.3 Musculatura Laríngea

Os músculos da laringe são divididos em dois agrupamentos: os músculos extrínsecos e os músculos intrínsecos. Os músculos extrínsecos da laringe são músculos que possuem uma inserção na laringe mas se conectam a uma outra estrutura, externa à laringe, enquanto os músculos intrínsecos são os músculos que possuem sua origem e inserção apenas na laringe [7].

#### 2.2.3.1 Músculos Extrínsecos da Laringe

A musculatura extrínseca da laringe não apresenta um interferência direta na fonação, mas modificam a laringe proporcionando alterações indiretas no processo de produção vocal [31], afetando principalmente as configurações do trato vocal. Os músculos extrínsecos se dividem em dois agrupamentos: o músculos infra-hioides e os músculos supra-hioides.

Os músculos infra-hioides se conectam na porção inferior do osso hioide e, ao se contrair, exercem a função de mover a laringe para baixo. De maneira análoga, os músculo supra-hioides se conectam na porção superior do osso hioide e exercem a função de elevar a laringe ao se contrair. Os principais músculos extrínsecos da laringe e as ações exercidas durante suas respectivas contrações são apresentados na Tabela 2.1, de acordo com o seu agrupamento.

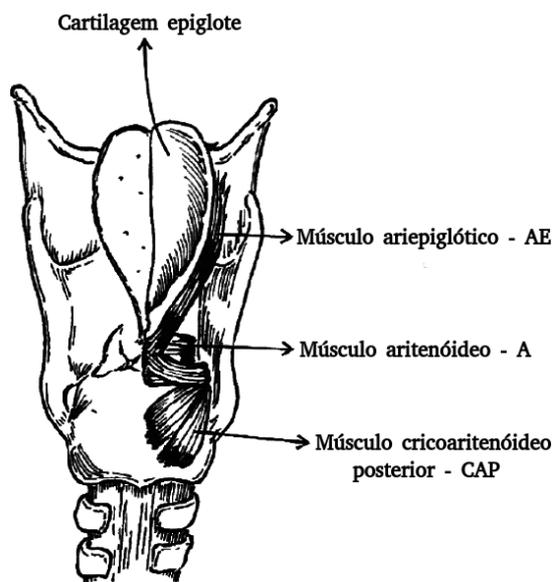
#### 2.2.3.2 Músculos Intrínsecos da Laringe

A musculatura intrínseca da laringe exerce uma influência mais direta e significativa durante o processo de fonação. Os principais músculos intrínsecos para a fonação exercem a função de abduzir, aduzir e/ou tensionar as pregas vocais durante a sua contração. Os principais músculos intrínsecos da laringe são os seguintes:

- **Cricoaritenoídeos Posteriores** - O par de músculos cricoaritenoídeos posteriores (CAP ou PCA) partem da região posterior da cartilagem cricoide e se conectam com os processos musculares das cartilagens aritenoídeas. A contração desses músculos faz com que as aritenoídeas girem posteriormente, proporcionando a abdução das pregas vocais. A ação desses músculos é essencial durante a respiração, abrindo a glote para a passagem do ar. Entretanto, durante a fala, a ação dos CAP é útil para a produção de alguns sons consonantais e para a emissão de sons mais soprosos, como a fala sussurrada. Existem também indícios de que uma leve contração dos CAP corrobore com a estabilidade da produção vocal [33].
- **Cricoaritenoídeos Laterais** - Os músculos pares cricoaritenoídeos laterais (CAL ou LCA) partem da borda lateral da cricóideia e conectam também com os processos musculares da aritenoídeas. A contração desses músculos desloca anteriormente os processos musculares, provocando o efeito de aproximar os processos vocais e, por consequência, aduzir as pregas vocais. Os músculos CAL são os principais adutores das pregas vocais, porém não proporcionam o seu fechamento completo, pois sua ação permite a adução completa somente da glote membranosa [7]. Entretanto, não é necessário o fechamento completo da glote para que haja a produção vocal [31].
- **Interaritenoídeo** - O músculo interaritenoídeo (IA), também conhecido como aritenoídeo ou ari-aritenoídeo, é um músculo que se divide em duas porções: a porção oblíqua, que possui duas partes que se originam, cada uma, na base de uma cartilagem aritenoídea e se insere no ápice da outra aritenoídea, e a porção transversa, que se insere nas regiões posteriores de ambas as aritenoídeas. A contração do IA exerce a função básica de aproximar as aritenoídeas, de modo que a contração da porção transversa aproxima os ápices e a contração da porção oblíqua aproxima as bases. Isso permite que ocorra a adução da glote cartilaginosa. Assim como os CAP, o músculo IA possui uma provável participação no controle da estabilidade da fonação [33].
- **Cricotireóídeos** - Os músculos cricotireóídeos (CT) são pares de músculos que também se dividem em duas porções: *pars recta* (PR) e *pars oblique* (PO) [7]. Ambas as porções partem da borda lateral da cricóideia, de modo que a porção PR, em uma inclinação quase vertical, se insere na borda inferior da tireóideia e a porção PO se insere na região anterior do corno inferior da tireóideia. A contração do CT faz com que a tireóideia se abaixe, se aproximando da cricóideia. Como um dos pontos de conexão das pregas vocais se encontra na tireóideia, esse movimento faz com que as pregas se estiquem e, por consequência, se tensionem. Na fonação, isso possui o efeito de aumentar a frequência fundamental da vibração glótica, de modo que os CT são os principais músculos empregados no controle da frequência fundamental [38].
- **Tireoaritenoídeos** - Os músculos tireoaritenoídeos (TA), ou músculos vocais, são os

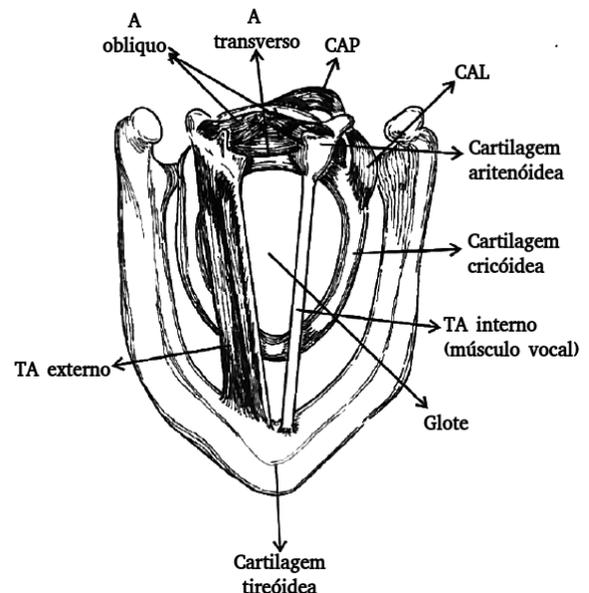
pares de músculos que compõem a estrutura das pregas vocais, sendo o corpo do modelo corpo-cobertura. Os TA possui dois feixes principais, sendo um dele mais lateral, conhecido como tireomuscular ou TA externo, e outro mais medial, conhecido com tireovocal ou TA interno. Os músculos TA internos partem do processos vocais, enquanto os TA externos partem dos processos musculares e ambos se inserem na tireóideia, mantendo a unidade muscular. Há também um feixe superior do TA que se estende até as pregas vestibulares e participa da medialização delas. A compressão do TA interno tensiona as pregas vocais, ao passo que a compressão do TA externo as encurta e relaxa.

Além dos principais músculos intrínsecos, existem também os músculos ariepiglóticos (AE) e os músculos tireoepiglóticos (TE) que atuam, respectivamente, para o fechamento e para a abertura da laringe por meio da epiglote. Os músculos intrínsecos da laringe encontram-se representados nas Figuras 2.11 e 2.12.



(Fonte: Behlau [30])

Figura 2.11 – Vista posterior da laringe, incluindo músculos AE, IA e CAP.



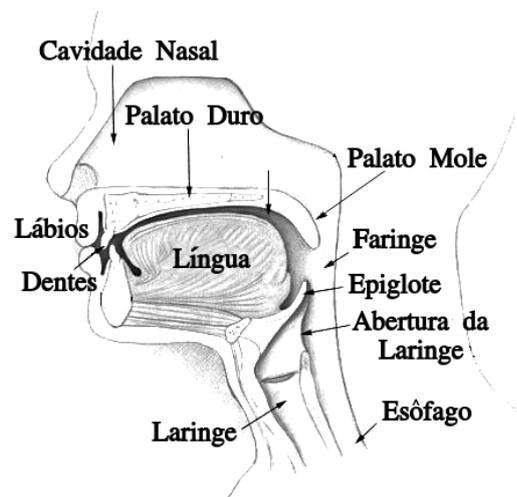
(Fonte: Behlau [30])

Figura 2.12 – Vista superior da musculatura intrínseca da laringe.

## 2.3 Sistema Articulatório-Ressonatório

O sistema articulatório-ressonatório consiste, essencialmente, no trato vocal, nos seus articuladores e nos músculos que se ativam para a atuação dos articuladores. O trato vocal é um tubo curvado e irregular com um formato variável e é composto por cavidades ressonatórias, partindo da região superior à laringe até aos lábios. Nesse tubo, que possui o comprimento de cerca de 17 cm em homens, de 14 a 15 cm em mulheres e de 8 a 9 cm em crianças, o som

produzido pelas pregas vocais na laringe é modificado através de constrictões executadas nesse tubo por seus articuladores. As principais cavidades ressonatórias do trato vocal são a faringe, a cavidade oral e a cavidade nasal (que pode ser acoplada e desacoplada ao trato). As cavidades do trato vocal e seus principais articuladores encontram-se representados na Figura 2.13.

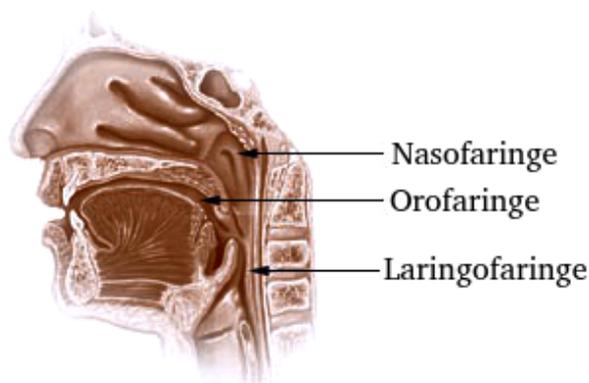


(Domínio Público)

Figura 2.13 – Trato Vocal.

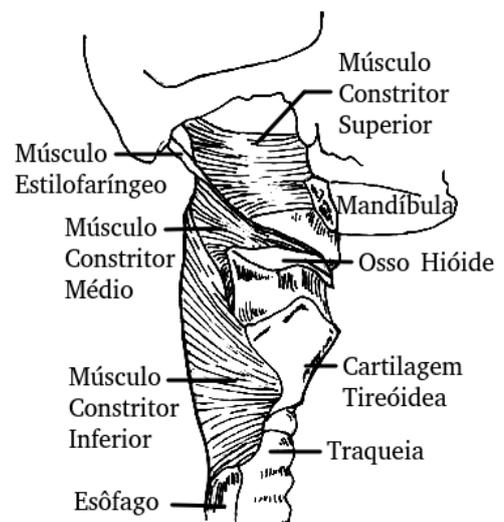
### 2.3.1 Faringe

Após sair da laringe, o ar vibrado pelas pregas vocais se desloca pela faringe, um tubo de 12 cm de comprimento e um diâmetro que varia ao longo de sua extensão (cerca de 2,5 cm em



(Domínio Público)

Figura 2.14 – Três partes da faringe.



(Adaptado de Raphael et al. [32])

Figura 2.15 – Músculos da Faringe.

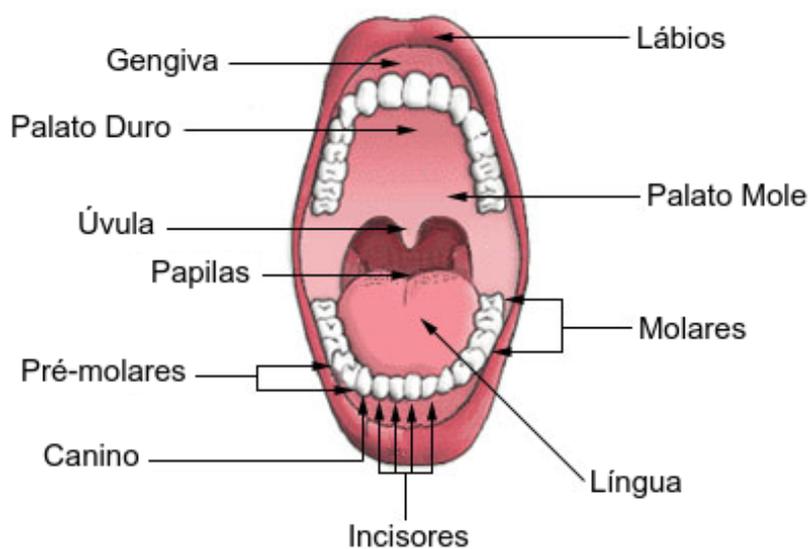
sua parte inferior e 4 cm em sua parte superior). A faringe é composta por músculos, tecidos conectivos e mucosa, sendo dividida em três partes: nasofaringe, orofaringe e laringofaringe. A nasofaringe se localiza atrás das cavidades nasais, a orofaringe atrás da cavidade oral e a laringofaringe atrás de parte da traqueia e da laringe, conectando-se ao esôfago, como ilustrado na Figura 2.14.

A articulação da faringe ocorre a partir da contração de seus músculos, gerando constrictões em sua extensão. Os principais músculos que compõem a faringe são o constritor inferior, o constritor médio, o músculo constritor superior e o estilofaríngeo, representados na Figura 2.15. A língua, os músculos faciais e os músculos laríngeos também exercem influência sobre as constrictões da faringe [7].

### 2.3.2 Cavidade Oral

A cavidade oral, por meio da qual a maior parte dos sons são emitidos, consiste no espaço limitado superiormente pelo palato (mole e duro), lateralmente pelas bochechas, inferiormente pela estrutura móvel da língua, anteriormente pela cavidade variável constringida pelos lábios e posteriormente pela cavidade faríngea. A cavidade oral possui diversos articuladores, ilustrados na Figura 2.16, que alteram sua forma ao longo do tempo durante a fala. Os principais de seus articuladores são os seguintes:

- **Língua:** A língua é o mais importante dos articuladores da cavidade oral, sendo participante da maior parte dos fonemas produzidos, formando constrictões com os dentes e com o palato duro. Sua estrutura possui uma musculatura rica, contendo músculos intrínsecos e extrínsecos que se contraem seletivamente de forma que a língua tenha



(Domínio Público)

Figura 2.16 – Cavidade oral e seus articuladores.

sustentação, mesmo sem conter nenhuma estrutura óssea ou cartilaginosa, e consiga se movimentar de maneira rápida e flexível.

- **Lábios:** Os dois lábios são compostos por músculo, membrana mucosa, tecido glandular, gordura e uma cobertura de epitélio. Os músculos faciais exercem uma influência significativa na articulação dos lábios, se inserindo nos lábios e influenciando em sua movimentação. A musculatura labial corrobora com a musculatura facial, permitindo grande flexibilidade e rapidez nos movimentos para os lábios. A articulação labial é importante para a produção das consoantes *p*, *b* e *m* e das vogais *u* e *o*.
- **Dentes:** Pouco atrás dos lábios, se encontram os dentes, estruturas separadas entre um agrupamento superior - maxilar - e um agrupamento inferior - mandíbula. Os dentes exercem uma função de articuladores imóveis com o quais a língua ou os lábios se conectam para formar diferentes sons, sendo os dentes incisores (superiores e inferiores) os principais participantes na articulação. Os dentes são importantes para a produção das consoantes *s* e *f* e das vogais *i* e *e*.
- **Palato Duro:** O palato duro é uma estrutura óssea forrada por epitélio localizada na parte superior da cavidade oral, logo atrás dos dentes. O palato duro participa na articulação como um ponto de contato ou de aproximação para a língua, sendo importante na produção de consoantes como *t*, *d*, *s*, *z*, *l*, *n* e *r*.
- **Palato Mole:** Também conhecido como véu palatino, o palato mole é composto exclusivamente de músculos e tecidos moles. A ação muscular do palato mole proporciona a abertura e o fechamento do orifício velofaríngeo, que é a via de conexão com o trato nasal. Assim o fechamento do véu palatino desacopla o trato nasal e a sua abertura acopla o trato nasal durante a fonação. A ação do véu palatino é essencial para a produção de sons nasais, como os das consoantes *m* e *n*. Entretanto, há também a participação do palato mole em constrições para formar as consoantes *g* e *k*. Problemas na atividade de abertura ou de fechamento do véu palatino podem causar, respectivamente, alterações de hiponasalidade ou hipernasalidade na fonação.

### 2.3.3 Cavidade Nasal

A cavidade nasal é formada por vários ossos do crânio e forrada por uma membrana mucosa. Essa estrutura interna é responsável por auxiliar no aquecimento, na filtragem de impurezas e na umidificação do ar inalado durante a inspiração. A cavidade nasal participa do processo de fonação quando acoplados ao trato vocal a partir da ação do véu palatino, que abre o orifício velofaríngeo permitindo a passagem do ar.

### 2.3.4 Ressonâncias do Trato Vocal

O som produzido na laringe percorre o trato vocal, sendo modificado de acordo com as frequências de ressonância do tubo. A estas frequências é designado o nome de formantes, sendo nomeada como primeira formante ( $F_1$ ) a de menor frequência, seguindo as demais formantes em ordem crescente ( $F_1 < F_2 < \dots < F_i < \dots$ ).

Assuma-se como exemplo um trato vocal com 17 cm de comprimento (comprimento médio de um trato vocal masculino) ajustado em posição neutra, i.e., as áreas de seção transversal do trato são iguais ao longo de todo o tubo. Neste caso, pode ser obtido que a primeira formante  $F_1 = 500$  Hz e que as demais formantes são múltiplos ímpares de  $F_1$  ( $F_2 = 1500$  Hz,  $F_3 = 2500$  Hz,  $F_4 = 3500$  Hz, ...)<sup>2</sup>.

O ajuste dos articuladores, ao alterar o formato do trato, altera também as formantes, deslocando suas frequências e modificando suas intensidades. Dessa maneira, os sons emitidos são caracterizados por suas formantes, sendo a primeira formante, normalmente, a mais intensa e as três primeiras formantes mais relevantes, tendo em vista que a maior parte da energia espectral encontra-se abaixo de 5000 Hz [7].

---

<sup>2</sup>Esses valores dependem da velocidade do som no meio [7, 39].

## CAPÍTULO

## 3

## Teoria Mioelástica-Aerodinâmica da Produção Vocal

Inicialmente pensada por Müller [40] e elaborada formalmente por van den Berg [5], a teoria mioelástica-aerodinâmica da fonação explica o processo ocorrido na laringe durante a produção vocal. Posteriormente, atualizada por Titze [6] e por Švec et al. [41], segue sendo a teoria mais aceita e difundida por apresentar uma melhor adequação teórico-experimental na maioria dos casos estudados, não somente para a fonação humana, mas também de mamíferos e de pássaros [42]. De acordo com essa teoria, a produção da voz ocorre a partir da interação entre as forças musculares (*mio*), a elasticidade do tecido (*elástica*), as pressões e os fluxos de ar (*aerodinâmica*) [7], de maneira que o ar proveniente dos pulmões, por meio da traqueia, atua sobre as pregas vocais, provocando nelas uma oscilação autossustentada<sup>1</sup>.

De acordo com a proposição original de van den Berg, podemos descrever o processo de produção vocal como segue. A fonação tem início com o fechamento das pregas vocais, condição necessária para que a fonação ocorra. Para isso, é necessária a ação dos músculos adutores (CAL e IA), não sendo obrigatória uma adução completa da glote, mas sendo essencial que a abertura glótica não seja superior a 3 mm [33]. Estando a glote aduzida, a abertura das pregas vocais ocorre devido ao aumento da pressão imediatamente abaixo da glote (pressão subglotal ou subglótica), como consequência do procedimento de respiração para a fonação descrito no Capítulo 2 (Subseção 2.1.1). Quando suficientemente elevada, essa pressão força a abertura das pregas vocais, permitindo que o ar escape para o sistema ressonador, onde o som é modificado. Uma vez abertas, o fluxo de saída de ar proporciona uma queda da pressão subglotal e as pregas vocais tendem a retornar no sentido medial, em consequência da ação de suas forças elásticas naturais e do princípio de Bernoulli, levando ao fechamento da glote. Após o fechamento das pregas, a pressão imediata abaixo da glote volta a subir e o processo se repete várias vezes por segundo. A frequência fundamental do som emitido é definida pelos ciclos de abertura e fechamento das pregas vocais, cuja frequência é dependente das massas e elasticidades efetivas dos tecidos das pregas vocais[5].

O princípio de Bernoulli enuncia que o aumento da velocidade de um fluido, em

---

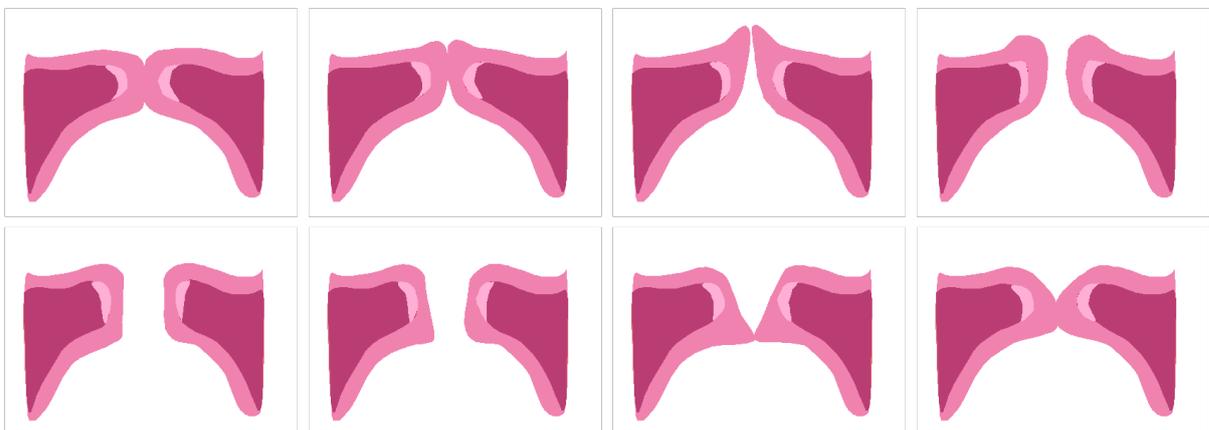
<sup>1</sup>Caso em que a frequência da oscilação não depende de um sinal periódico muscular, aerodinâmico ou de impulsos neurais (como argumentado em outras teorias anteriores) [5, 6].

condições adiabáticas, implica a redução de sua pressão, de acordo com a lei de Bernoulli,

$$P + \frac{1}{2}\rho v^2 + \rho z g = k, \quad (3.1)$$

em que  $P$  e  $v$  equivalem, respectivamente, à pressão e à velocidade do fluido,  $\rho$  é a densidade do fluido,  $g$  é a aceleração da gravidade,  $z$  é a altura do ponto em que o fluido é analisado acima de um plano de referência no eixo de direção da gravidade e  $k$  é uma constante. Esse efeito ocorre na fonação pois, ao passar entre as pregas vocais, o ar sofre um aumento de sua velocidade, tendo em vista que a glote representa um estreitamento para o ar advindo da traqueia. O aumento da velocidade do ar na glote proporciona uma redução na pressão entre as pregas, causando um efeito de sucção que as aproxima. Quanto mais estreita a fenda entre as pregas vocais, maior a redução da pressão e, conseqüentemente, maior a força de sucção no sentido medial. A ocorrência do Efeito Bernoulli no processo de produção vocal, averiguada em [43], é um dos elementos centrais propostos por van den Berg na teoria mioelástica-aerodinâmica e foi amplamente empregado na elaboração de modelos matemáticos, utilizados para estudar e simular o comportamento das pregas vocais durante a fonação [8]. Em geral, o efeito da gravidade é desprezível e ignorado nesses modelos, devido à pequena variação vertical ( $z$ ) que ocorre na glote.

Baseado em fotografias de movimento em alta velocidade e em eletromiografias, van den Berg fez observações a respeito de características da função glótica durante a fonação e argumentou em favor da relação entre os fenômenos observados e o controle mioelástico exercido pelo organismo nas pregas vocais diante da atuação aerodinâmica sobre elas [5]. Fatores como o comportamento da diferença de fase vertical, ilustrada na Figura 3.1, e da diferença de fase horizontal na abertura glótica são observado pelo autor e explicados com base na teoria. Essas diferenças de fase também são chamadas de onda da mucosa. A teoria mioelástica-aerodinâmica também explica o controle da frequência fundamental, destacando a influência aerodinâmica da pressão subglótica sofrida sobre a taxa vibratória das pregas,



(Domínio Público)

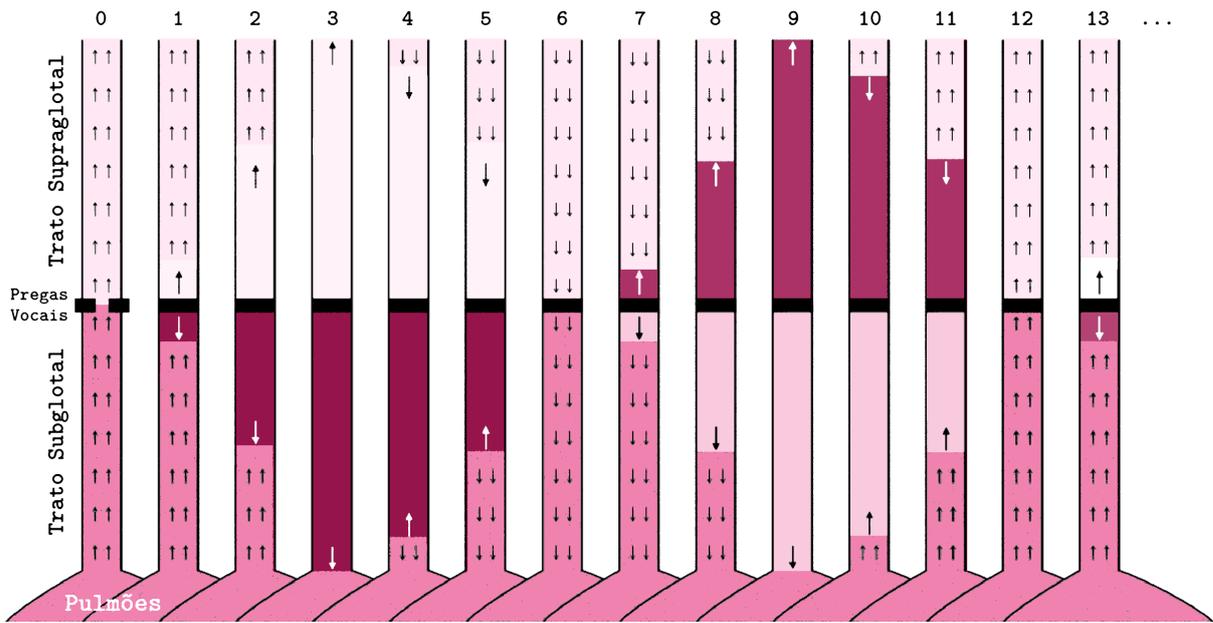
Figura 3.1 – Diferença de fase vertical na abertura e fechamento glótico.

bem como justifica o efeito de acoplamento entre a glote e o trato vocal, apontando o efeito de carga acústica que o trato vocal exerce, modificando a vibração das pregas vocais [5].

A teoria mioelástica-aerodinâmica foi, posteriormente, revisada por Titze que, com base em dados obtidos experimentalmente por meio de averiguação na produção vocal humana, em modelos matemático-computacionais propostos e em trabalhos recentes na literatura, observou alguns resultados divergentes dos previstos pela teoria proposta por van den Berg. Titze observou discrepâncias entre o esperado controle aerodinâmico da frequência fundamental e dados clínicos e, com o auxílio de experimentos simulados com modelos matemáticos, investigou suas causas. Titze pôde concluir que o controle da frequência fundamental da produção vocal é majoritariamente elástica e pontuou que o efeito aerodinâmico na frequência fundamental ainda não era bem explicado pela teoria mioelástica-aerodinâmica, identificando alguns possíveis fatores de erro, mas destacando também a importância dos elementos já consolidados da teoria [6]. É dado, ainda, um destaque à importância da onda da mucosa no processo de fonação e à necessidade de que a teoria fosse aprimorada para a inclusão de outros fatores ainda não formulados, como a perda de energia nos tecidos das pregas e os efeitos da inércia do ar, da turbulência do ar na glote e da impedância acústica do trato vocal.

Alguns experimentos posteriores foram realizados para a medição das pressões subglóticas e supraglóticas (pressão imediatamente acima das pregas vocais). Para isso, microssores de pressão recém-desenvolvidos foram utilizados, de forma que os sensores eram introduzidos através da glote por meio de um cateter, mantendo um dos sensores abaixo e outro acima da glote. Nesses experimentos, foram encontrados resultados divergentes dos esperados pela teoria [44, 45, 46, 47, 48, 49]. Diferentemente do que fora postulado por van den Berg, foram observados, durante o fechamento glotal, valores maiores de pressão subglótica do que no instante de abertura. Além disso, durante o período em que a glote permanecia fechada, ao invés de um crescimento da pressão, que culminaria num pico no momento de abertura da glote, observou-se uma oscilação amortecida nos valores medidos.

Com base nos valores medidos nos diversos estudos, Schutte e Miller apontam para a necessidade de uma reformulação da teoria, elencando elementos necessários para uma melhor explicação do processo fonatório [49], incluindo os fatores sugeridos por Titze [6]. É sugerido por eles, também, que o fechamento repentino da glote aliado ao efeito das ressonâncias da traqueia e do trato vocal ocasiona o aumento da pressão subglotal e influencia as oscilações nas pressões subglóticas e supraglóticas [46, 49]. Essa ideia foi corroborada ao longo dos anos e foi descrita com maior nível de detalhes por Chen e Miller [50], como ilustrado na Figura 3.2. Na passagem da fase 0 para a fase 1, ocorre o fechamento repentino da glote. Devido à inércia do ar, passa a ocorrer uma concentração de ar abaixo da glote, causando um pico na pressão subglotal, e ocorre uma diluição maior do ar acima da glote, causando um valor mínimo na pressão supraglotal. Das fases 2 a 6 o ar se propaga pelo trato vocal e pela

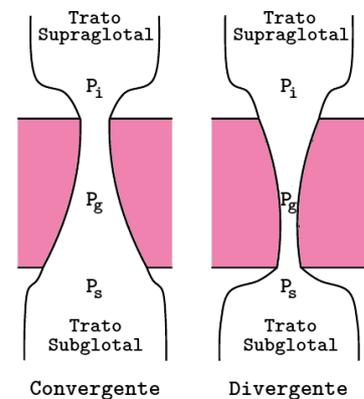


(Adaptado de Svec et al. [41])

Figura 3.2 – Ilustração das ressonâncias dos tratos subglotal e supraglotal causados pelo fechamento da glote. As cores mais escuras representam o ar mais comprimido e as mais claras o ar mais diluído. Por efeito de simplificação, os tratos são ilustrados com o mesmo comprimento.

traqueia em ida e volta, ressonando de acordo com as formantes subglotais [51] e do trato vocal [52]. Dessa maneira o ar segue se propagando pelos tubos, e se amortecendo, até que a glote se abra novamente, iniciando um novo ciclo glotal [50, 41].

Desde então, diversos trabalhos foram publicados, com experimentos de medidas clínicas e modelos matemáticos, visando compreender os fenômenos ainda não bem explicados pela teoria. Recentemente Švec et al. [41] realizam uma revisão da teoria mioelástica-aerodinâmica, tendo como base os avanços obtidos nesses trabalhos. Os pesquisadores destacam que, apesar da discrepância entre as medidas da pressão subglotal e os valores previstos, a teoria mioelástica-aerodinâmica não é invalidada, mas que as medidas permitem que a teoria seja observada diante de uma nova perspectiva. Eles, então, ressaltam a importância da inércia do ar [54, 55] e da diferença de fases vertical (onda da mucosa) [55, 56] como elementos essenciais ao processo de produção vocal.



(Adaptado de Story. [53])

Foi observado que a pressão subglotal não é o elemento principal que leva as pregas vocais a abrir e fechar a glote. Esse elemento principal é a pressão intraglotal, i.e.,

Figura 3.3 – Geometrias convergente e divergente da glote.

a pressão entre as pregas vocais. Essa pressão é influenciada pelas pressões subglotal, supraglotal e de Bernoulli, de uma maneira dependente da geometria da glote [55]. A essa geometria são atribuídos o nome convergente, para quando a porção inferior das pregas vocais encontra-se mais aberta que a porção superior, e divergente, quando a porção mais aberta é a porção superior (ver Figura 3.3). Dessa maneira, o requisito essencial para a oscilação autossustentada é a ocorrência da pressão intraglotal ser maior no momento de abertura glótica do que no momento de fechamento glótico. Para que isso ocorra, é necessário que a coluna de ar presente no trato vocal se mova com atraso temporal em relação à abertura das pregas vocais, devido à inércia do trato vocal [54, 53], e que a glote apresente geometria convergente na abertura glótica e geometria divergente (ou menos convergente) durante o fechamento glotal [56, 53].

Švec et al. resumem a teoria mioelástica-aerodinâmica da fonação, essencialmente, com os seguintes pontos [41]:

- Quando é iniciada a expiração para a fonação, a adução glótica provoca um aumento da pressão da subglotal, fazendo com que esta se torne maior que a pressão supraglotal;
- A pressão subglotal provoca a abertura das porções inferiores das pregas vocais. As porções superiores também se abrem, com um pequeno atraso temporal, de modo que a glote assuma a forma convergente;
- A abertura da glote ocorre devido à pressão intraglotal, ao passo que o fechamento da glote ocorre em decorrência da elasticidade das pregas vocais, atuando para reverter o movimento de abertura. Quanto mais tensionadas estiverem as pregas vocais, mais rápido será o fechamento da glote, proporcionando uma maior frequência de oscilação das pregas e, conseqüentemente, uma maior frequência fundamental da voz produzida;
- A diferença de fases vertical pode fazer com que a porção inferior da glote comece a se fechar primeiro. Dessa maneira a glote assume, no fechamento, uma forma divergente (ou uma forma menos convergente que a etapa de abertura);
- A forma divergente (ou menos convergente) proporciona uma queda na pressão intraglotal, fazendo com que as pregas vocais se aproximem (com um efeito de sucção) e possam se fechar. Entretanto, o fechamento não é necessário para a manutenção do movimento cíclico;
- Devido à elasticidade do tecido das pregas, elas voltam a se mover em sentido lateral. Dessa maneira, é iniciada novamente a abertura glótica e o ciclo se repete.

### 3.1 Modelos Matemático-Computacionais das Pregas Vocais

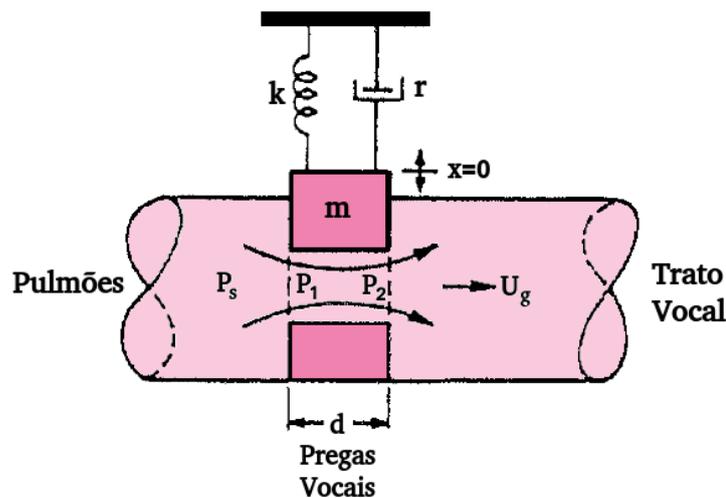
Em busca de avanços no estudo teórico da produção vocal e na síntese de voz, diversos modelos matemático-computacionais foram propostos com base na teoria mioelástica-aerodinâmica [16, 2, 8]. Para isso, esses modelos se pautaram na mecânica conhecida dos sistemas massa-mola amortecidos, para a representação da oscilação das pregas vocais, sendo também equacionada a dinâmica do ar e as pressões aéreas envolvidas no processo de fonação. Os principais modelos propostos na literatura são apresentados a seguir.

#### 3.1.1 Modelos de Uma Massa

Proposto por Flanagan e Landgraf [23], o primeiro modelo matemático-computacional baseado na teoria mioelástica-aerodinâmica consistia na representação de cada prega vocal utilizando uma única massa, acoplada a uma estrutura fixa e acrescentando um efeito de amortecimento, como ilustrado na Figura 3.4. Nesse modelo, a prega vocal é equacionada como um oscilador mecânico de acordo com:

$$m\ddot{x} + r\dot{x} + kx = F(t), \quad (3.2)$$

em que  $m$  é a massa da prega,  $k$  a constante elástica,  $r$  o fator de amortecimento,  $F(t)$  a força aplicada ao sistema e  $x(t)$  o deslocamento da massa da prega ao longo do tempo. O modelo é proposto considerando a simetria entre as pregas vocais, sendo necessário resolver



(Adaptado de Flanagan e Landgraf [23])

Figura 3.4 – Modelo de Uma Massa.

as equações do modelo apenas para uma prega. Entretanto, é possível adaptar o modelo para casos assimétricos.

As pressões localizadas imediatamente na entrada e na saída da glote,  $P_1$  e  $P_2$  respectivamente (ver Figura 3.4), são calculadas utilizando a pressão subglotal  $P_s$  e a pressão ocasionada pelo princípio de Bernoulli  $P_B$ , tomando como referência medições experimentais do efeito Bernoulli na produção vocal [43]. Assim,  $P_1$  e  $P_2$  são calculadas como segue:

$$\begin{cases} P_1 = (P_s - 1,37P_B), & (3.3) \\ P_2 = -0,5P_B. & (3.4) \end{cases}$$

A pressão de Bernoulli é calculada por:

$$P_B = \frac{1}{2}\rho \left( \frac{|U_g|}{A_g} \right)^2, \quad (3.5)$$

em que  $A_g$  equivale à área da glote, calculada como  $A_g = A_{g0} + lx$  (em que  $A_{g0}$  é a área inicial da glote, e  $l$  é o comprimento da glote), e  $U_g$  ao fluxo de ar na entre as pregas vocais, que é calculado a partir da interação com o trato vocal<sup>2</sup>. A expressão da Eq. (3.5) deriva da Eq. (3.1), assumindo o efeito da gravidade como desprezível<sup>3</sup>.

O cálculo da força aplicada ao sistema é feito utilizando a média das pressões de entrada e saída da glote, expresso da seguinte maneira:

$$F(t) = \frac{ld}{2}(P_1 + P_2), \quad (3.6)$$

em que  $d$  representa a espessura das pregas vocais.

Durante o fechamento da glote, ocorre a colisão entre as pregas vocais. Duas condições de contorno para esse evento são apresentadas como possíveis nesse modelo. A primeira condição implica assumir que a massa sofre uma perda instantânea de seu momento ao colidir. Assumir essa condição implica fechamentos glotais breves. A segunda condição consiste em assumir a interrupção do fluxo de ar na glote durante o fechamento, porém acrescentando uma viscosidade extra ao modelo, desacelerando as pregas aos poucos. Dessa maneira,  $r$  é substituído na Eq. (3.2) por  $(r + r')$ , em que  $r'$  representa o amortecimento viscoso ocasionado pela colisão. Em ambos os casos, assume-se  $F(t) = \frac{ld}{2}(P_s)$  durante o fechamento glotal.

Diferentes modelos de trato vocal podem ser acoplados ao sistema, sendo que no trabalho original em que esse modelo foi proposto, foi utilizado um circuito acústico para modelar o trato, como apresentado em Flanagan [27], a ser detalhado posteriormente na Seção 3.2. A área e o fluxo glotal foram utilizados para calcular a impedância glótica, a ser integrada no sistema.

<sup>2</sup>O cálculo de  $U_g$  como proposto por Flanagan e Landgraf [23] é apresentado na Seção 3.2.

<sup>3</sup>Para isso, a velocidade do fluido é calculada como  $\left( \frac{|U_g|}{A_g} \right)^2$

### 3.1.1.1 Variações do Modelo

Posteriormente, outros modelos baseados em uma única massa foram propostos [57, 55, 58, 59], buscando manter a simplicidade do equacionamento do sistema apresentado no modelo de Flanagan e Landgraf [23], porém acrescentando o efeito da diferença de fase vertical (ver Figura 3.1). O modelo proposto por Gupta et. al [57] utiliza uma massa com um formato fixo não retangular, mas um formato de paralelogramo com a base maior apontando para o sentido oposto da gravidade. Dessa maneira, considerando a superposição entre as massas, a porção superior se fecha antes da inferior e se abre após ela. Outro modelo de uma massa foi proposto por Titze [55] em que o formato da massa é variável. Para isso, o deslocamento no centro vertical da glote é utilizado no modelo e as áreas glotais nas regiões inferiores e superiores da glote são calculadas utilizando o deslocamento do centro, respectivamente, com um adiantamento e com um atraso, representando a onda da mucosa. O modelo desenvolvido por Liljencrants [58] propôs a representação das pregas vocais por uma massa rotacional, i.e., uma massa que se desloca na direção convencional, mas que também pode girar em torno de seu próprio eixo. Essa modificação adiciona um pouco mais de complexidade no modelo massa-mola, passando a utilizar duas equações diferenciais para modelar as pregas, porém permite representar com mais acurácia a oscilação glotal. Há também o modelo proposto por Drioli [59], que se assemelha ao de Titze [55]. Nesse modelo, é utilizada uma massa representando a porção inferior da glote e o movimento executado por ela é propagado com atraso temporal para o restante das pregas vocais, emulando, assim, a onda da mucosa.

### 3.1.2 Modelos de Duas Massas

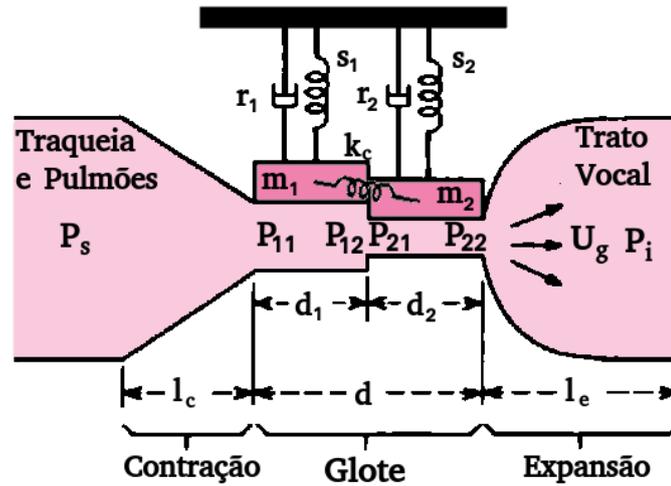
Estendendo a ideia do modelo a uma massa, um novo modelo foi proposto por Ishizaka e Flanagan [24], utilizando duas massas para representar as pregas vocais. Um dos principais objetivos foi corrigir algumas inadequações fisiológicas do modelo anterior, como a ausência da defasagem vertical, as regras de pressão para a interação entre glote e o trato vocal e o comportamento da glote durante o fechamento glotal. Esse modelo representa as pregas vocais utilizando duas massas,  $m_1$  e  $m_2$ , acopladas a molas e amortecedores, estando acopladas entre si por uma outra mola, como ilustrado na Figura 3.5. Nesse modelo é feito o uso também da estrutura simétrica para simplificar a solução das equações do modelo.

Esse modelo é equacionado pelo seguinte par de equações:

$$\begin{cases} m_1 \frac{d^2 x_1}{dt^2} + r_1 \frac{dx_1}{dt} + s_1(x_1) + k_c(x_1 - x_2) = F_1, \\ m_2 \frac{d^2 x_2}{dt^2} + r_2 \frac{dx_2}{dt} + s_2(x_2) + k_c(x_2 - x_1) = F_2. \end{cases} \quad (3.7)$$

$$\begin{cases} m_2 \frac{d^2 x_2}{dt^2} + r_2 \frac{dx_2}{dt} + s_2(x_2) + k_c(x_2 - x_1) = F_2. \end{cases} \quad (3.8)$$

em que  $k_c$  representa a mola de acoplamento entre as massas e  $x_i(t)$ ,  $r_i$ ,  $s_i(x_i)$  e  $F_i(t)$  equivalem, respectivamente, ao deslocamento, ao amortecimento, a uma função não linear para a



(Adaptado de Ishizaka e Flanagan [24])

Figura 3.5 – Modelo de Duas Massas.

força elástica e à força aplicada à  $i$ -ésima massa ( $m_i$ , em que  $i = 1$  representa a massa inferior e  $i = 2$  a massa superior).

A elasticidade das massas é representada como uma função de  $x_i$  pelo fato de que, nesse modelo, são utilizadas molas não-lineares, buscando uma melhor representação do movimento realizado pelas pregas vocais. É adicionada também uma força elástica não linear ao instante da colisão, para acrescentar mais realismo ao fechamento da glote no modelo. Dessa maneira, a força elástica na massa  $m_i$  ( $i = 1, 2$ ) é calculada da seguinte maneira:

$$s_i(x_i) = \begin{cases} k_i(x_i + \eta_{k_i}x_i^3), & \text{para } x_i > -\frac{A_{g0_i}}{2l_g} \\ k_i(x_i + \eta_{k_i}x_i^3) + h_i \left( \left( x_i + \frac{A_{g0_i}}{2l_g} \right) + \eta_{h_i} \left( x_i + \frac{A_{g0_i}}{2l_g} \right)^3 \right), & \text{para } x_i \leq -\frac{A_{g0_i}}{2l_g} \end{cases}, \quad (3.9)$$

em que  $k_i$  e  $\eta_{k_i}$  são, respectivamente, a rigidez linear e o coeficiente de não-linearidade da mola,  $h_i$  e  $\eta_{h_i}$  são, respectivamente, a rigidez linear e o coeficiente de não-linearidade da colisão,  $A_{g0_i}$  é a área glotal inicial e  $l_g$  é o comprimento da glote. O amortecimento das massas ( $r_i$ ) durante a colisão é abordado da mesma maneira que a segunda abordagem proposta no modelo de Flanagan e Landgraf [23], acrescentando um amortecimento viscoso ao valor inicial, sem a interrupção imediata do momento das pregas.

Como ilustrado na Figura 3.5,  $P_s$ ,  $P_{11}$ ,  $P_{12}$ ,  $P_{21}$ ,  $P_{22}$  e  $P_i$  são, respectivamente, a pressão subglotal, a pressão na entrada da área da primeira massa, a pressão na saída da área da primeira massa, a pressão na entrada da área da segunda massa, a pressão na saída da área da segunda massa e a pressão na primeira seção do trato vocal. Considerando o efeito Bernoulli de acordo com a mesma lógica utilizada no modelo de uma massa (ver Eq. (3.5)), a

distribuição de pressões ao longo da glote é descrita da seguinte maneira:

$$\left\{ \begin{array}{l} P_s - P_{11} = 1.37 \frac{\rho}{2} \left( \frac{U_g}{A_{g1}} \right)^2 + \int_0^{l_c} \frac{\rho}{A_c(y)} dy \frac{dU_g}{dt}, \end{array} \right. \quad (3.10)$$

$$\left\{ \begin{array}{l} P_{11} - P_{12} = 12 \frac{\mu l_g^2 d_1}{A_{g1}^3} U_g + \frac{\rho d_1}{A_{g1}} \frac{dU_g}{dt}, \end{array} \right. \quad (3.11)$$

$$\left\{ \begin{array}{l} P_{12} - P_{21} = \frac{\rho}{2} U_g^2 \left( \frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2} \right), \end{array} \right. \quad (3.12)$$

$$\left\{ \begin{array}{l} P_{21} - P_{22} = 12 \frac{\mu l_g^2 d_2}{A_{g2}^3} U_g + \frac{\rho d_2}{A_{g2}} \frac{dU_g}{dt}, \end{array} \right. \quad (3.13)$$

$$\left\{ \begin{array}{l} P_{22} - P_i = -\rho \left( \frac{U_g}{A_{g2}} \right)^2 \frac{A_{g2}}{A_i} \left( 1 - \frac{A_{g2}}{A_1} \right), \end{array} \right. \quad (3.14)$$

em que  $\mu$  é o coeficiente de viscosidade do ar,  $A_{g1}$  e  $A_{g2}$  são as áreas das massas  $m_1$  e  $m_2$ , respectivamente,  $A_c(y)$  é a área da região de transição da traqueia para a glote (contração na Figura 3.5),  $A_i$  é a área da primeira seção do trato vocal e  $d_1$  e  $d_2$  são as espessuras das massas  $m_1$  e  $m_2$ .

Para calcular a pressão em cada massa, é considerada a média entre as pressões em sua entrada e sua saída. Logo, as pressões nas massas são calculadas como:

$$\left\{ \begin{array}{l} P_{m1} = \frac{1}{2} (P_{11} + P_{12}), \end{array} \right. \quad (3.15)$$

$$\left\{ \begin{array}{l} P_{m2} = \frac{1}{2} (P_{21} + P_{22}), \end{array} \right. \quad (3.16)$$

em que  $P_{m1}$  e  $P_{m2}$  são as pressões em  $m_1$  e  $m_2$ . A força incidente para cada massa é calculada pela expressão  $F_i = P_{m_i} l_g d_i$ .

Com base na distribuição das pressões na glote, é utilizada uma analogia com circuitos elétricos para calcular as pressões  $P_{11}$ ,  $P_{12}$ ,  $P_{21}$  e  $P_{22}$ . Fazendo as seguintes equivalências:

$$R_c = 1.37 \frac{\rho}{2} \left( \frac{|U_g|}{A_{g1}^2} \right), \quad L_c = \int_0^{l_c} \frac{\rho}{A_c(y)} dy,$$

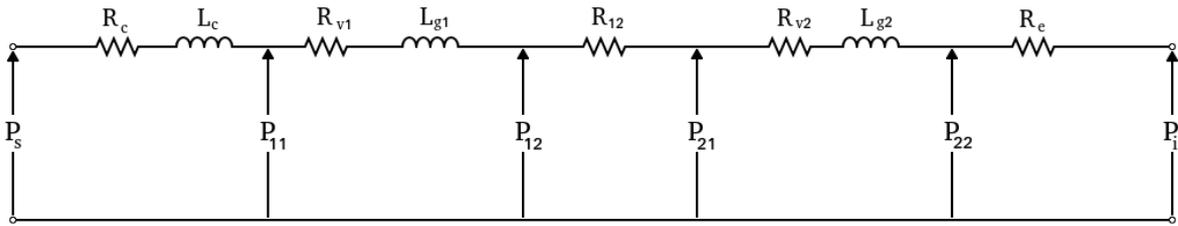
$$R_{v1} = 12 \frac{\mu l_g^2 d_1}{A_{g1}^3}, \quad L_{g1} = \frac{\rho d_1}{A_{g1}},$$

$$R_{12} = \frac{\rho}{2} \left( \frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2} \right) |U_g|,$$

$$R_{v2} = 12 \frac{\mu l_g^2 d_2}{A_{g2}^3}, \quad L_{g2} = \frac{\rho d_2}{A_{g2}},$$

$$R_e = -\rho \left( \frac{1}{A_1 A_{g2}} \right) \left( 1 - \frac{A_{g2}}{A_1} \right) |U_g|,$$

e assumindo a equivalência de representar pressão como tensão e fluxo como corrente, a glote pode ser aproximada pelo circuito da Figura 3.6.



(Baseado em Ishizaka e Flanagan [24])

Figura 3.6 – Circuito Equivalente da Glote

Utilizando as relações de malha do circuito, assumindo que  $L_c$  tem efeito desprezível e substituindo  $P_{11}$ ,  $P_{12}$ ,  $P_{21}$  e  $P_{22}$  nas Eqs. (3.15) e (3.16), podem ser obtidas as pressões incidentes em cada massa do sistema, como segue:

$$\begin{cases} P_{m_1} = P_s - 1.37 \frac{\rho}{2} \left( \frac{U_g}{A_{g1}} \right)^2 - \frac{1}{2} \left( R_{v1} U_g + L_{g1} \frac{dU_g}{dt} \right), & (3.17) \\ P_{m_2} = P_{m1} - \frac{1}{2} \left( (R_{v1} + R_{v2}) U_g + (L_{g1} + L_{g2}) \frac{dU_g}{dt} \right) - \frac{\rho}{2} U_g^2 \left( \frac{1}{A_{g2}^2} - \frac{1}{A_{g1}^2} \right). & (3.18) \end{cases}$$

Assim como no modelo de uma massa, o trabalho original desse modelo utilizou também um circuito acústico para modelar o trato vocal [27]. Entretanto, de maneira similar, outros modelos de trato vocal podem ser integrados ao sistema, sendo feitas as devidas adequações.

### 3.1.2.1 Variações do Modelo

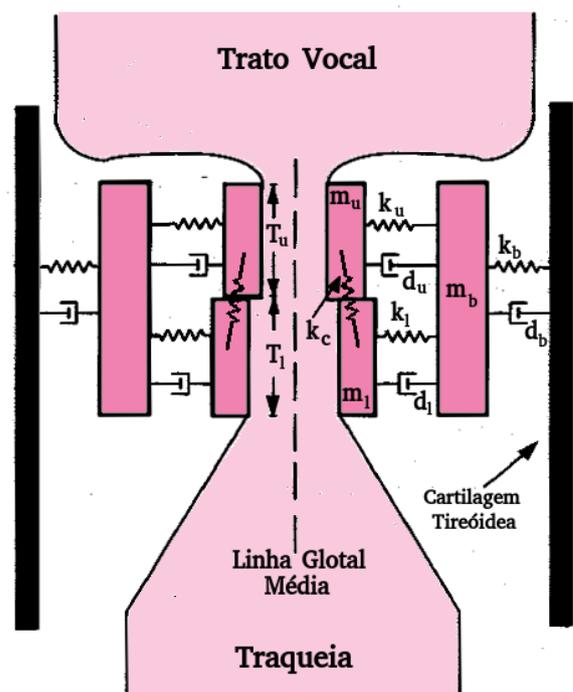
Diversas variações do modelo de duas massas foram propostos posteriormente [34, 60, 61, 62, 63, 64, 65, 17, 66, 67], buscando aperfeiçoar a adequação fisiológica obtida pelo modelo para a síntese de diversos tipos de fonação. Três variações do modelo foram propostas por Koizumi et al. [60], implementando diferentes ajustes da relação de molas e amortecedores, desacoplando a mola superior ( $m_2$ ) de uma superfície rígida (deixando-a acoplada somente a  $m_1$ ) e permitindo a movimentação vertical de  $m_2$ . Pereira [34, 62] e Miller et al. [61] apresentaram um modelo baseado no modelo de duas massas da prega vocal, acrescentando a presença de uma massa superior para representar as pregas ventriculares, suavizando a área glotal entre as duas massas e refinando as equações do balanço de pressões, que resultou em melhores resultados na síntese de vozes sussurradas. O modelo proposto por Pelorson et al. [63] apresentou uma suavização da geometria das pregas, similar à apresentada posteriormente no modelo de Lous et al. [65], e um refinamento mais preciso no equacionamento das pressões, baseando se nas equações de Navier-Stokes e incluindo efeitos de escoamento turbulento no modelo. Steinecke e Herzog [64] propuseram uma simplificação aos equacionamentos

de pressão, visando facilitar algumas análises matemáticas do modelo. Essa simplificação veio a ser utilizada por Zhang e Jiang [17] em um modelo proposto para poder representar a presença de um póliplo unilateral nas pregas vocais. Para isso foi adicionada uma massa extra a uma das pregas vocais, gerando necessariamente uma vibração assimétrica. Recentemente foi proposta uma variação desse modelo, que busca incluir a diferença de fase horizontal ao sistema, por Birkholz et al. [66, 67], sugerindo um formato fixo na direção pósterio-anterior para as massas.

### 3.1.3 Modelos *Body-Cover*

Tomando como base o modelo corpo-cobertura das pregas vocais [35], um novo modelo foi proposto por Story e Titze [25], estendendo a ideia do modelo de duas massas. Os modelos *Body-Cover* (ou Corpo-Cobertura) são formulados com duas camadas de massas, uma representando a cobertura de tecido mucoso das pregas vocais e outra representando o corpo muscular das pregas vocais.

O modelo de Story e Titze [25] busca aperfeiçoar o sistema de síntese de voz, tornando-o mais realista que seus antecessores, porém mantendo o uso de uma estrutura simples e de baixo custo computacional. Também visa estabelecer algumas correlações do modelo com condições fisiológicas de diferentes modos de fonação. Para a simplificação do modelo em apenas duas camadas, a camada intermediária é desconsiderada, sendo o ligamento vocal<sup>4</sup> dividido entre as camadas do corpo e da cobertura. Para a adequação das massas, considera-se que a cobertura consiste no epitélio e nas camadas superficial e intermediária da lâmina própria e que o corpo consiste na camada profunda da lâmina própria e no músculo tireoaritenoideo (TA).



(Adaptado de Story e Titze [25])

Figura 3.7 – Modelo Corpo-Cobertura.

Esse modelo utiliza três massas, sendo duas delas para representar a cobertura e uma para o corpo. Essas massas são acopladas por molas e amortecedores com uma estrutura fixa que representa a cartilagem tireóidea e entre si, como ilustrado na Figura 3.7.

Para simplificar a formulação matemática do problema, assume-se a existência de

<sup>4</sup>Composto pelas camadas intermediária e profunda da lâmina própria. Ver 2.2.2.1 no Capítulo 2.

uma simetria entre as duas pregas vocais, sendo possível, entretanto, a extensão do modelo para casos assimétricos. As equações de movimento para as três massas ( $m_u$ ,  $m_l$  e  $m_b$ ) são dadas em termos de balanços de força, da seguinte maneira:

$$\begin{cases} m_u \ddot{x}_u = F_{ku} + F_{du} - F_{kc} + F_{eu} + F_{uCol}, & (3.19) \\ m_l \ddot{x}_l = F_{kl} + F_{dl} + F_{kc} + F_{el} + F_{lCol}, & (3.20) \\ m_b \ddot{x}_b = F_{kb} + F_{db} - F_{ku} - F_{du} - F_{kl} - F_{dl}. & (3.21) \end{cases}$$

em que:

- $x_u$ ,  $x_l$  e  $x_b$  são os deslocamentos de cada massa;
- $F_{ku}$ ,  $F_{ku}$  e  $F_{ku}$  são as forças exercidas pelas molas, dadas por:

$$F_{ku} = -k_u \left( (x_u - x_{u0}) - (x_b - x_{b0}) + \eta_u ((x_u - x_{u0}) - (x_b - x_{b0}))^3 \right), \quad (3.22)$$

$$F_{kl} = -k_l \left( (x_l - x_{l0}) - (x_b - x_{b0}) + \eta_l ((x_l - x_{l0}) - (x_b - x_{b0}))^3 \right), \quad (3.23)$$

$$F_{kb} = -k_b \left( (x_b - x_{b0}) + \eta_b (x_b - x_{b0})^3 \right); \quad (3.24)$$

- $x_{u0}$ ,  $x_{l0}$  e  $x_{b0}$  são as posições iniciais de cada massa, sendo relacionadas com o nível de adução glótica;
- $k_u$ ,  $k_l$  e  $k_b$  são as constantes elásticas das molas;
- $\eta_u$ ,  $\eta_l$  e  $\eta_b$  são os coeficientes de não-linearidade das molas;
- $F_{kc}$  é a força da mola (linear) de acoplamento entre  $m_u$  e  $m_l$ , dada por:

$$F_{kc} = -k_c ((x_l - x_{l0}) - (x_u - x_{u0})); \quad (3.25)$$

- $k_c$  é a rigidez da mola de acoplamento;
- $F_{du}$ ,  $F_{du}$  e  $F_{du}$  são as forças devidas ao amortecimento, dadas por:

$$F_{du} = -d_u (\dot{x}_u - \dot{x}_b), \quad (3.26)$$

$$F_{dl} = -d_l (\dot{x}_l - \dot{x}_b), \quad (3.27)$$

$$F_{db} = -d_b \dot{x}_b; \quad (3.28)$$

- $d_u$ ,  $d_l$  e  $d_b$  são os coeficientes de amortecimento para as massas, que são calculados como segue:

$$d_u = 2\zeta_u \sqrt{m_u k_u}, \quad (3.29)$$

$$d_l = 2\zeta_l \sqrt{m_l k_l}, \quad (3.30)$$

$$d_b = 2\zeta_b \sqrt{m_b k_b}; \quad (3.31)$$

- $\zeta_u, \zeta_l$  e  $\zeta_b$  são as taxas de amortecimento das massas;
- $F_{uCol}$  e  $F_{lCol}$  são as forças elásticas agindo nas massas durante a colisão, dadas por:

$$F_{uCol} = \begin{cases} -h_{uCol}((x_u - x_{uCol}) + \eta_{uCol}(x_u - x_{uCol})^3), & \text{se } x_u \leq x_{uCol}, \\ 0, & \text{se } x_u > x_{uCol}, \end{cases} \quad (3.32)$$

$$F_{lCol} = \begin{cases} -h_{lCol}((x_l - x_{lCol}) + \eta_{lCol}(x_l - x_{lCol})^3), & \text{se } x_l \leq x_{lCol}, \\ 0, & \text{se } x_l > x_{lCol}; \end{cases} \quad (3.33)$$

- $x_{uCol}$  e  $x_{lCol}$  são os valores dos deslocamentos em que a colisão ocorre. No caso simétrico, esses valores serão equivalentes ao deslocamento em que se atinge a linha glotal media (é conveniente definir  $x_{uCol} = 0$  e  $x_{lCol} = 0$ ).
- $h_{uCol}$  e  $h_{lCol}$  são as constantes elásticas da colisão;
- $\eta_{uCol}$  e  $\eta_{lCol}$  são os coeficientes elásticos não-linearidade da colisão;
- Durante a colisão  $\zeta_u$  e  $\zeta_l$  tem seu valor incrementado, de forma que seus valores são atualizados da seguinte maneira:

$$\zeta_u = \begin{cases} \zeta_u + 0.4, & \text{se } x_u \leq x_{uCol}, \\ \zeta_u, & \text{se } x_u > x_{uCol}, \end{cases} \quad (3.34)$$

$$\zeta_l = \begin{cases} \zeta_l + 0.4, & \text{se } x_l \leq x_{lCol}, \\ \zeta_l, & \text{se } x_l > x_{lCol}, \end{cases} \quad (3.35)$$

- $F_{eu}$  e  $F_{el}$  são as forças externas geradas pelo fluxo de ar na glote e que podem ser calculadas com o equacionamento da pressão durante a oscilação.

A área glotal é calculada para as regiões da massa superior e da massa inferior pelas seguintes expressões:

$$a_u = \begin{cases} 0, & \text{se } x_u \leq x_{uCol}, \\ 2x_u L_g, & \text{se } x_u > x_{uCol}, \end{cases} \quad (3.36)$$

$$a_l = \begin{cases} 0, & \text{se } x_l \leq x_{lCol}, \\ 2x_l L_g, & \text{se } x_l > x_{lCol}, \end{cases} \quad (3.37)$$

em que  $L_g$  é o comprimento da glote.

O equacionamento das pressões utilizado nesse modelo assume como verdadeiras as seguintes premissas [25]:

1. No local em que ocorre o diâmetro glotal mínimo o escoamento de ar sofre uma separação, i.e., até a região com a menor área glotal ocorre um tipo de escoamento do ar e a partir dessa região outro tipo;
2. O escoamento de ar de acordo com a equação de Bernoulli ocorre da região subglotal até a região com a menor área glotal;
3. A separação do escoamento, partindo da região de menor área glotal, proporciona um jato de ar de diâmetro constante. Nessa região a pressão é considerada constante.
4. A recuperação de pressão na saída da glote é dada de acordo com o equacionamento de Ishizaka e Matsudaira [68].

Partindo desses pressupostos, são obtidas as seguintes equações para calcular a pressão no fluxo de Bernoulli (antes da separação do escoamento) e no fluxo de jato de ar (após a separação do escoamento):

$$P = \begin{cases} P_s - (P_s - P_i) \frac{a^{-2} - a_s^{-2}}{a_m^{-2}(1 - k_e) - a_s^{-2}}, & \text{fluxo de Bernoulli,} \\ P_i - (P_s - P_i) \frac{k_e a_m^{-2}}{a_m^{-2}(1 - k_e) - a_s^{-2}}, & \text{fluxo de jato de ar,} \end{cases} \quad (3.38)$$

$$P = \begin{cases} P_i - (P_s - P_i) \frac{k_e a_m^{-2}}{a_m^{-2}(1 - k_e) - a_s^{-2}}, & \text{fluxo de jato de ar,} \end{cases} \quad (3.39)$$

em que  $P_s$  é a pressão subglotal,  $P_i$  a pressão supraglotal,  $a$  é a área entre as massas da região em que se está sendo calculada,  $a_s$  é a área da seção transversal da região subglotal<sup>5</sup>,  $a_m$  é a área mínima da glote ( $\min(a_u, a_l)$ ) e  $k_e$  é o coeficiente de recuperação de pressão [68], que é calculado utilizando a área da seção transversal da região supraglotal<sup>6</sup>  $a_i$  como segue:

$$k_e = \frac{2a_m}{a_i} \left( 1 - \frac{a_m}{a_l} \right). \quad (3.40)$$

Como  $k_e$  e  $a_s^{-2}$  são valores pequenos comparados às demais grandezas, seu efeito pode ser desprezado, reduzindo as Eqs. (3.38) e (3.39) a:

$$P = \begin{cases} P = P_s - (P_s - P_i) \left( \frac{a_m}{a} \right)^2, & \text{fluxo de Bernoulli,} \\ P = P_i, & \text{fluxo de jato de ar.} \end{cases} \quad (3.41)$$

$$P = P_i, \quad \text{fluxo de jato de ar.} \quad (3.42)$$

As pressões na massa superior e na massa inferior ( $P_u$  e  $P_l$ ) podem ser calculadas para diferentes configurações das pregas. Como o modelo corpo-cobertura utiliza 2 massas, existem 5 configurações possíveis, ilustradas na Figura 3.8, que resultam em diferentes pressões em cada massa, de acordo com as Eqs. (3.41) e (3.42).

<sup>5</sup>Em geral, considera-se como a área da seção transversal do final da traqueia.

<sup>6</sup>Em geral, considera-se como a área da seção transversal do início do trato vocal.

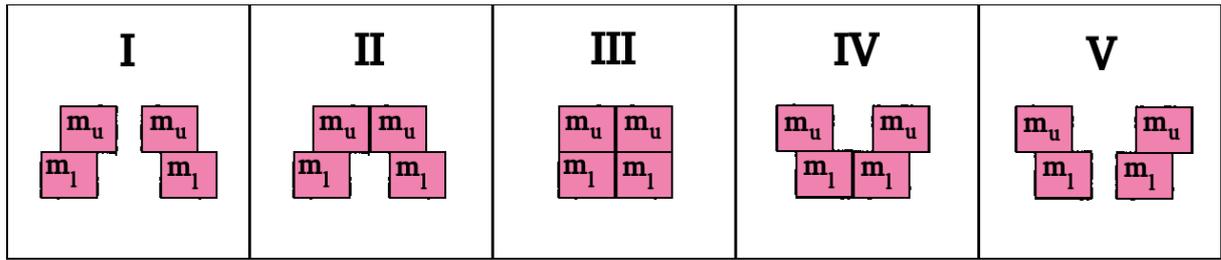


Figura 3.8 – Cinco possíveis configurações glotais do modelo corpo-cobertura.

Na configuração **I** a área mínima se encontra na região da massa superior ( $a_m = a_u$ ), assim até a massa inferior o fluxo é de Bernoulli e, acima dela, o fluxo se separa formando um jato de ar. Logo, nesse caso  $P_l = P_s - (P_s - P_i) \left(\frac{a_u}{a_l}\right)^2$  e  $P_u = P_i$ .

Na configuração **II** ocorre o fechamento da parte superior da glote, o que impede a existência de fluxo. Dessa maneira, haverá somente a pressão subglótica incidindo na massa inferior. Logo,  $P_l = P_s$  e  $P_u = 0$ .

Na configuração **III** toda a glote se encontra fechada, o que impede o fluxo e resulta na ausência de pressão agindo entre as massas. Nesse caso,  $P_l = 0$  e  $P_s = 0$ .

Na configuração **IV** ocorre o fechamento da parte inferior da glote, o que impede a existência de fluxo. Dessa maneira, haverá somente a pressão supraglótica incidindo na massa superior. Logo,  $P_l = 0$  e  $P_u = P_i$ .

Na configuração **V**, a área mínima se encontra na região da massa inferior ( $a_m = a_l$ ), de modo que o fluxo de Bernoulli se encerra na entrada da glote e a partir da glote o fluxo se separa formando um jato de ar. Assim, nesse caso  $P_l = P_i$  e  $P_s = P_i$ .

Dessa maneira, o cálculo das forças aerodinâmicas sobre as massas é realizado pelas seguintes equações

$$F_{eu} = P_u L_g T_u \quad (3.43)$$

$$F_{el} = P_l L_g T_l \quad (3.44)$$

em que  $T_u$  e  $T_l$  são as espessuras verticais das partes superior e inferior, respectivamente.

O fluxo glotal é calculado com base nas equações formuladas por Titze [69], da seguinte maneira:

$$u = \left(\frac{a_m c}{k_t}\right) \left(\frac{-a_m}{A^*} \pm \sqrt{\left(\frac{a_m}{A^*}\right)^2 \pm \left(\frac{4k_t}{c^2 \rho}\right) (P_s^+ - P_i^-)}\right), \quad (3.45)$$

em que  $k_t = 1 - k_e$  [70, 71] é o coeficiente de pressão transglotal,  $\rho$  é a densidade do ar,  $c$  é a velocidade do som no ar,  $P_s^+$  e  $P_i^-$  são as pressões incidentes acima e abaixo da glote, respectivamente, que dependem do modelo de trato vocal e de tubo subglotal utilizado.  $A^*$  é denominada área efetiva do trato vocal para a carga acústica da glote, sendo calculada, utilizando área da primeira seção do trato vocal ( $A_s$ ) e a área da última seção do tubo subglótico

( $A_i$ ), como segue:

$$\frac{1}{A^*} = \frac{1}{A_s} + \frac{1}{A_i}. \quad (3.46)$$

Diferenciando-se do que foi proposto nos modelos a uma massa [23] e a duas massas [24], no trabalho em que esse modelo foi proposto não é utilizado um circuito acústico para representar o trato vocal, mas um modelo de linha de transmissão analógica por reflexão de ondas do trato vocal [26, 72], sendo acoplada também a traqueia como um trato subglotal. Apenas para simular o efeito de radiação, que ocorre na saída da voz pelos lábios, foi utilizada a analogia de circuitos, modelando como uma impedância indutiva e resistiva [27].

Os autores também propuseram parâmetros ao modelo para representar os 4 ajustes das pregas vocais analisados por Hirano<sup>7</sup> [35], ilustrados na Figura 2.10. Esses parâmetros propostos por Story e Titze [25] para simular cada caso encontram-se ilustrados na Tabela 3.1.

Tabela 3.1 – Valores para as 4 configurações propostas por Hirano.

(Adaptado de Story e Titze [25])

Parâmetros	Caso (a)	Caso (b)	Caso (c)	Caso (d)
$m_u$ (g)	0,01	0,01	0,01	0,01
$m_l$ (g)	0,01	0,01	0,01	0,01
$m_b$ (g)	0,05	0,05	0,05	0,015
$T_u$ (cm)	0,15	0,15	0,15	0,15
$T_l$ (cm)	0,15	0,15	0,15	0,15
$k_u$ (N/m)	3,5	3,5	3,5	79,0
$k_l$ (N/m)	5,0	5,0	5,0	80,0
$k_b$ (N/m)	20,0	700,0	100,0	200,0
$k_c$ (N/m)	0,5	2,0	2,0	2,0
$x_{u0}$ (cm)	0,0179	0,0179	0,0179	0,0179
$x_{l0}$ (cm)	0,018	0,018	0,018	0,018
$x_{b0}$ (cm)	0,30	0,30	0,30	0,30
$P_l$ (kPa)	0,8	0,8	0,8	0,8

### 3.1.3.1 Equacionamento Atualizado e Regras Musculares

Posteriormente, foram propostas atualizações a esse modelo por Titze [71] e por Titze e Story [28], propondo três alterações relevantes: (1) substituir as duas massas da cobertura por uma massa rotacional, como proposta por Lijencrants [58], para permitir uma abertura e um fechamento mais suaves da glote; (2) refinar o equacionamento da pressão, adequando-o ao novo modelo da cobertura e acrescentando o critério de separação do fluxo proposto por Pelorson et al. [63]; (3) criar regras para controlar os parâmetros do modelo de acordo com a atividade de músculos intrínsecos da laringe.

<sup>7</sup>Como descrito em 2.2.2.1, no Capítulo 2.

Ao substituir a representação da cobertura *Body-Cover* por uma massa rotacional, obtém-se uma nova dinâmica para o modelo, porém mantendo o mesmo número de graus de liberdade para o movimento das massas. Para o deslocamento da cobertura, em acordo com a ilustração da Figura 3.9, assumimos uma rotação  $\theta$ , em torno de um ponto nodal  $z_n$ , e uma translação  $\xi$ , resultando nas seguintes equações [71]

$$\begin{cases} I_c \ddot{\theta} + B_c \dot{\theta} + \kappa \theta = T_a, \\ m \ddot{\xi} + b (\dot{\xi} - \dot{\xi}_b) + k (\xi - \xi_b) = F_a, \end{cases} \quad (3.47)$$

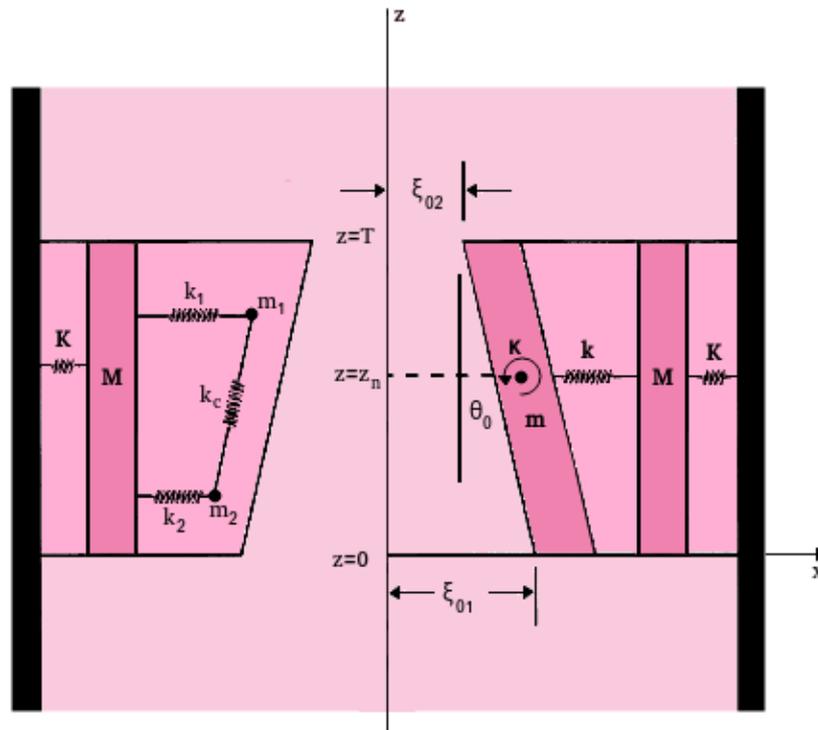
$$(3.48)$$

em que  $I_c$  equivale ao momento de inercia da cobertura,  $B_c$  ao amortecimento rotacional,  $\kappa$  à elasticidade rotacional,  $T_a$  ao torque aerodinâmico,  $m$  à massa da cobertura,  $b$  ao amortecimento translacional,  $k$  à elasticidade translacional,  $F_a$  a força aerodinâmica no ponto nodal e  $\xi_b$  ao deslocamento do corpo. Mantendo uma lógica similar ao modelo original, o movimento do corpo pode ser equacionado como segue:

$$M \ddot{\xi}_b + b (\dot{\xi}_b - \dot{\xi}) + k (\xi_b - \xi) + K \xi_b + B \dot{\xi}_b = 0, \quad (3.49)$$

em que  $M$ ,  $B$  e  $K$  representam, respectivamente, a massa, o amortecimento e a elasticidade do corpo.

Assim como ilustrado na Figura 3.9, é possível fazer uma equivalência utilizando duas massas na representação da cobertura, sem perder a característica contínua da glote (que



(Adaptado de Titze e Story [28])

Figura 3.9 – Modelo Corpo-Cobertura com Massa Rotacional de Liljencrants.

permite um equacionamento mais suave das pressões entre as pregas vocais). Para isso, os parâmetros para cada massa podem ser calculadas de maneira a depender do ponto nodal  $z_n$  e da espessura da glote  $T$ , como segue [28]:

$$m_1 = m \frac{z_n}{T}, \quad (3.50)$$

$$m_2 = m \left(1 - \frac{z_n}{T}\right), \quad (3.51)$$

$$k_1 = k \frac{z_n}{T}, \quad (3.52)$$

$$k_2 = k \left(1 - \frac{z_n}{T}\right), \quad (3.53)$$

$$b_1 = b \frac{z_n}{T}, \quad (3.54)$$

$$b_2 = b \left(1 - \frac{z_n}{T}\right). \quad (3.55)$$

e a mola de acoplamento pode ser obtida pela seguinte expressão:

$$k_c = \frac{1}{4} \left( \frac{m}{I_c} \kappa - k \right) \quad (3.56)$$

O equacionamento para o balanço de forças e torque, para as pressões envolvidas, fluxo de ar na glote e a área glotal encontra se detalhado no Apêndice A.

Para definir e controlar os parâmetros do modelo, Titze e Story desenvolveram regras musculares, baseando-se nas atividade dos músculos intrínsecos TA, CT e LCA. Os efeitos da contração/relaxamento desses músculos são quantificados para que seja possível calcular, a partir de seus níveis de ativação, os valores das massas, elasticidades e amortecimentos, dentre outros parâmetros do modelo [28].

Assumindo as ativações normalizadas, com valores entre 0 e 1, dos músculos TA ( $a_{TA}$ ), CT ( $a_{CT}$ ) e LCA ( $a_{LC}$ )<sup>8</sup>, Titze e Story propuseram, então, as seguintes regras para os parâmetros do modelo baseados na atividade muscular [28]:

- **Regra do prolongamento:** Tomando como base experimentos realizados em laboratório com laringes caninas<sup>9</sup> [73], propõe que o prolongamento das pregas vocais pode ser expresso partindo da seguinte expressão:

$$\epsilon = G(Ra_{CT} - a_{TA}) - Ha_{LC}, \quad (3.57)$$

em que  $\epsilon$  é a tensão longitudinal nas pregas vocais e G, R e H são constantes experimentais denominadas, respectivamente, ganho de prolongamento, taxa de torque e fator

<sup>8</sup>Ao músculo LCA é permitido assumir valores negativos, para representar, em conjunto, a atividade do músculo PCA realizando a abdução da glote. Dessa maneira, os valores para  $a_{LC}$  poderiam variar entre -1 e 1, sendo que os valores negativos representam a atividade do PCA.

<sup>9</sup>A atividade vocal das pregas vocais caninas em muito se assemelha à humana.

de tensão adutora. O comprimento resultante das pregas vocais  $L$  pode ser calculado como segue:

$$L = L_0 (1 + \epsilon), \quad (3.58)$$

sendo que  $L_0$  é o comprimento de repouso das pregas vocais. Em média, para homens  $L_0 = 1,6\text{cm}$  e para mulheres  $L_0 = 1,0\text{cm}$ ;

- **Regra do ponto nodal:** Essa regra define a posição do ponto nodal  $z_n$  com base na contração do TA que faz com que as pregas vocais se aduzam mais em sua porção inferior do que em sua porção superior. Isso reduz a convergência das pregas (como será visto na regra da convergência) e o ponto nodal se desloca para cima, ocasionando maior vibração na porção inferior da glote. O ponto nodal se associa ao ponto do *mucosal upheaval*, observado por Yumoto et al. [74], e é calculado como segue:

$$z_n = (1 + a_{TA}) \frac{T}{3}; \quad (3.59)$$

- **Regras da espessura e da profundidade:** O encurtamento das pregas vocais tende a aumentar sua espessura, bem como o prolongamento possui o efeito oposto. Em geral, a maior parte das variações de comprimento causa variação de espessura. Dessa maneira, foi proposta a seguinte regra para a espessura:

$$T = \frac{T_0}{1 + 0,8\epsilon}, \quad (3.60)$$

em que  $T_0$  é a espessura de repouso das pregas vocais.

As regras para a profundidade são divididas entre a profundidade do corpo e a profundidade da cobertura, sendo dadas, respectivamente, pelas seguintes expressões:

$$D_b = \frac{a_{TA}D_{mus} + 0,5D_{lig}}{1 + 0,2\epsilon}, \quad (3.61)$$

$$D_c = \frac{D_{muc} + 0,5D_{lig}}{1 + 0,2\epsilon}, \quad (3.62)$$

em que  $D_{muc}$  equivale à profundidade da mucosa (ou cobertura),  $D_{lig}$  equivale à profundidade do ligamento vocal (ou camada de transição) e  $D_{mus}$  equivale à profundidade da porção interna do músculo TA (ou corpo)<sup>10</sup>;

- **Regra da adução:** A regra proposta para adução define a abertura pré-fonatória para a porção superior da glote  $\xi_{02}$  (ver Figura 3.9) de acordo com a seguinte expressão:

$$\xi_{02} = 0,25L_0 (1 - 2a_{LC}). \quad (3.63)$$

Dessa maneira, para  $a_{LC} = 0,5$  temos a glote com adução completa (totalmente fechada), i.e.,  $\xi_{02} = 0$ . Para  $a_{LC} < 0,5$ , a adução é fraca e a glote mantém aberturas ( $\xi_{02} > 0$ ). Já para  $a_{LC} > 0,5$ , a adução é forte e as pregas vocais são pressionadas uma contra a outra ( $\xi_{02} < 0$ )<sup>11</sup>. Quando  $a_{LC}$  assume valores negativos, considera-se que as glote

<sup>10</sup>Como explanado em 2.2.2.1 no Capítulo 2.

<sup>11</sup>Valores negativos da adução representam as pregas vocais pressionadas uma contra a outra.

apresenta um caráter abdutor;

- **Regra da convergência:** A convergência das pregas vocais é regido em essência pela atividade do músculo TA. Com isso, a regra proposta para convergência é dada como segue:

$$\xi_c = \xi_{01} - \xi_{02} = T(0,05 - 0,15a_{TA}). \quad (3.64)$$

Com essa regra, o ângulo de convergência da glote é dado de acordo com a seguinte expressão:

$$\tan(\theta) = \frac{\xi_c}{T}, \quad (3.65)$$

de forma que os valores de  $\theta$  podem variar entre  $-6^\circ$  e  $3^\circ$ . Para  $a_{TA} = 0,33$  temos que a glote assume uma postura retangular, i.e.,  $\xi_c = 0$ .

Os parâmetros diretos do modelo são obtidos, com base nos valores calculados pelas regras apresentadas acima, de acordo com as seguintes expressões:

$$M = \rho L T D_b, \quad (3.66)$$

$$K = 2\mu_b L T / D_b + \pi^2 \sigma_b (D_b / L) T, \quad (3.67)$$

$$B = 0,2(KM)^{\frac{1}{2}}, \quad (3.68)$$

$$m = \rho L T D_c, \quad (3.69)$$

$$k = 2\mu_c (L T / D_c) + \pi^2 \sigma_c (D_c / L) T, \quad (3.70)$$

$$b = 0,2(km)^{\frac{1}{2}}, \quad (3.71)$$

$$I_c = \rho L T D_c T^2 \left[ \frac{1}{3} - \frac{z_n}{T} \left( 1 - \frac{z_n}{T} \right) \right], \quad (3.72)$$

$$\kappa = \frac{1}{2} \mu_c L T D_c + \pi^2 \sigma_c (D_c / L) T^3 \left[ \frac{1}{3} - \frac{z_n}{T} \left( 1 - \frac{z_n}{T} \right) \right], \quad (3.73)$$

$$B_c = 0,2(\kappa I_c)^{\frac{1}{2}}. \quad (3.74)$$

em que  $\rho$  é a densidade do tecido<sup>12</sup>,  $\mu_b$  e  $\mu_c$  são os módulos de cisalhamento do corpo e da cobertura, respectivamente<sup>13</sup>, e  $\sigma_b$  e  $\sigma_c$  são as tensões do corpo e da cobertura, respectivamente, dadas por:

$$\sigma_b = [0,5\sigma_{lig} D_{lig} + \sigma_{mus} D_{mus}] / D_b, \quad (3.75)$$

$$\sigma_c = [\sigma_{muc} D_{muc} + 0,5\sigma_{lig} D_{lig}] / D_c, \quad (3.76)$$

sendo  $D_{muc}$ ,  $D_{lig}$  e  $D_{mus}$  as profundidades e  $\sigma_{muc}$ ,  $\sigma_{lig}$  e  $\sigma_{mus}$  as tensões das camadas, respectivamente, da mucosa, do ligamento e do feixe tireovocal do músculo TA [28].

<sup>12</sup>Titze e Story [28] definem  $\rho = 1040\text{kg/m}^3$ , mantendo a concordância com medidas laboratoriais de ondas da mucosa [75, 56].

<sup>13</sup>Com base em medições feitas em laringes extirpadas [76, 77], temos que  $\mu_b = 1000\text{Pa}$  e  $\mu_c = 500\text{Pa}$  [28].

Tabela 3.2 – Constantes usadas para o cálculo das Equações (3.77) e (3.78).

(Fonte: Titze e Story [28])

Camada	$\epsilon_1$	$\epsilon_2$	$\sigma_0$ (kPa)	$\sigma_2$ (kPa)	$C$	$\sigma_m$ (kPa)	$\epsilon_m$	$b$
Mucosa	-0,5	0,35	0,5	30,0	4,4	--	--	--
Ligamento	-0,5	-0,00	0,4	1,393	17,0	--	--	--
Músculo TA	-0,5	-0,05	1,0	1,50	6,5	105	0,4	1,07

A tensão do músculo possui componentes ativos e passivos, sendo dados pela seguinte expressão:

$$\sigma_{mus} = a_{TA}\sigma_{am} \max(0, 1 - b(\epsilon - \epsilon_m)^2) + \sigma_p, \quad (3.77)$$

em que  $\sigma_{am}$  é tensão máxima ativa das fibras do músculo TA, e  $\sigma_p$  é a tensão passiva. As tensões da mucosa e do ligamento são apenas o valor da tensão passiva. As tensões  $\sigma_{muc}$ ,  $\sigma_{lig}$  e  $\sigma_{mus}$  podem ser calculadas usando os valores da Tabela 3.2 e de acordo com a seguinte expressão<sup>14</sup>:

$$\sigma_p = \begin{cases} 0, & \text{para } \epsilon < \epsilon_1, \\ -\frac{\sigma_0}{\epsilon_1}(\epsilon - \epsilon_1), & \text{para } \epsilon_1 \leq \epsilon \leq \epsilon_2, \\ -\frac{\sigma_0}{\epsilon_1}(\epsilon - \epsilon_1) + \sigma_2 [e^{C(\epsilon - \epsilon_2)} - C(\epsilon - \epsilon_2) - 1], & \text{para } \epsilon > \epsilon_2, \end{cases} \quad (3.78)$$

em que  $\sigma_0$  é a tensão quando  $\epsilon = 0$ ,  $\sigma_2$  é um fator de escala,  $\epsilon_1$  é o nível de deformação em que a porção linear da tensão passiva chega a zero e  $\epsilon_2$  é o nível de deformação em que a curva exponencial da tensão passiva se inicia [28].

### 3.1.3.2 Variações do Modelo

Diferentes modelos *Cover-Body* também foram propostos [78, 79, 80, 81, 82, 18], para atender à necessidade de modelos que representem o comportamento glotal em diferentes condições fisiológicas. O modelo multi-massas proposto por Titze [78, 79] consistiu na utilização de 16 massas compondo a estrutura *body-cover* das pregas vocais, sendo 8 massas no corpo e 8 massas na cobertura. O número elevado de massas trouxe uma maior capacidade de adequação fisiológica para o modelo, porém acrescentou uma maior complexidade nos equacionamentos e um maior custo computacional. O principal objetivo desse modelo era de possibilitar o estudo mais acurado dos modos de oscilação das pregas vocais. Numa abordagem mais simples, Tokuda et al. [80] propuseram um modelo corpo-cobertura com 4 massas, utilizando 3 massas na cobertura, para uma melhor adequação fisiológica com a alteração de registros na produção vocal. Zanãrtu et al. [81] propõe uma extensão do modelo

<sup>14</sup> $\epsilon$  é dado pela regra do prolongamento, descrita anteriormente.

com o acréscimo de uma abertura fixa encontrada na região cartilaginosa da glote e Galindo et al. [82] expande essa ideia, adicionando a diferença de fase horizontal, de maneira similar ao proposto por Birkholz et al [66, 67] para o modelo de duas massas, por meio de uma geometria triangular no sentido póstero-anterior. Santos<sup>15</sup> et al. [18] propõe uma adaptação do modelo atualizado por Titze e Story [71, 28] para a representação das pregas vocais com a presença de um pólipô, adaptando a ideia de Zhang e Jiang [17] a um equacionamento mais detalhado para a aerodinâmica glotal.

### 3.1.4 Outros Modelos: Método dos Elementos Finitos

O crescente avanço computacional e o desenvolvimento de novos métodos no decorrer dos anos possibilitou a implementação de modelos mais complexos das pregas vocais. Alguns dos modelos que melhor conseguem representar a complexidade vibratória das pregas vocais são baseados no método dos elementos finitos [83, 84, 16, 85, 86, 87, 88, 89]. Dentre exemplos relevantes na literatura que empregam o métodos dos elementos finitos para a simulação das pregas vocais, pode-se destacar os modelos propostos por Alipour-Haghighi e Titze [83] e Alipour-Haghighi et al. [84], por Rosa [16] e Rosa et al. [85], por Xue et al. [86, 87] e por Greiss et al. [88, 89].

## 3.2 Modelos de Propagação de Ondas no Trato Vocal

O trato vocal recebe o sinal produzido nas pregas vocais e o modifica por meio das ressonâncias características da geometria formada pela movimentação e posicionamento de seus articuladores. Para que a propagação do som no trato vocal seja bem simulada, modelos acurados de seu formato são necessários e, em geral, conhecer a área transversal do trato é de grande importância [39, 90, 91].

Dentre diversos modelos propostos para o equacionamento e a simulação do efeito do trato vocal, dois métodos se destacam por serem frequentemente aplicados para a síntese de voz com modelos matemáticos computacionais das pregas vocais: o modelo de circuito acústico [27], que é baseado na analogia com circuitos elétricos, e o modelo de linha de transmissão analógica por reflexão de ondas [26]. Em ambos os métodos, utiliza-se a área transversal discretizada, como ilustrado na Figura 3.10, de modo que a propaga-

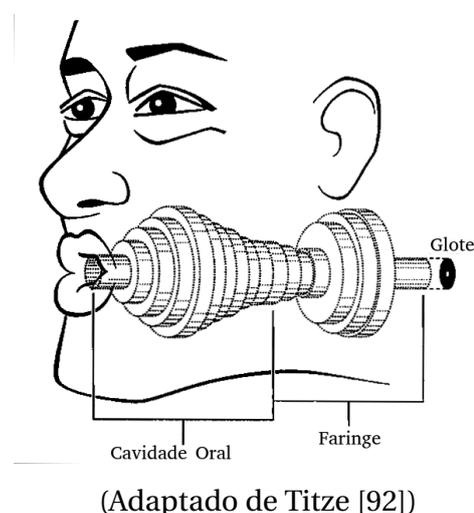


Figura 3.10 – Discretização do Trato Vocal.

<sup>15</sup>Autor desta dissertação.

ção da onda no trato vocal é implementada na sequência de tubos concatenados resultantes da discretização.

### 3.2.1 Circuito Acústico

O modelo de circuito acústico [27, 23] consiste em realizar uma analogia das equações de propagação de onda no trato vocal com as de um circuito elétrico RLC, de maneira que o fluxo de ar seja equivalente à corrente elétrica e a pressão do ar equivalente à tensão elétrica [39]. Assim, para cada tubo discretizado do trato vocal, é definida uma malha RLC equivalente, permitindo simular o trato vocal de acordo com as relações de circuito.

Em geral, como no modelo de Ishizaka e Flanagan [24], para a excitação do trato vocal é utilizado um circuito elétrico análogo ao modelo massa-mola das pregas vocais, com os parâmetros variáveis em função da área entre as massas. Assim, podemos representar o circuito acústico do trato vocal, acoplado à excitação glotal, como ilustrado na Figura 3.11.

Os parâmetros do circuito podem ser calculados com base na área e no comprimento de cada seção transversal  $j$  do trato vocal, para um trato dividido em  $n$  seções, de acordo com as seguinte expressões [24]:

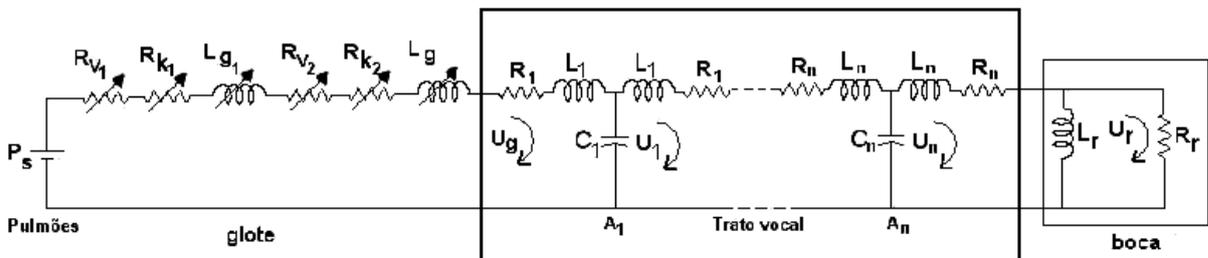
$$R_j = \frac{s_j}{A_j^2} \sqrt{\rho \mu \omega / 2}, \quad j = 1, 2, \dots, n, \quad (3.79)$$

$$C_j = \frac{\ell_j A_j}{\rho c^2}, \quad j = 1, 2, \dots, n, \quad (3.80)$$

$$L_j = \frac{\rho \ell_j}{2A_j}, \quad j = 1, 2, \dots, n, \quad (3.81)$$

em que  $c$  é a velocidade do som,  $A_j$  é a área da  $j$ -ésima seção transversal,  $s_j$  é o perímetro da circunferência da  $j$ -ésima seção e  $\omega$  é a frequência em radianos por segundo.

Para representar o efeito da radiação, que ocorre na boca, a resistência e a indutância



(Fonte: Cataldo et al. [2])

Figura 3.11 – Circuito acústico do trato vocal.

equivalentes são calculadas como segue:

$$R_r = \left( \frac{128\rho c}{9\pi^2 A_n} \right), \quad (3.82)$$

$$L_r = \frac{8\rho}{3\pi} \sqrt{\pi A_n}, \quad (3.83)$$

em que  $A_n$  é a área da última seção transversal do trato, que corresponde à própria boca.

A propagação das ondas no trato vocal pode ser obtida equacionando o circuito, incluindo a malha da glote. Para isso, pode ser utilizado o método das correntes de malha, resultando nas seguintes expressões [24]:

- Malha  $U_g$ :

$$(R_{k1} + R_{k2}) |U_g| U_g + (R_{v1} + R_{v2}) U_g + (L_{g1} + L_{g2}) \frac{dU_g}{dt} + L_1 \frac{dU_g}{dt} + R_1 U_g + \frac{1}{C_1} \int_0^t (U_g - U_1) dt - P_s = 0; \quad (3.84)$$

- Malha  $U_1$ :

$$(L_1 + L_2) \frac{dU_1}{dt} + (R_1 + R_2) U_1 + \frac{1}{C_2} \int_0^t (U_1 - U_2) dt + \frac{1}{C_1} \int_0^t (U_1 - U_g) dt = 0. \quad (3.85)$$

Esse padrão é repetido até chegar a última seção do trato (penúltima malha);

- Malha  $U_n$ :

$$(L_n + L_R) \frac{d(U_n)}{dt} + R_n U_n - L_R \frac{d(U_R)}{dt} + \frac{1}{C_n} \int_0^t (U_n - U_{n-1}) dt = 0; \quad (3.86)$$

- Malha  $U_r$  (radiação):

$$L_R \frac{d(U_R - U_n)}{dt} + R_R U_R = 0. \quad (3.87)$$

Para calcular o sinal de pressão emitido na fonação, basta calcular a tensão no resistor  $R_R$ , utilizando  $U_R$ .

### 3.2.2 Linha de Transmissão Analógica por Reflexão de Ondas

O modelo de linha de transmissão analógica por reflexão de ondas é baseado no conceito de relações de onda e coeficientes de reflexão, comumente utilizados em sistemas de rádio e de radar. Esse modelo é também conhecido como modelo de Kelly e Lochbaum (ou modelo K-L) [26], em homenagem aos autores do primeiro trabalho a empregar essa ideia, e é uma abordagem de grande utilidade por combinar uma simplicidade conceitual e computacional que possui uma relação muito próxima aos princípios físicos do fenômeno da propagação de onda no trato.

No modelo K-L, se assume a ocorrência de uma propagação de onda planar unidimensional. Dessa maneira, é possível utilizar a variação das áreas de seção transversal do tubo para obter um agrupamento de coeficientes de transmissão. Assume-se que, na junção entre dois tubos quaisquer do trato vocal ( $A_i$  e  $A_{i+1}$ ) as pressões e fluxos podem ser expressas em termos de ondas que se propagam para frente e para trás no tubo, como ilustrado na Figura 3.12. As ondas  $P_i^+$  e  $P_{i+1}^-$  são antecessoras no tempo das ondas  $P_{i+1}^+$  e  $P_i^-$ . Estas, por sua vez, são calculadas usando aquelas, como segue:

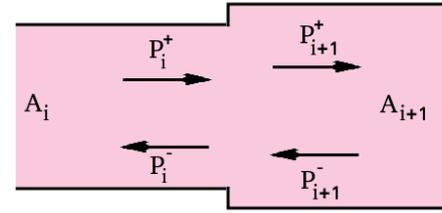


Figura 3.12 – Ondas de pressão propagadas para frente e para trás.

$$P_{i+1}^+ = (1 + r_i)P_i^+ - r_i P_{i+1}^-, \quad (3.88)$$

$$P_i^- = (1 - r_i)P_{i+1}^- + r_i P_i^+, \quad (3.89)$$

em que  $r_i$  é o coeficiente de reflexão da junção entre a  $i$ -ésima e a  $i + 1$ -ésima seção transversal do trato vocal, dada por:

$$r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}}. \quad (3.90)$$

É necessário que a defasagem do tempo da propagação para a transição seja respeitada. Uma maneira simples de realizar isso para a simulação digital é assumir que a taxa de amostragem do sistema seja  $F_a = \frac{c}{2D}$  ou  $F_a = \frac{c}{D}$ , em que  $D$  é o comprimento de cada seção. Isso equivale a criar um atraso, respectivamente, de  $z^{-1/2}$  ou  $z^{-1}$  na representação da transformada Z [93]. Em geral, a primeira opção é a mais utilizada, por permitir o uso de uma menor taxa de amostragem e por realizar a propagação de onda de duas seções transversais do trato vocal a cada interação [94]. Entretanto, essa escolha requer que sejam contabilizadas, a cada interação, primeiramente a pressão  $P_1^+$ , dada por [94]:

$$P_1^+ = U_g \frac{\rho c}{A_1} + r_g P_1^-, \quad (3.91)$$

em que  $r_g$  é um coeficiente de reflexão glótica, sendo seguida pelas seções pares do trato e, por fim, as seções ímpares. Na primeira interação do método, todas as pressões são consideradas nulas.

A radiação que ocorre nos lábios não foi abordada por Kelly e Lochbaum [26]. Apesar disso, diferentes abordagens podem ser utilizadas para acrescentar o efeito de radiação ao modelo, dentre as quais podemos destacar: (1) utilizar um circuito RL no sinal resultante do modelo K-L [90]; (2) assumir, de maneira simplificada, um coeficiente de reflexão para a radiação, de modo que  $P_n^- = r_l P_n^+$  e a pressão do som emitido  $P_{\text{out}} = P_n^+ + P_n^-$  [94].

O modelo K-L, como proposto por Kelly e Lochbaum [26], é um modelo que desconsidera perdas de energia da onda na propagação. Entretanto, esse modelo serve como

uma estrutura básica para a propagação de onda, sendo possível acrescentar uma maior complexidade à propagação, acrescentando o efeito das perdas que ocorrem no interior do trato vocal, como feito nos trabalhos de Liljencrants [72] e de Story [90]. Também pode ser simulado o tubo da traqueia utilizando o modelo K-L, propagando a pressão pulmonar até chegar à região subglotal.

## CAPÍTULO

## 4

## Qualidade Vocal, Disfonias e Patologias Vocais

Este capítulo possui o objetivo de apresentar os conceitos de qualidade vocal, disfonias e patologias vocais. Nele também é realizada a revisão bibliográfica da síntese de voz com qualidade vocal alterada (disfônica), descrevendo os principais trabalhos, os métodos empregados e os resultados obtidos.

### 4.1 Qualidade Vocal

Diversas pesquisas foram e são conduzidas buscando estudar a incidência e as características de diversos tipos de qualidade vocal. Entretanto, não existe um termo único e amplamente aceito, nas diversas áreas de estudo, para o termo qualidade vocal [7]. O termo pode assumir diferentes significados para públicos diferentes, como pontuado por Childers e Lee [95], de forma que pode representar diferenças articulatórias para um foneticista, pode se definir diferentes registros vocais para um cantor, pode descrever características da voz, como sopro, rouquidão e aspereza, entre outros e pode se relacionar com os padrões vibratórios das pregas vocais.

De acordo com Behlau et al. [96], qualidade vocal pode ser definida como o conjunto de características que definem a voz no âmbito da percepção, se relacionado à impressão total que uma voz gera no ouvinte. Ferrand [7] aponta a qualidade vocal como um conceito complexo e que envolve diversos aspectos da produção vocal, sendo regada pela vibração das pregas vocais e pelo ajuste do trato vocal.

### 4.2 Voz Normal, Disfonias e Patologias Vocais

A ideia de voz normal é, frequentemente, definida como a ausência de problemas na voz [7]. Entretanto Colton e Casper [9] pontuam que não existe uma definição aceita de voz normal, sendo essa indefinição um problema para a definição de objetivos em tratamentos e até mesmo a avaliação de alterações vocais. Diversos autores propuseram formas de definir

voz normal e diferenciá-la da voz alterada, caracterizando os tipos de vozes por meio de parâmetros acústicos como frequência fundamental, *jitter*, *shimmer* e medidas de proporção de ruído, dentre outros [7]. Também houve a proposição de escalas e protocolos para avaliação da qualidade vocal de acordo com a percepção do especialista [97, 98].

Às vozes alteradas é designado o nome de disфонia [7]. Num sentido mais amplo, pode-se definir disфонia como

"[...] um distúrbio da comunicação oral, no qual a voz não consegue cumprir seu papel básico de transmissão da mensagem verbal e emocional de um indivíduo" [99].

Tal definição permite englobar todo tipo de alteração vocal, com as mais diversas causas e efeitos, sejam essas alterações patológicas ou não.

#### 4.2.1 Classificação Etiológica das Disfonias

A classificação de disfonias pode ser realizada de acordo com diversos critérios, porém a classificação etiológica se destaca pelo fato de classificar as disfonias pelos fatores que causam a alteração observada [99]. Inúmeras variações de métodos de classificação etiológica foram propostos na literatura. Dentre esses métodos, encontra-se o método de Behlau e Pontes [100], que classifica as disfonias em 3 categorias: disfonias funcionais, disfonias orgânicas e disfonias organofuncionais.

As disfonias funcionais são caracterizadas pela presença de alterações vocais comportamentais, i.e., as alterações vocais não decorrem de uma lesão ou uma patologia vocal [99]. Essas alterações podem ser causadas pelo uso incorreto da voz, inaptações vocais ou alterações psicogênicas [100].

As disfonias orgânicas se caracterizam pela presença de lesões ou doenças que impedem o uso adequado da voz, i.e., são alterações de causa patológica. Essas disfonias podem ter dois tipos de causas principais [99]:

- Alterações nos órgãos da comunicação, como doenças congênitas, lesões traumáticas, doenças inflamatórias, neoplasias e problemas auditivos;
- Alterações em outros órgãos do corpo, como distúrbios endocrinológicos, síndromes, desordens neurológicas, entre outras.

As disfonias organofuncionais são caracterizadas pela presença de lesões secundárias nas pregas vocais causadas por disfonias funcionais. Em outras palavras, as alterações comportamentais no uso da voz podem causar lesões nas pregas vocais que deem uma nova característica à disфонia [100].

### 4.3 Avaliação da Qualidade Vocal

A avaliação da voz por meio da análise perceptivo-auditiva toma como base a percepção do fonoaudiólogo e é o método mais comumente utilizado para a avaliação da qualidade vocal. Para a sua realização, diversas escalas e protocolos são empregados. A escala GRBASI [97] e o protocolo CAPE-V [98] são os procedimentos mais utilizados para a avaliação clínica e científica, respectivamente.

A GRBASI é uma escala de 4 pontos<sup>1</sup> que avalia os seguintes parâmetros de uma voz: Grau Geral da Disfonia - G; Rugosidade - R; Soprosidade - B; Astenia - A; Tensão - S; Instabilidade - I [97]. Já o CAPE-V (*Consensus Auditory-Perceptual Evaluation of Voice*) é um protocolo para avaliação vocal que envolve regras para as vozes a serem avaliadas e propõe uma avaliação, em escala visual-analógica, dos seguintes parâmetros: gravidade geral da alteração; rugosidade; soprosidade; tensão; *pitch*; *loudness* [98].

Entretanto, a avaliação baseada na percepção de um especialista proporciona uma subjetividade indesejada ao processo e não permite uma conexão direta com as causas das alterações vocais. Kreiman e Gerratt [11] pontuam que, apesar de seu extenso uso, a avaliação perceptivo-auditiva promove uma alta dependência do avaliador, de forma que a falta de parâmetros objetivos façam com que a experiência, o treinamento e fatores pessoais e emocionais do especialista influenciem significativamente na avaliação. Este fato torna questionável a confiabilidade desse procedimento de análise. Alguns trabalhos pontuam que algumas características acústicas da voz permitem diferenciar alterações vocais com maior acurácia que a percepção de especialistas [101, 7]. De igual forma, o uso de análise-pela-síntese para, com base em parâmetros acústicos como instrumento de controle, também se mostra mais promissor para a avaliação de alterações vocais [10, 11]. Apesar disso, a avaliação perceptivo-auditiva ainda é assumida como a principal ferramenta (padrão ouro) para avaliação da voz [96, 102, 103].

### 4.4 Síntese de Vozes Alteradas

Utilizando diferentes abordagens, diversos trabalhos foram publicados em que foi realizada a síntese computacional de vozes alteradas, seja na perspectiva de sintetizar qualidades vocais específicas ou de representar as características particulares de alguma patologia.

Esta seção encontra-se dividida entre dois tipos de síntese de vozes, que aqui denominam-se Síntese Digital (4.4.1) e Síntese Mecânica-Acústica (4.4.2) de vozes alteradas. Por síntese digital, compreende-se o processo de geração de amostras vocais com base na teoria fonte-

---

<sup>1</sup>Avalia o nível de cada parâmetro com valores de 0 a 3, representando: 0 - ausência da característica avaliada; 1 - grau leve; 2 - grau moderado; 3 - grau severo.

filtro digital [52], em que o sinal glotal é compreendido como uma entrada para um filtro digital, que representa o trato vocal. Com o termo síntese mecânica-acústica, propõe-se embarcar processos que levem em conta a mecânica da oscilação glotal e/ou a ressonância acústica no trato vocal para a produção das amostras de voz.

#### 4.4.1 Síntese Digital de Vozes Alteradas

Em 1990, Klatt e Klatt [104] propuseram um modelo de síntese de diferentes modos de fonação com base em novo um modelo paramétrico para representar o sinal glotal (KLGLOT88) e um sintetizador digital de formantes. No ano seguinte, utilizando uma abordagem similar à proposta, Lalwani e Childers [105] implementaram um sintetizador digital de vozes patológicas, com o qual produziram vozes soprosas, crepitantes, roucas e rugosas. Nesse trabalho também foi proposto um modelo paramétrico para o fluxo glotal, sendo observado, através da avaliação perceptiva de voluntários, que o desempenho de naturalidade das amostras foi superior aos dos modelos de Liljencrants-Fant (LF) [106] e o KLGLOT88 [104].

O trabalho de Bangayan et al. [107], em 1997, utilizou o sintetizador de Klatt e Klatt [104] para gerar amostras de voz patológica. A ideia central do trabalho consistiu em sintetizar amostras que buscassem copiar vozes patológicas e encaminhar para avaliação de especialistas. Os ouvintes receberam diversos pares de amostra natural e sintetizada, devendo avaliar, em uma escala de 1 a 7, o quão adequada ao sinal original foi a síntese. Apesar de obter bons resultados, foram apontadas limitações no sintetizador devido à baixa disponibilidade de formantes ajustáveis e foram sugeridas alterações para melhorar a síntese de vozes patológicas.

Baseando-se parcialmente nessa proposta, Gerratt e Kreiman [10], em 2001, implementaram um novo sintetizador, com uma disponibilidade maior de parâmetros para controle. Nesse trabalho, foi realizada a análise-pela-síntese de vozes alteradas como ferramenta de avaliação. Para isso, os especialistas ajustavam os parâmetros do sintetizador buscando que o sinal produzido se assemelhasse ao máximo possível com a amostra de voz avaliada. Entretanto, o excesso de parâmetros para a análise-pela-síntese tornaria difícil o ajuste feito pelo especialista. Por isso, foi estimada a maior parte dos parâmetros do sinal para a ressíntese, deixando apenas o SNR<sup>2</sup> como único parâmetro disponível para o ajuste. Os resultados desse experimento demonstraram uma acurácia e uma concordância inter e intra avaliador significativamente maior do que na análise perceptivo-auditiva sobre as mesmas amostras de voz.

Em 2005, Kreiman e Gerratt [108] estenderam a aplicação desse sintetizador, disponibilizando para os especialistas o ajuste do NSR<sup>3</sup>, *jitter* e *shimmer* das amostras sintetizadas.

---

<sup>2</sup> *Signal-to-noise ratio*

<sup>3</sup> *Noise-to-signal ratio* (inverso do SNR)

Porém, apesar de uma elevada concordância no ajuste do NSR ser novamente observada, os ouvintes não concordaram entre si quanto aos níveis de *jitter* e *shimmer*. Os resultados sugeriram que, apesar de suas importantes contribuições espectrais, o *jitter* e o *shimmer* não conseguem ser totalmente identificados e separados pelos ouvintes, de modo que esse parâmetros não apresentam uma relação clara com a qualidade vocal percebida.

#### 4.4.2 Síntese Mecânica-Acústica de Vozes Alteradas

Em 1976, Ishizaka e Isshiki [109] foram pioneiros na síntese computacional de vozes alteradas, utilizando um modelo de duas massas com assimetria entre as pregas para simular condições patológicas das pregas vocais. O uso de tensões desbalanceadas, i.e., diferentes valores de massas e rigidez das molas em cada prega, permitiu observar comportamentos similares a vibrações glóticas observadas em experimentos fisiológicos com desbalanço de tensão patológico. O controle do desbalanço entre as tensões permitiu a síntese de vozes percebidas como rugosas, soprosas, roucas e diplofônicas. Entretanto, também foi observado que pequenos desequilíbrios nas tensões não provocam alterações perceptivas na voz.

Steinecke e Herzel [64], em 1995, também utilizaram o desbalanço de tensão nas pregas vocais, aplicando a dinâmica não-linear para seu estudo, observando o comportamento das bifurcações no modelo. Foi utilizado nesse trabalho o modelo de duas massas de Ishizaka e Flanagan [24], sendo feitas adaptações ao equacionamento de pressão e de força de colisão. Além de possibilitar uma melhor compreensão da dinâmica não-linear laríngea, as condições analisadas do modelo assimétrico utilizado permitiram produzir vozes diplofônicas e crepitantes.

Um modelo da laringe baseado no método dos elementos finitos foi proposto para síntese de voz por Rosa [16], em 2002. O modelo proposto também foi utilizado para fazer algumas reflexões a respeito de patologias laríngeas. Para isso, algumas alterações foram feitas à estrutura original proposta no modelo, permitindo simular vozes alteradas pelas seguintes patologias: câncer de laringe, paralisia laríngea unilateral e sulco vocal.

Em 2004, Zhang e Jiang [17] propuseram um modelo baseado no trabalho de Steinecke e Herzel [64], porém com a adição de uma massa extra para a representação de um pólipó vocal e com a tensão das pregas balanceadas<sup>4</sup>. Com esse modelo foram executadas análises de diagramas de bifurcação para o estudo da dinâmica de oscilação das pregas vocais com a presença de um pólipó. Efeitos de aperiodicidade e alterações no fechamento glotal foram observados nas simulações das pregas vocais com pólipó.

Zhang et al. [110], em 2005, apresentaram uma formulação para a representação da voz de indivíduos com doença de Parkinson, com base no equacionamento descrito por Steinecke e Herzel [64] para o modelo de duas massas. Para uma boa representação das

---

<sup>4</sup>Dessa maneira, a única fonte de perturbação assimétrica seria apenas o pólipó.

características fonatórias de um indivíduo com Parkinson, foram propostas: a) o aumento da rigidez das pregas vocais, com base no aumento de frequência observado em pacientes; b) uma oscilação temporal no valor da rigidez das pregas do tipo  $K = a + b \sin(\omega_T t)$ , para a representação do tremor vocal típico. Foi sugerido o uso de frequências de oscilação de 5Hz. Com base nesse modelo também foi desempenhada uma análise de bifurcação da atividade das pregas vocais. Posteriormente, em 2008, Zhang e Jiang [111] aperfeiçoaram o modelo para o tremor vocal, permitindo adicionar mais complexidade à oscilação produzida na rigidez vocal e expandindo a aplicação para outros tremores além dos causados pela doença de Parkinson.

O desbalanço de tensões foi utilizado por Xue et al. [112], em 2010, aplicado ao estudo da paralisia laríngea unilateral. A síntese foi baseada no modelo de Ishizaka e Flanagan [24], aplicando um equacionamento aerodinâmico mais refinado, derivado diretamente das equações de Navier-Stokes. A simulação permitiu constatar que o desbalanço nas pregas torna necessária uma maior pressão subglotal para que a fonação ocorra e requer um maior esforço aerodinâmico para aumentar a intensidade vocal. Constatou-se também que a frequência fundamental passa a depender, quase unicamente, da prega vocal que oscila normalmente. A simulação desse sistema produziu vozes de qualidade soprada.

Um modelo multi-massas foi proposto e utilizado na análise do controle simétrico e assimétrico da rigidez de suas molas e das suas massas, em 2012, por Fraile et al. [113]. O aumento da rigidez das molas no caso simétrico exerceu influência significativa apenas no aumento da  $f_0$ , ao passo que no caso assimétrico houve uma influência moderada no aumento da  $f_0$  e do *jitter* e uma influência significativa no aumento do *shimmer*. Por outro lado, o aumento da massa no caso assimétrico só se diferenciou do caso simétrico por exercer uma menor influência na redução da  $f_0$ .

Fraj et al. [114] apresentaram em 2012 um sintetizador de vozes alteradas que, seguindo a equação de Titze [69], realiza o cálculo do fluxo glotal utilizando um modelo paramétrico de área glotal, e representa o trato vocal pelo modelo K-L de tubos concatenados [26]. Na produção de vozes alteradas, para o ruído glótico se utilizou ruído gaussiano aditivo branco, e foram propostos modelos para o *jitter*, o *shimmer* e o tremor vocal. Com os sinais sintetizados em diferentes níveis dos parâmetros modelados, foram feitos experimentos com ouvintes voluntários para analisar a capacidade de identificar sons humanos e sintetizados e para analisar os estímulos sintetizados de acordo com os parâmetros GRB da escala GRBASI [97]. Os resultados obtidos mostraram que o sintetizador foi capaz de representar bem a voz humana, gerando dificuldades na separação entre vozes naturais e sintetizadas com ouvintes experientes e inexperientes. Também foi possível observar que as avaliações de rugosidade foram dependentes do ruído glotal, do *shimmer* e frequência da fonação, as avaliações da soprosoidade dependeram principalmente do ruído glótico e as avaliações do grau geral da disфонia foram dependentes do ruído glotal, do *shimmer* e da interação entre os dois.

Em 2013, regras para a produção de vozes alteradas com base no modelo físico das pregas vocais foram propostas por Lucero et al. [13]. Para isso, foi utilizado um modelo massa-mola similar ao modelo de uma massa com propagação da onda da mucosa de Titze [55]. As regras apresentadas se propõem a representar ruído glótico e *jitter* ao modificar, diretamente, o fluxo glotal e a rigidez da prega, respectivamente. Para o ruído glótico, é proposta a adição de um  $\Delta U_g$  ao fluxo glotal  $U_g$ , de acordo com a seguinte expressão:

$$\Delta U = \beta \epsilon U_g, \quad (4.1)$$

em que  $\beta$  é um fator de escala e  $\epsilon$  é uma variável aleatória que pode assumir valores de  $-1$  e  $1$ , com probabilidade  $0,5$  cada um, que é filtrada por um filtro passa-baixas com banda passante de  $1000\text{Hz}$ . Para o *jitter*, é proposta a adição de um  $\Delta K$  à rigidez elástica da prega  $K$ , como

$$\Delta K = \alpha \epsilon K, \quad (4.2)$$

em que  $\alpha$  é um fator de escala e  $\epsilon$  é a mesma variável aleatória do caso anterior, filtrada por um filtro passa-baixas com banda passante de  $150\text{Hz}$ . Também é proposta uma regra para a assimetria das pregas vocais, equivalente ao proposto por Ishizaka e Isshiki [109].

Baseando-se na ideia de Fraj et al. [114] de aplicar a síntese de vozes alteradas para a avaliação perceptivo-auditiva, Englert et al. [15], em 2017, utiliza as regras propostas por Lucero et al. [13] para a síntese de ruído glótico *jitter*, trocando a variável aleatória  $\epsilon$  por uma distribuição normal com desvio padrão  $\sqrt{\frac{1}{f_s}}$ , em que  $f_s$  é a taxa de amostragem. Entretanto, assumiu-se que a regra do ruído glótico seria equivalente a um modelo para a soproidade (B da GRBASI) e que a regra do *jitter* seria um modelo para a rugosidade (R da GRBASI). Além disso também foi proposta uma regra para a tensão (S da GRBASI), sendo esta modelada com o aumento da rigidez da mola, o aumento da pressão subglótica e a redução da área glotal pré-fonatória (maior adução glótica). Esse trabalho também se difere do de Fraj et al. [114] justamente por usar os modelos diretamente para representar RBS e comparar às avaliações perceptivo-auditivas, ao invés de identificar relações entre as alterações e a percepção dos voluntários. A análise do realismo da síntese também foi investigada, por meio da avaliação de classificar um dado estímulo como voz humana ou sintetizada. O resultado das avaliações indicaram que houve um maior realismo nas vozes femininas e melhor classificação perceptivo-auditiva para vozes soprosas sintetizadas.

Em 2020, Santos<sup>5</sup> et al. [18] propuseram uma adaptação aos modelos body-cover, com base na formulação de Zhang e Jiang [17], para a inclusão da presença de um pólipso nas pregas vocais, como antecipado em 3.1.3.2 no Capítulo 3. Os sinais gerados pelo modelo foram usados como base para a realização de conjecturas analíticas sobre a natureza oscilatória dos pólipos. Com base na análise do sinal produzido, é levantada a hipótese de que a entropia do sinal e dimensão intrínseca local podem ser ferramentas úteis para a detecção da

<sup>5</sup>Autor desta dissertação.

presença de pólipos com base na amostra de voz. Para testar essa ideia, foi feita a estimação da entropia e da dimensão intrínseca de amostras de vozes humanas com e sem a presença de pólipos, sendo as estimativas obtidas dadas como padrões a serem classificadas por um classificador "ingênuo". Os resultados obtidos sugeriram que o modelo proposto pode ser uma boa ferramenta para aprimorar detectores de pólipos vocais. Mais detalhes sobre esse trabalho são apresentados no Capítulo 6.

## CAPÍTULO

## 5

---

## Dimensão Intrínseca e Entropia Diferencial

A análise de dados por dimensão intrínseca é um campo crescente de aplicações relevantes para redução de dimensão de dados [115, 18]. Proposta pela primeira vez por Bennett [116], a noção de dimensão intrínseca é por ele apresentada como (em tradução livre):

"[...] a dimensionalidade de uma coleção de sinais é definida como o número de parâmetros livres requeridos por um gerador de sinais hipotético capaz de gerar uma aproximação próxima de cada sinal na coleção" [116].

Em outras palavras, pode se dizer que a dimensão intrínseca de um conjunto de dados observados equivale à quantidade de graus de liberdade que serão necessários para gerar esse conjunto [117].

Quando os dados são modelados como variáveis aleatórias (VA), a dimensão intrínseca pode ser associada com a dimensão efetiva do espaço de distribuição de probabilidade das variáveis definidas [22]. Essa modelagem também torna parâmetros como a entropia diferencial relevante [118, 115], tendo em vista que a entropia diferencial equivalerá à cardinalidade efetiva (ou volume) do espaço de probabilidade da VA [22].

Dessa maneira, tanto a estimação da entropia diferencial como da dimensão intrínseca assumem papéis relevantes no campo de reconhecimento de padrões. Isso é ainda mais evidente quando se assume que a projeção de dados em espaços de baixa dimensionalidade e alta entropia pode trazer melhorias ao desempenho em tarefas como classificação, regressão, agrupamento e visualização de dados [119, 120, 121, 117, 122, 18].

### 5.1 Método de Montalvão para Análise Visual de Dimensão Intrínseca e Entropia Diferencial

Diversos métodos comumente utilizados para a estimação de dimensão intrínseca tomam como base a dimensão de correlação [119]. Entretanto, um problema comum nesses métodos é o fato de que há uma dependência de uma escala de análise, que produz variações na dimensão estimada. Montalvão [22], em 2019, apresenta um método em que a dimensão

intrínseca e a entropia diferencial são estimadas de maneira conjunta, tirando proveito da dependência da escala como uma ferramenta para a resolução do problema.

O ponto de partida desse método encontra-se no método de estimação de entropia, usando contagem de coincidências, de Ma [123]. Ma utiliza uma abordagem similar à dos estimadores de dimensão de correlação, de forma que, para um conjunto de  $N$  observações  $\{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)\}$  e um raio limiar, a dimensão de correlação pode ser definida como segue:

$$d = \lim_{r \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\log C(r)}{\log r}, \quad (5.1)$$

em que

$$C(r) = \frac{1}{N^2} \sum_{\forall i \neq j} I(\|\mathbf{x}(i) - \mathbf{x}(j)\| \leq r) \quad (5.2)$$

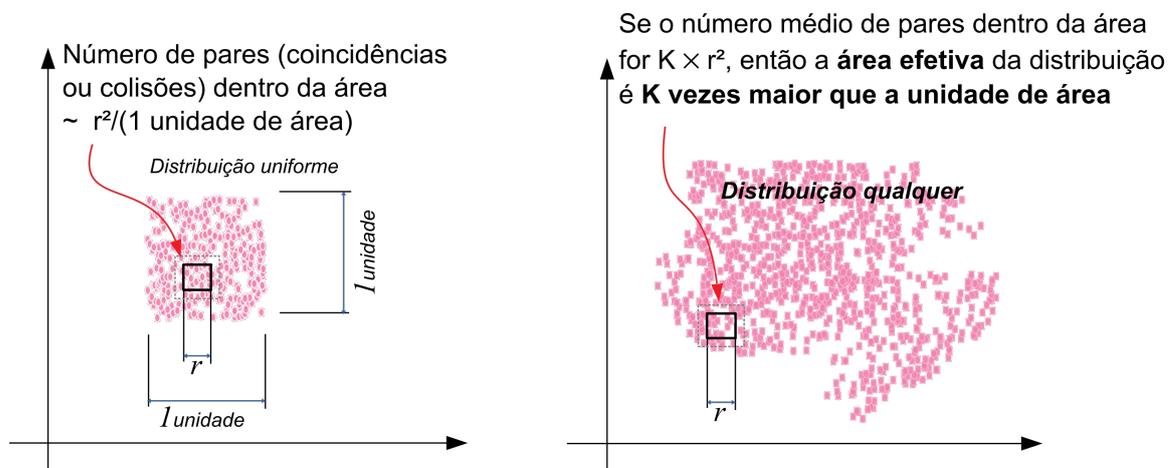
e  $I$  representa uma função do tipo

$$\begin{cases} I(\kappa) = 1, & \text{se } \kappa \text{ for uma condição verdadeira,} \\ I(\kappa) = 0, & \text{caso contrário.} \end{cases}$$

Pode ser dito também, de outra maneira, que a função  $I$  identifica se  $\mathbf{x}(i)$  e  $\mathbf{x}(j)$  se encontram dentro de uma mesma hipersfera de raio  $r$  [124, 22], o que pode ser denominado coincidência. Mantendo essa mesma noção de coincidências, para o mesmo conjunto de observações, a entropia quadrática pode ser estimada como

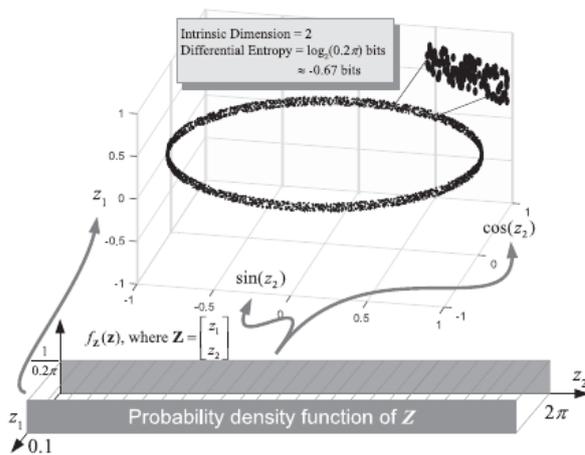
$$\hat{H}(r) = -\log \left( \frac{2}{N(N-1)} \sum_{\forall i < j} I(\|\mathbf{x}(i) - \mathbf{x}(j)\| \leq r) \right). \quad (5.3)$$

Desta vez  $I$ , detecta coincidências em um hipercubo [22]. A contagem de coincidências para a estimação da entropia quadrática é ilustrada na Figura 5.1, para diferentes distribuições.



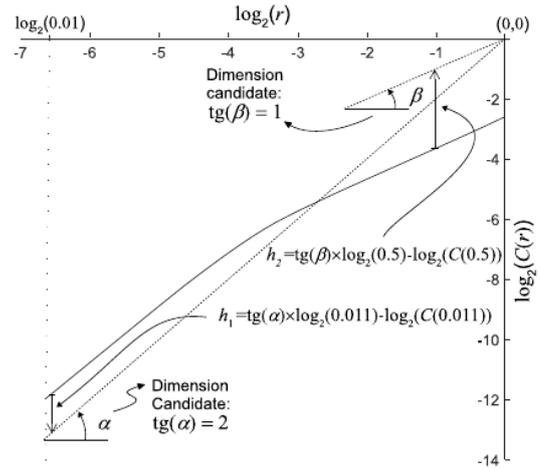
(Fonte: Montalvão [125])

Figura 5.1 – Ilustração da contagem de coincidências para estimação da entropia quadrática (ou entropia de colisão), que é estimada a partir do logaritmo de  $K$ .



(Fonte: Montalvão [22])

Figura 5.2 – Os pontos representam a variável aleatória  $X_c$ , que é gerada a partir de uma variável aleatória latente  $\mathbf{Z} = [0.1U_1; 2\pi U_2]$ , em que  $U_1$  e  $U_2$  são variáveis aleatórias uniformes, que assumem valores entre 0 e 1.



(Fonte: Montalvão [22])

Figura 5.3 – Ilustração da estimação conjunta de dimensão intrínseca e entropia diferencial para 3000 instancias da variável  $X_c$ . É possível observar que diferentes valores de  $r$  podem resultar em estimações diferentes. É possível observar, nesse caso, duas tendências principais.

Montalvão [22] argumenta que, ao utilizar uma mesma definição de coincidência,  $-\log(C(r))$  e  $\hat{H}(r)$  são equivalentes, como pode ser observado das equações (5.2) e (5.3). Utilizando a ideia explicada por Montalvão et al. [126] de que, para variáveis aleatórias contínuas e valores de  $r$  suficientemente pequenos, a entropia quadrática pode ser estimada como

$$\hat{h}(r) = \hat{H}(r) + \log r^{\hat{d}}.$$

Montalvão então propõe substituir  $\hat{H}(r)$  por  $-\log(C(r))$ , o que resulta em

$$\log C(r) = \hat{d} \log r - \hat{h}(r). \tag{5.4}$$

É possível observar que, ao analisar essa equação, sob a perspectiva gráfica do  $\text{plot } \log C(r) \times \log r$ , espera-se encontrar uma reta cuja inclinação é dada por dimensão intrínseca  $\hat{d}$  e que intercepta o eixo  $\log C(r)$  no valor negativo da entropia quadrática estimada  $\hat{h}(r)$  [22].

As Figuras 5.2 e 5.3 ajudam a ilustrar a estimação por meio do método apresentado, sendo este aplicado, como exemplo, para uma variável aleatória  $X_c$  no espaço tridimensional, com uma estrutura intrínseca bidimensional, mas com forte domínio de uma forma circular unidimensional.

É possível observar que o método resulta em duas tendências principais, das quais podem ser estimadas, graficamente, a dimensão intrínseca e a entropia diferencial. Mon-

talvão [22] demonstra que é possível (e apresenta um método para) utilizar duas diferentes tendências para estimar a variância de sinais de ruído.

Essa abordagem apresentada se utiliza de uma característica considerada como um problema nas técnicas de estimação de dimensão intrínseca e cria uma ferramenta útil para a estimação conjunta com a entropia diferencial, bem como para outras aplicações que, como a estimação da variância de sinais de ruído [22], fogem ao escopo deste trabalho.

## CAPÍTULO

## 6

## Modelo das Pregas Vocais com um Pólipo Vocal e Aplicações Potenciais

A presença de pólipos nas pregas vocais é um dos tipos mais comuns de lesão das pregas vocais [20]. Essas lesões têm sua causa no trauma, muitas vezes em decorrência do abuso, mal uso ou uso excessivo da voz [127] ou por atividades físicas violentas não relacionadas com a voz [128]. A presença de um pólipo afeta a vibração das pregas, podendo causar fechamento incompleto durante a fonação e ocasionar padrões oscilatórios irregulares em ambas as pregas (afetadas ou não) [19]. O fonação produzida pelas pregas com pólipo vocal pode ser percebida como rugosa e, em alguns casos, soprosa [128], devido ao movimento aperiódico na glote e a um eventual vazamento de ar das pregas [20].

Este capítulo apresenta o modelo para as pregas vocais com pólipo proposto por Santos<sup>1</sup> et al. [18] em trabalho publicado em 2020. Esse modelo é utilizado para a geração de amostras de voz com e sem a presença de pólipos, sendo estas amostras utilizadas para realizar conjecturas analíticas úteis para aprimorar classificadores de pólipos em vozes humanas.

### 6.1 Modelo *Body-Cover* das Pregas Vocais com um Pólipo

O modelo proposto adapta a estrutura apresentada por Zhang e Jiang [17] ao modelo *body-cover* de 3 massas de Story [25]. Para cada prega vocal, o modelo utiliza uma massa para o corpo e duas massas para a cobertura, acopladas por uma mola linear, e, na prega afetada pela lesão (a prega vocal do lado direito foi arbitrariamente escolhida), uma massa para o pólipo, como ilustrado na Figura 6.1.

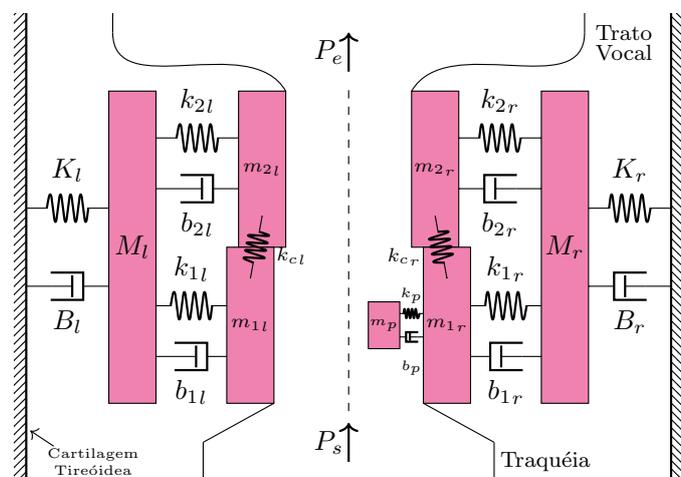


Figura 6.1 – Modelo *Body-Cover* das Pregas Vocais com Pólipo Vocal

<sup>1</sup>Autor desta dissertação.

### 6.1.1 Equações de Movimento

As equações de movimento para as três massas de cada prega vocal  $\alpha$  ( $\alpha = r, l$ ) e para a massa do pólipo são dadas como segue:

$$m_{2\alpha}\ddot{x}_{2\alpha} = F_{k_{2\alpha}} + F_{b_{2\alpha}} - F_{k_{c\alpha}} + F_2 + F_{Col_{2\alpha}}, \quad (6.1)$$

$$m_{1r}\ddot{x}_{1r} = F_{k_{1r}} + F_{b_{1r}} + F_{k_{cr}} + F_1 + F_{Col_{1r}}, \quad (6.2)$$

$$-F_{k_p} - F_{b_p},$$

$$m_{1l}\ddot{x}_{1l} = F_{k_{1l}} + F_{b_{1l}} + F_{k_{cl}} + F_1 + F_{Col_{1l}}, \quad (6.3)$$

$$m_p\ddot{x}_p = F_{k_p} + F_{b_p} + F_{Col_p}, \quad (6.4)$$

$$M_\alpha\ddot{X}_\alpha = F_{K_\alpha} + F_{B_\alpha} - F_{k_{1\alpha}} + F_{d_{1\alpha}} + F_{k_{2\alpha}} + F_{d_{2\alpha}}, \quad (6.5)$$

em que:

- $x_{i\alpha}$  ( $i = 1, 2$ ) são os deslocamentos das massas da cobertura,  $X_\alpha$  é o deslocamento de cada massa do corpo e  $x_p$  é o deslocamento da massa do pólipo;
- $F_{k_{i\alpha}}$ ,  $F_{k_p}$  e  $F_{K_\alpha}$  são as forças das molas agindo nas massas e  $F_{k_{c\alpha}}$  é a força de acoplamento entre as massa da cobertura, dadas por:

$$F_{k_{i\alpha}} = -k_{i\alpha} \left( (x_{i\alpha} - x_{i\alpha_0}) - (X_\alpha - X_{\alpha_0}) \right) + \eta_{i\alpha} \left( (x_{i\alpha} - x_{i\alpha_0}) - (X_\alpha - X_{\alpha_0}) \right)^3, \quad (6.6)$$

$$F_{k_p} = -k_p \left( (x_p - x_{p_0}) - (x_{1r} - x_{1r_0}) \right) + \eta_p \left( (x_p - x_{p_0}) - (x_{1r} - x_{1r_0}) \right)^3, \quad (6.7)$$

$$F_{K_\alpha} = -K_\alpha \left( (X_\alpha - X_{\alpha_0}) + H (X_\alpha - X_{\alpha_0})^3 \right), \quad (6.8)$$

$$F_{k_{c\alpha}} = -k_p \left( (x_{1\alpha} - x_{1\alpha_0}) - (x_{2\alpha} - x_{2\alpha_0}) \right), \quad (6.9)$$

em que  $x_{i\alpha_0}$ ,  $x_{p_0}$  e  $X_{\alpha_0}$  são as posições iniciais das massas,  $k_{i\alpha}$ ,  $k_p$ ,  $K_\alpha$  e  $k_{c\alpha}$  são as constantes das molas e  $\eta_{i\alpha}$ ,  $\eta_p$  e  $H$  são os coeficientes não lineares das molas;

- $F_{b_{i\alpha}}$ ,  $F_{b_p}$  e  $F_{B_\alpha}$  são as forças de amortecimento agindo em cada massa, dadas por:

$$F_{b_{i\alpha}} = -b_{i\alpha} (\dot{x}_{i\alpha} - \dot{X}_\alpha), \quad (6.10)$$

$$F_{b_p} = -b_p (\dot{x}_p - \dot{x}_1), \quad (6.11)$$

$$F_{B_\alpha} = -B_\alpha \dot{X}_\alpha, \quad (6.12)$$

- $F_{Col_{i\alpha}}$  e  $F_{Col_p}$  são as forças agindo nas massas devido à colisão com outras massas, dadas por:

$$F_{Col_{i\alpha}} = -\Theta(-a_i, a_{i_0}) k_{Col_{i\alpha}} (x_{i\alpha} + \eta_{Col_{i\alpha}} x_{i\alpha}^3), \quad (6.13)$$

$$F_{Col_p} = -\Theta(-a_p, a_{p_0}) k_{Col_p} (x_p + \eta_{Col_p} x_p^3), \quad (6.14)$$

em que

$$\Theta(x, x_0) = \begin{cases} \tanh\left(50\frac{x}{x_0}\right), & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (6.15)$$

é a função de colisão, proposta por Steinecke e Herzel [64],  $a_i$  e  $a_p$  são as áreas entre as massas de cada prega e entre o pólipo e a massa inferior da prega do lado esquerdo, respectivamente,  $a_{i_0}$  e  $a_{p_0}$  são as áreas iniciais,  $k_{Col_{i\alpha}}$  e  $k_{Col_p}$  são as constantes de rigidez das colisões e  $\eta_{Col_{i\alpha}}$  e  $\eta_{Col_p}$  são os coeficientes não-lineares de colisão;

- $F_i$  são as forças aerodinâmicas agindo nas massas da cobertura, devido a pressão do ar na glote.

As áreas entre as massas dos dois lados das pregas vocais encontram-se ilustradas na Figura 6.2 e são definidas como segue:

$$a_i = l(x_{ir} + x_{il}), \quad (6.16)$$

$$a_p = l_p(x_p + x_{1l}), \quad (6.17)$$

em que  $l$  é o comprimento da glote e  $l_p$  é o comprimento do pólipo.

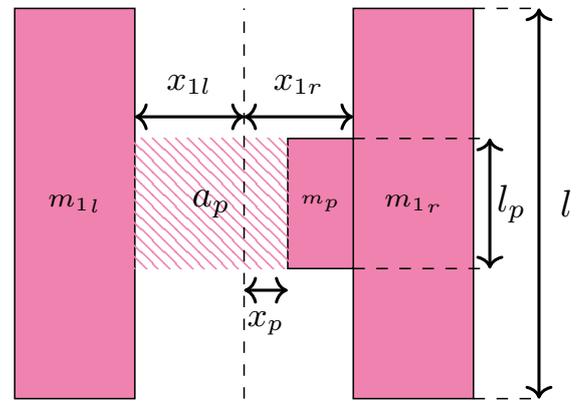


Figura 6.2 – Ilustração das áreas  $a_i$  e  $a_p$ .

### 6.1.2 Aerodinâmica Glotal

As forças externas agindo nas pregas vocais, devido à pressão do ar, são obtidas ao adaptar as equações apresentadas por Titze [71] para o modelo *body-cover*, detalhadas no Apêndice A. Para a glote aberta, i.e.,  $a_1 > 0$  e  $a_2 > 0$ , as pressões no interior da glote são computadas como

$$P_1 = \begin{cases} P_s - \frac{a_d}{a_1} P_{kd}, & a_2 > a_1 \\ P_s - \left(\frac{a_2}{a_1}\right)^2 P_{kd}, & a_2 \leq a_1 \end{cases}, \quad (6.18)$$

$$P_2 = \begin{cases} P_s - P_{kd}, & a_2 > a_1 \\ P_s - \frac{a_2}{a_1} P_{kd}, & a_2 \leq a_1 \end{cases}, \quad (6.19)$$

em que  $P_s$  é a pressão subglotal,  $a_d$  e  $P_{kd}$  são, respectivamente, a área e a pressão cinética no ponto de separação dos fluxos. Aqui, a área é definida como [25]

$$a_d = \max(0, \min(a_1, a_2)) \quad (6.20)$$

e a pressão cinética é dada por [71]

$$P_{kd} = \frac{P_s - P_e}{1 - k_e} \quad (6.21)$$

em que  $k_e$  é o coeficiente de recuperação de pressão [24], como formulado por Titze [71], e  $P_e$  é a pressão no início da epilaringe (região imediatamente acima da glote, indo até a epiglote), que depende do acoplamento com o trato vocal.

Para o fechamento glotal, as pressões agindo nas pregas são obtidas da por

$$P_1 = \begin{cases} P_s, & a_2 \leq 0 \text{ and } a_1 > 0 \\ P_h, & a_1 \leq 0 \end{cases}, \quad (6.22)$$

$$P_2 = \begin{cases} P_e, & a_1 \leq 0 \text{ and } a_2 > 0 \\ P_h, & a_2 \leq 0 \end{cases}, \quad (6.23)$$

em que  $P_h$  é a pressão hidrostática, definida como a média entre  $P_s$  e  $P_e$ , i.e.

$$P_h = \frac{P_s + P_e}{2}. \quad (6.24)$$

As forças aerodinâmicas que agem nas pregas são dadas por:

$$F_i = lT_iP_i, \quad (6.25)$$

em que  $T_i$  representa a espessura de cada massa da cobertura.

### 6.1.3 Fluxo Glotal e Acoplamento do Trato Vocal

O fluxo glotal é computado de acordo com a equação de Titze [69] (Eq. 3.45), assumindo  $a_m = a_d$ . A turbulência no fluxo de ar na glote é modelado a partir do calculo do número de Reynolds, como proposto por Samlan e Story [129].

O trato vocal e a traqueia são acoplados às pregas vocais usando o modelo K-L [26]. O modelo foi implementado como proposto por Story [90], usando a formulação para um tubo sem perdas e aplicando um fator de atenuação para representar as perdas. As áreas de seção transversal da traqueia e do trato vocal foram obtidas e discretizadas por Story [91] e estão disponíveis com o *software open source LeTalker 1.22* [130]. Para o estudo deste capítulo, foram utilizados apenas áreas de tratos vocais das vogais /a/, /i/ e /u/.

### 6.1.4 Regras Musculares e Demais Parâmetros do Modelo

As regras propostas por Titze e Story [28], que se encontram destrinchadas em 3.1.3.1 no Capítulo 3, para relacionar atividades musculares com parametros do modelo foram utilizadas. Para a geração das amostras, foram utilizados  $a_{ct}$  e  $a_{ta}$  variando de 0,05 a 0,85 com passos de 0,1 e  $a_{lc}$  variando de 0,425 a 0,575 com passos de 0,25. Os coeficientes não-lineares assumem os valores propostos por Story e Titze [25] e por Ishizaka e Flanagan[24],  $k_p$  é mantido fixo em  $140 \text{ Kg/s}^2$  e os outros parâmetros relacionados ao pólipo são definidos como proposto

por Zhang e Jiang [17]. O modelo resultante foi utilizado para gerar amostras normais (ao remover a massa do pólipo) e patológicas. As amostras cuja simulação não produziu oscilação sustentada foram descartadas.

## 6.2 Detecção de Pólipo Vocal a partir da Dimensão Intrínseca e Entropia Diferencial

Para apresentar uma ilustração prática da utilidade do modelo, foram simuladas mais de 800 amostras de vozes masculinas e femininas<sup>2</sup>, com e sem a presença de pólipo, com base no intervalo definido para as regras musculares [28].

Essas amostras foram cuidadosamente estudadas, diante de diferentes perspectivas de suas características temporais e espectrais. Dessa maneira, foi possível observar que a presença de um pólipo induz flutuações irregulares (aleatórias) de amplitude ao longo do tempo, como ilustrado na Figura 6.3. Esta flutuação sugeriu a possibilidade da detecção de pólipos por meio da entropia e dimensão intrínseca locais dos sinais.

Seja  $s(n)$ ,  $n = 0, 1, 2, \dots, N$ , para  $s \in \mathbb{R}$ ,  $N \in \mathbb{N}$ , a representação de  $N$  amostras de um sinal de som, com taxa de amostragem constante  $f_s$  amostras por segundo. Para uma modelagem adequada dos sinais de fala, assume-se que  $\{s(0), s(1), \dots, s(N)\}$  é uma instância de um processo estocástico, no qual se espera que a presença de um pólipo perturbe o modelo, como pode, de fato, ser notada claramente, através da ilustração na Figura 6.3, a perturbação aleatória existente na amplitude do sinal após o *onset*.

Para os fins propostos neste capítulo, assume-se que essa perturbação ao modelo imposta pelo pólipo é suficiente para gerar uma mudança detectável na amostra de voz, mesmo quando o pólipo é apenas uma pequena perturbação na estrutura das pregas vocais. Essa suposição é inspirada na habilidade auditiva, que alguns especialistas treinados possuem, de

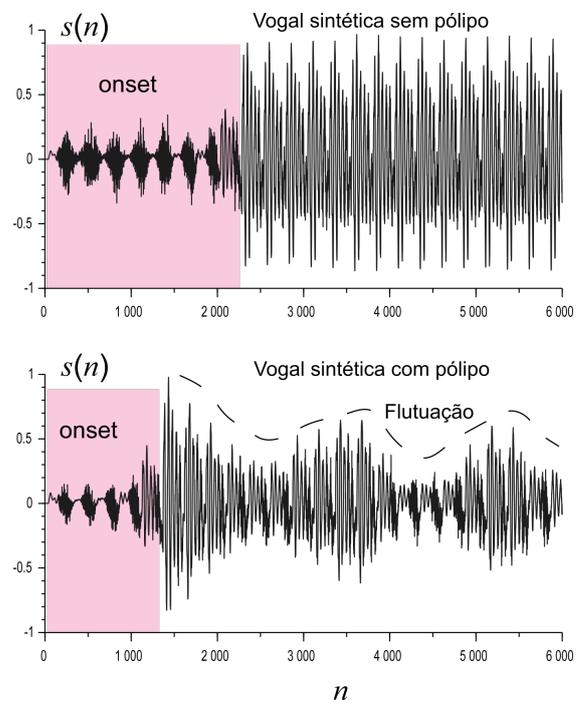


Figura 6.3 – Sinais representando instâncias de  $s(n)$  das amostras com (acima) e sem (abaixo) pólipo. O intervalo de *onset* da fala e amplitude causada pelo pólipo são indicados.

<sup>2</sup>As amostras encontram-se disponíveis em [biochaves.com/polyp-samples/](http://biochaves.com/polyp-samples/)

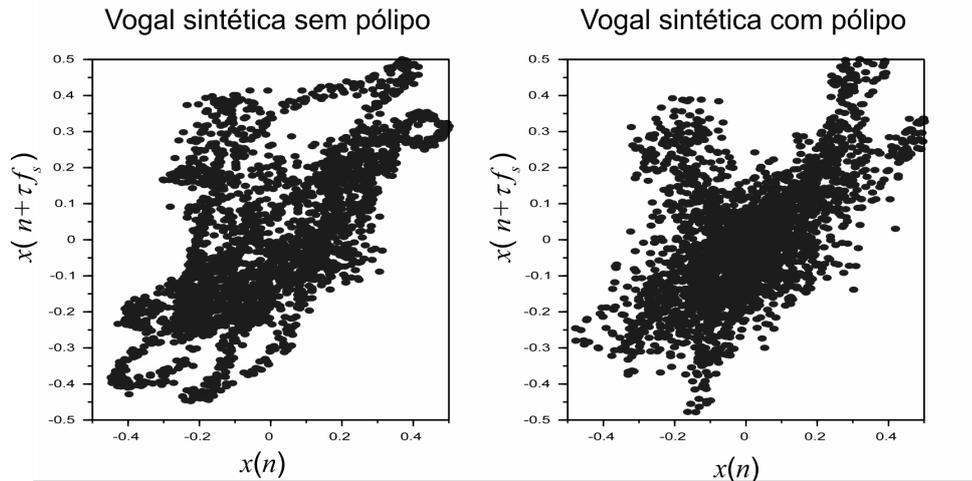


Figura 6.4 – Os pontos representam o par de coordenadas  $(s(n), s(n + \tau f_s))$  para as vogais sintéticas sem (esquerda) e com (direita) a presença de pólipos.

detectar alterações vocais, como as induzidas por pólipos, em pacientes através da percepção auditiva da fala.

Utilizando os sinais simulados, com base no modelo apresentado na Seção 6.1, para realizar experimentos como os ilustrados na Figura 6.3, foi possível destacar uma diferença estatística entre vozes com e sem a presença de pólipos, que é uma consequência da flutuação aleatória ilustrada. Mais especificamente, sabe-se que um sinal perfeitamente periódico produz uma trajetória no espaço de fase equivalente a um loop fechado unidimensional, ao passo que se espera que ruídos aleatórios deformem tal trajetória. Utilizando sinais com e sem pólipos, a Figura 6.4 ilustra projeções 2D dessas trajetórias. Para uma boa ilustração, o atraso de tempo  $\tau$  é definido com 0.001 segundos para ambos os sinais. Para uma melhor comparação, ambos os conjuntos de pontos foram limitados ao espaço de um quadrado unitário. Isto equivale à normalização de todas amostras do sinal, limitando as ao intervalo  $[-0.5, 0.5]$ . Tal normalização também foi aplicada a todos os sinais analisados neste capítulo.

Mesmo mediante a perspectiva limitada possibilitada por essas projeções 2D (Figura 6.4), fica claro que os pontos das trajetórias associadas à presença de pólipos nas pregas vocais são menos afastados uns dos outros do que as trajetórias das vozes sem pólipos. Assim, duas conjecturas formam a base do método proposto para a detecção de pólipos a partir do sinal de voz, como seguem:

1. Após o *onset*, as vogais sustentadas sem pólipos são quasi-periódicas, logo estas formam trajetórias localmente 1D no espaço de fases, i.e., elas produzem *loops* fechados;
2. As vogais sustentadas cuja dinâmica é perturbada pela presença de um pólipo, após o *onset*, produzem trajetórias menos bem definidas no espaço de fase, possivelmente colapsando em si mesmas.

Uma consequência relevante do item 2 é o fato de que não é mais esperado que hajam trajetórias unidimensionais. Para o aproveitamento disto, foi adaptado o método proposto por Montalvão [22], descrito no Capítulo 5, para a estimação conjunta da dimensão intrínseca local e da entropia diferencial de uma instância de um sinal estocástico, que nesse caso é um segmento do sinal de voz. Nessa adaptação, foram obtidos padrões 6D (vetores) de amostras consecutivas do sinal, como

$$\mathbf{p}(n) = [s(n), s(n+1), \dots, s(n+5)], \quad n = 0, 1, 2, \dots, N-5.$$

Somente os padrões contendo o trecho do sinal após o *onset* foram utilizados para a análise.

Para um conjunto de  $P$  padrões, em que  $P$  corresponde a cerca de 50ms do sinal (por exemplo,  $P = 2500$  padrões para  $f_s = 44100$  amostras por segundo), de uma dada amostra de voz, foi realizada a análise conjunta descrita por Montalvão [22]. Dessa análise, duas medidas estatísticas em escala logarítmica foram obtidas através da contagem direta de coincidências. As duas medidas em escala logarítmica são  $\log_2(r)$  e  $\log_2(C(r))$ , em que  $r$  corresponde ao comprimento da aresta de um cubo (ou hipercubo) dentro do qual dois padrões são ditos coincidentes, e  $C(r)$  corresponde à taxa de coincidência.

Para os fins esperados nesse experimento, dado que  $u(r) = \log_2(r)$  e  $v(r) = \log_2(C(r))$  são duas funções do comprimento de aresta  $r$ , logo a dimensão intrínseca local pode ser estimada como

$$\hat{d}(r) = \frac{dv(r)}{du(r)}.$$

Consequentemente, a entropia diferencial, pode ser estimada como

$$\hat{h}(r) = \hat{d}(r) - v(r).$$

As duas estimativas encontram-se ilustradas na Figura 6.5, para dois sinais sintéticos.

Através do estudo de curvas como as representadas na Figura 6.5, para as amostras de vogais sintéticas, foi observado que a escala correspondente a  $\log_2(r) \approx -3$ , logo para  $r \approx 2^{-3}$ , uma diferença em termos de dimensão intrínseca e entropia diferencial se destaca consistentemente entre sinais com e sem pólipos nas pregas vocais. Como descrito anteriormente, essa diferença era esperada analiticamente, e ela pode ser atribuída à flutuação aleatória na amplitude. Isto induz o aumento da dimensão intrínseca local (em pequenas escalas de análise  $r$ ), e força a diminuição da entropia diferencial (volume efetivo [22]). Ambos os efeitos são ilustrados na Figura 6.5.

Para testar essa abordagem, como uma espécie de prova de conceito, foram utilizadas amostras de vogais humanas sadias (sem nenhum tipo de alteração vocal) e com a presença de pólipos, obtidos na base de dados aberta *Saarbruecken Voice Database* [131]. Como ilustrado na Figura 6.6, cada ponto corresponde a uma amostra de voz, sendo suas coordenadas dadas pelas estimativas da entropia diferencial e da dimensão intrínseca, utilizando uma escala de análise de  $r = 2^{-3}$ . Foram utilizadas apenas amostras da vogal /a/ sustentada em tom natural.

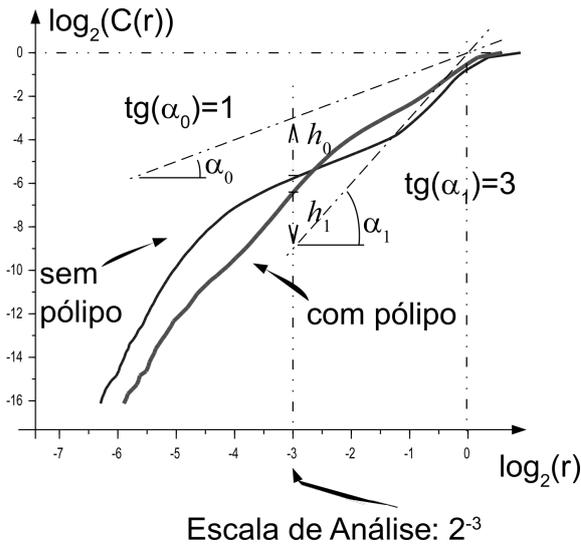


Figura 6.5 – Duas curvas a partir das quais a estimação conjunta da dimensão intrínseca (curves slopes) e da entropia diferencial (distâncias from corresponding constant slope lines) pode ser obtida para um intervalo de escalas de análise. Para a escala de análise de  $2^{-3}$ , a dimensão intrínseca e a entropia diferencial para sinais com e sem pólipo são estimadas como as tangentes dos ângulos  $\alpha_0$  e  $\alpha_1$ , e as distâncias  $h_0$  e  $h_1$ , respectivamente.

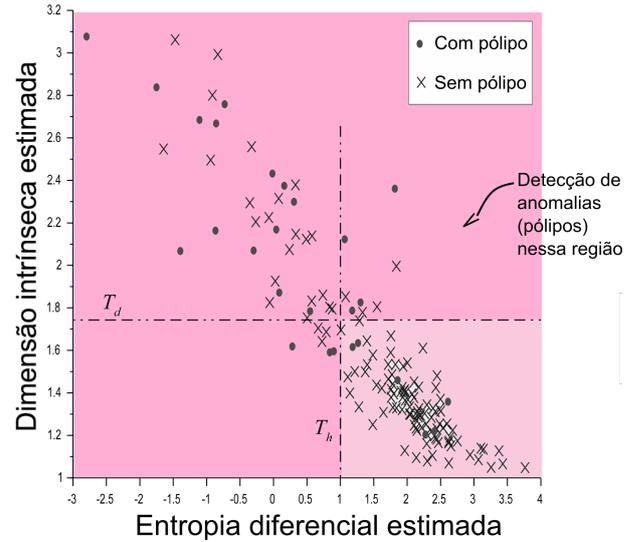


Figura 6.6 – Ilustração dos pontos correspondentes aos sinais reais de voz de indivíduos com (bolinhas) e sem (cruzeiros) a presença de pólipos nas pregas vocais. As coordenadas de cada ponto são dadas pela entropia diferencial e pela dimensão intrínseca estimadas para uma escala de análise de  $2^{-3}$ . Dois limiares para a entropia e dimensão intrínseca,  $T_h = 1$  e  $T_d = 1.75$  respectivamente, são indicados, gerando uma regra simples de detecção de anomalias (pólipos).

Para a remoção do *onset*, todas as amostras foram analisadas apenas a partir da amostra  $s(3000)$ .

Por fim, um classificador simples foi implementado, apenas para verificar o conceito analisado, de acordo com as seguintes regras:

- Se a entropia diferencial de um dado sinal for maior que um limiar  $T_h$  e a dimensão intrínseca for menor que um limiar  $T_d$ , então o sinal é classificado como normal;
- Caso contrário, o sinal é classificado como anômalo (com pólipo, neste caso).

Na Figura 6.6 é ilustrado o resultado do classificador para  $T_h = 1$  e  $T_d = 1.75$ . Neste caso, cerca de 22% das amostras de voz com pólipos resultaram em falso negativo (classificadas como normais), enquanto cerca de 23% dos sinais normais resultaram em falso positivo (classificadas como anômalas). Tais resultados apontam para uma baixa probabilidade de

que esse classificador tenha obtido bons resultado por acaso, o que sugere que essa pode ser uma ferramenta válida e útil para a detecção de pólipos.

## CAPÍTULO

## 7

## Sintetizador de Voz Controlado pela Ativação Muscular das Pregas Vocais

Visando gerar um ferramenta útil, com base nos modelos apresentados para as pregas vocais e para o trato vocal, foi desenvolvido um sintetizador de voz que permite gerar amostras com bases nos parâmetros de controle muscular propostos por Titze e Story [28]. O *software open source LeTalker 1.22* [130] também é desenvolvido com base nos mesmos modelos e regras. Entretanto, o *software* implementado neste trabalho disponibiliza novas opções, como escolha de gênero (alterado pelo comprimento das pregas vocais), escolha de tipo de transições nos parâmetros, possibilidade de um trato vocal uniforme e a implementação das regras para vozes alteradas de Lucero [13, 15]<sup>1</sup>. Além disso, o *software* implementado possui um tempo de execução, para sintetizar uma amostra, muito inferior ao do *LeTalker 1.22*.

A interface do *software*, provisoriamente nomeado como *Glottal Generator*, encontra-se ilustrada na Figura 7.1. Os códigos do sintetizador e dos comandos entre internos foram escritos em *Python 3.8* e a os elementos gráficos da interface foram gerados com o *Qt Designer*

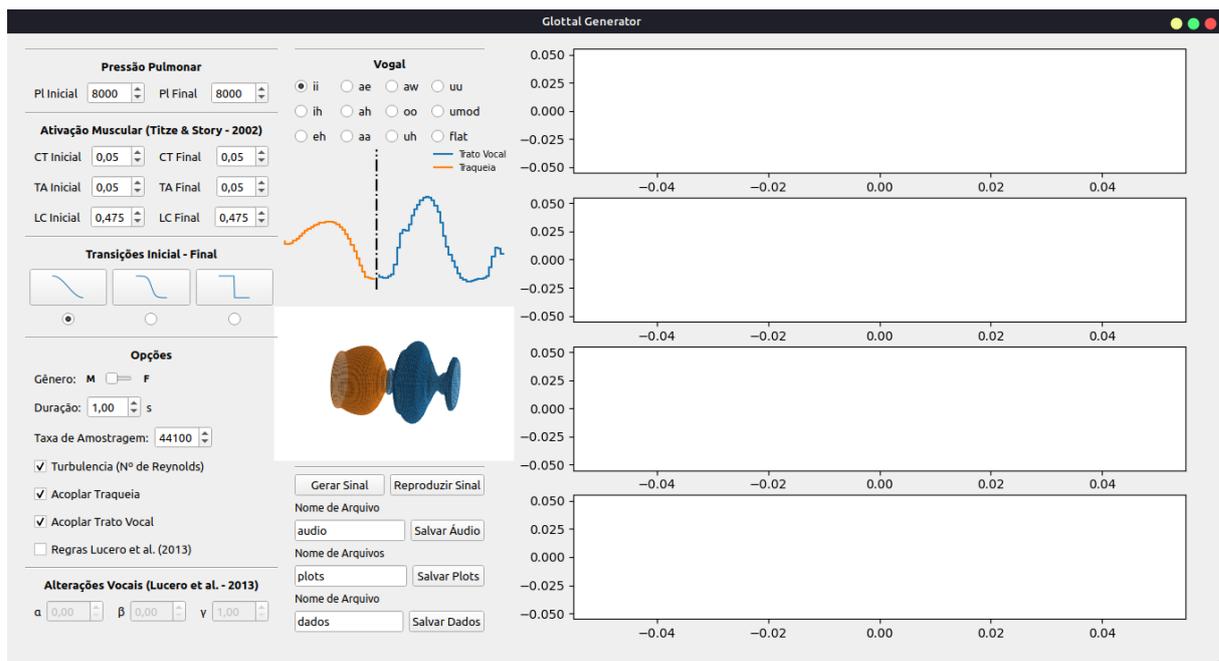


Figura 7.1 – Interface do *Glottal Generator*.

<sup>1</sup>Aqui o parâmetro multiplicador das elasticidades da mola, para simular a tensão (S), é denominado  $\gamma$ .

e a biblioteca *PyQt5* para conectar o código em *Python* com a interface. Para uma execução mais rápida do código, foi utilizada a biblioteca *Numba*, que realiza compilação *Just-in-time* das funções especificadas. Além disso, foram utilizadas as bibliotecas *numpy*, *sys*, *cv2*, *time*, *struct*, *matplotlib*, *wave*, *csv* e *sounddevice*.

A Figura 7.2 destaca os controles para os níveis de pressão pulmonar e para os níveis de ativação muscular. A esses parâmetros é permitida a variação ao longo do tempo, de forma que, quando os valores inicial e final de um desses parâmetros não são iguais, um vetor ao longo do tempo é gerado de acordo com a função de transição selecionada. Da esquerda para a direita as funções de transição são: (1) uma meia onda cossenoidal ( $f_{\cos}(n)$ ); (2) uma tangente hiperbólica ( $f_{\tanh}(n)$ ); (3) uma função degrau que altera seu valor abruptamente no meio do intervalo de tempo total do sinal ( $f_{\text{sign}}(n)$ ). Essas funções são calculadas pelas expressões

$$f_{\cos}(n) = \frac{v_i - v_f}{2} \cos\left(2\pi \frac{n}{N}\right) + \frac{v_i + v_f}{2}, \quad (7.1)$$

$$f_{\tanh}(n) = \frac{v_i - v_f}{2} \tanh\left(8 \frac{n - \frac{N}{2}}{N}\right) + \frac{v_i + v_f}{2}, \quad (7.2)$$

$$f_{\text{sign}}(n) = \frac{v_i - v_f}{2} \text{sign}\left(\frac{n - \frac{N}{2}}{N}\right) + \frac{v_i + v_f}{2}. \quad (7.3)$$

em que  $v_i$  e  $v_f$  são os valores inicial e final, respectivamente, do parâmetro para o qual o vetor é calculado,  $n$  é o tempo discreto e  $N$  é o total de amostras do sinal.

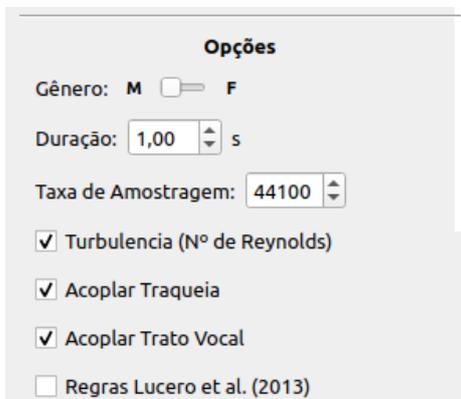


Figura 7.3 – Opções de Síntese.

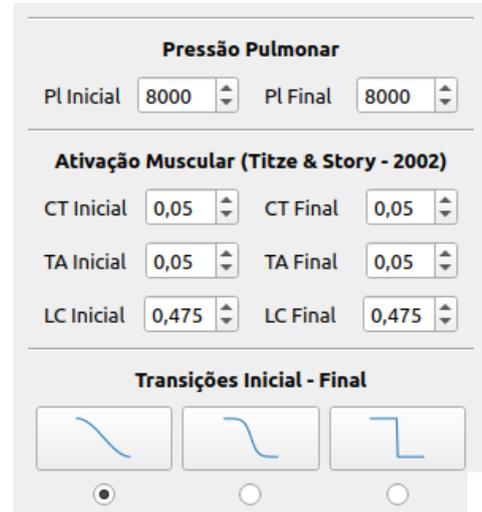


Figura 7.2 – Controles da pressão pulmonar e dos parâmetros musculares, que podem variar ao longo do tempo, de acordo com a transição escolhida.

Na Figura 7.3 são focadas as opções disponibilizadas ao usuário na interface do *Glottal Generator*. É permitido ao usuário escolher o gênero (essa opção altera  $L_0$  do modelo das pregas vocais entre 1.6cm e 1.0cm; **M** e **F**, respectivamente), a duração das amostras, o número de amostras por segundo, se a traqueia e o trato vocal serão acoplados e se as regras de Lucero [13, 15] serão aplicadas.

Na Figura 7.4 são ressaltadas as entradas para aplicar as regras de Lucero [13, 15] para a síntese de vozes alteradas. Para que esses campos possam ser editados, é necessário ativar as regras nas opções (Figura 7.3). Como explicado no Capítulo 4 e no início deste capítulo,  $\alpha$

Para que esses campos possam ser editados, é necessário ativar as regras nas opções (Figura 7.3). Como explicado no Capítulo 4 e no início deste capítulo,  $\alpha$

produz vozes rugosas,  $\beta$  vozes soprosas e  $\gamma$  vozes tensas, de acordo com os valores aplicados.

A Figura 7.5 realça o menu de seleção do trato vocal, através das vogais desejadas. Abaixo da seleção de vogais encontram-se dois gráficos. O primeiro ilustra a função de áreas transversais discretizadas do trato vocal e da traqueia, ao passo que o segundo representa o contorno tridimensional discretizados dos tratos e da traqueia, estimado de acordo com os valores das áreas. Os gráficos se atualizam imediatamente após à seleção de uma nova vogal.

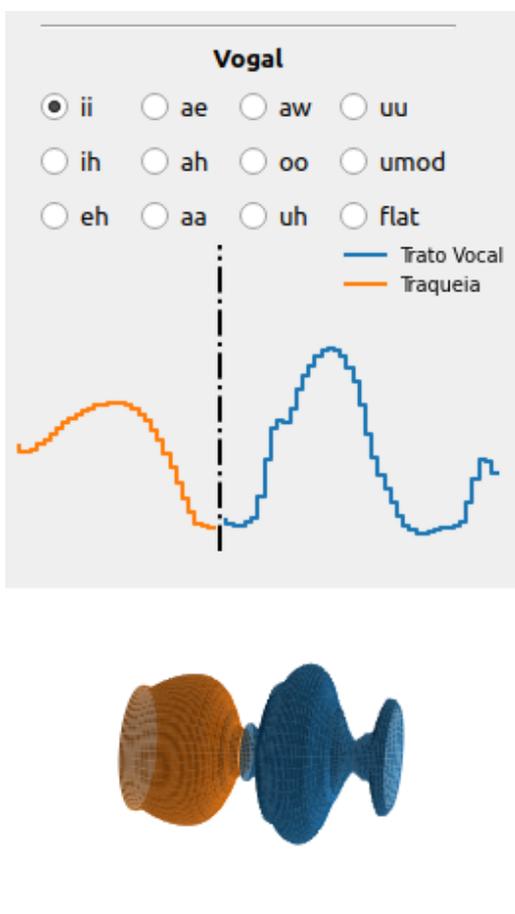


Figura 7.5 – Seleção do trato vocal, de acordo com a vogal desejada.

voz é salva em um arquivo '.wav' na pasta em que a interface se encontra, com o nome de arquivo atribuído pelo usuário. Quando o botão "Salvar Plots" é selecionado, duas figuras em formato '.png' são geradas e arquivadas, adicionando um sufixo ao nome atribuído pelo usuário. Uma figura gerada é a ilustração das áreas e dos tubos do trato vocal, que é salva com o sufixo "\_areas" e a outra figura possui em si os gráficos gerados, sendo salva com o sufixo "\_curvas". Dessa maneira, para o nome de arquivo "plots", por exemplo, os arquivos resultan-

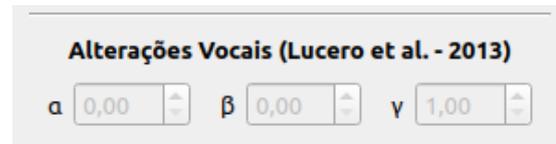


Figura 7.4 – Controle dos valores para a regra da síntese de vozes alteradas.

Os valores das áreas dos tratos vocais foram obtidos e discretizados por Story et al. [91] e disponibilizados com o *LeTalker 1.22* [130]. Adicionalmente, foi inserida a possibilidade, no *Glottal Generator*, de sintetizar amostras utilizando um tubo uniforme no trato vocal (representado como *flat* na interface). As áreas de cada uma das áreas disponíveis no simulador encontram-se representadas na Figura 7.6.

Por fim, na Figura 7.7 são ilustrados os botões presentes na interface para sintetizar o sinal, para reproduzir o áudio e para salvar os arquivos. Ao selecionar o botão "Gerar Sinal" o modelo é simulado, utilizando o método de Runge-Kutta de 4ª ordem, e os gráficos são plotados na área disponível para eles, como ilustrado na Figura 7.1. Os gráficos, de cima para baixo, são referentes ao sinal de voz, ao fluxo glotal, à área glotal e aos deslocamentos das massa. Selecionando o botão "Reproduzir Sinal", o sinal de voz é reproduzido e, ao selecionar o botão "Salvar Áudio", essa amostra de

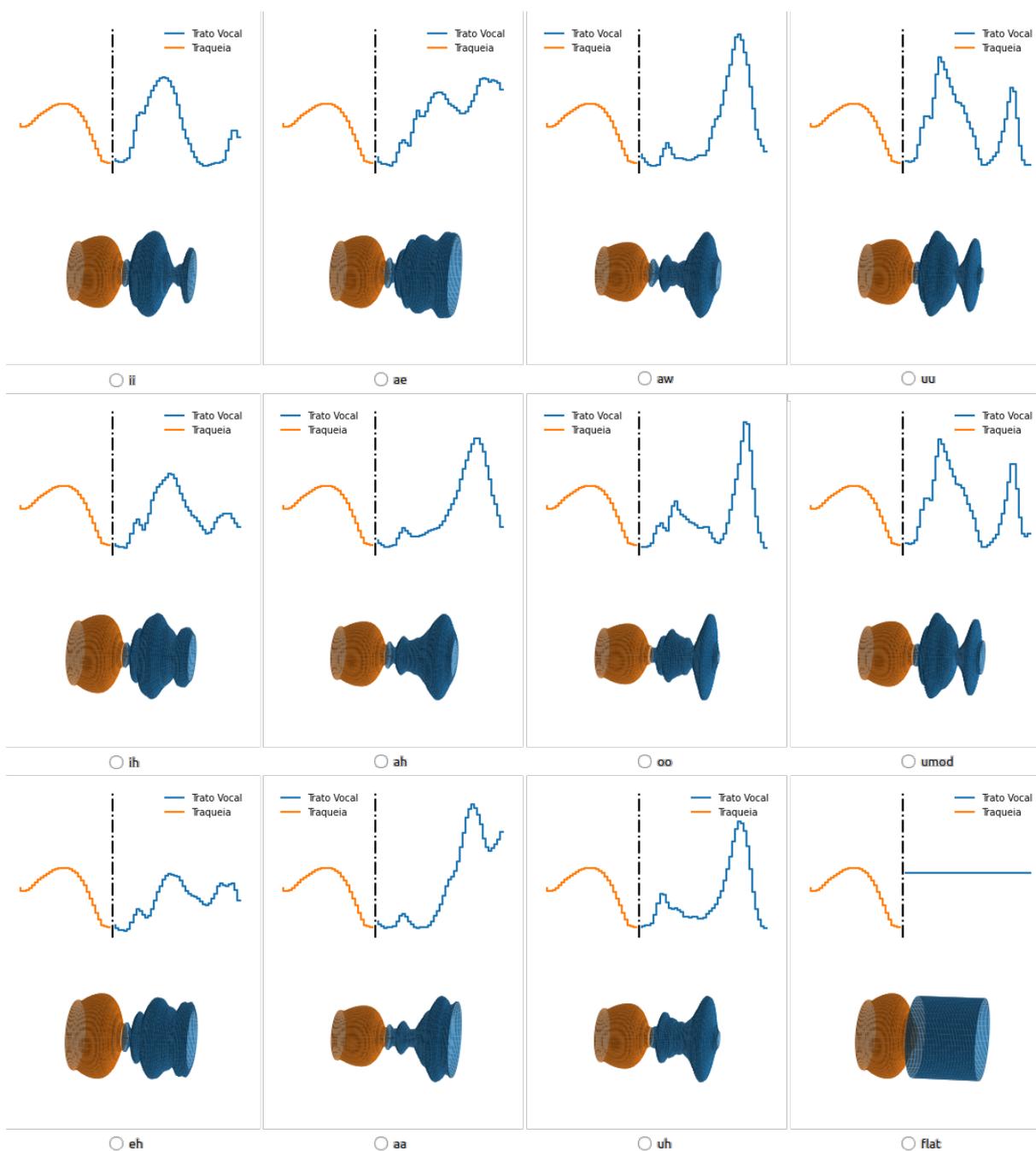


Figura 7.6 – Todas as áreas disponíveis para a síntese de amostras.

tes seriam "plots\_areas.png" e "plots\_curvas.png". A Figura 7.8 contém um exemplo de plot salvo das curvas geradas na interface. Ao selecionar salvar dados, os dados computados pelo modelo são disponibilizados em um arquivo '.csv', salvo com o nome atribuído pelo usuário. São disponibilizados no arquivo os sinais da voz, do fluxo glotal, do ruído da turbulência, da área glotal, dos deslocamentos das massas, da pressão subglotal, da pressão supraglotal e do tempo.

O uso de ferramentas como sintetizador implementado é útil para poder estudar a dinâmica vocal, podendo analisar o comportamento das pregas vocais durante a modificação

de parâmetros, assim como permite produzir amostras de voz alteradas, tanto pelas regras de Lucero [13, 15] como ao utilizar valores das regras musculares que forcem condições atípicas na produção vocal (como uma adução exacerbada, obtida com baixos valores de  $a_{lc}$ , que tende a gerar amostras mais soprosas, por exemplo). Essa ferramenta também apresenta o potencial de ser aplicável à avaliação da qualidade vocal, permitindo ao avaliador gerar amostras similares à voz humana que se deseja avaliar e poder obter informações relevantes dos dados usados para produzir o sinal sintético, de maneira similar à aplicação de Gerratt e Kreiman [10] com um simulador acústico.



Figura 7.7 – Botões utilizados para gerar o sinal, reproduzir seu áudio e salvar os arquivos disponíveis.

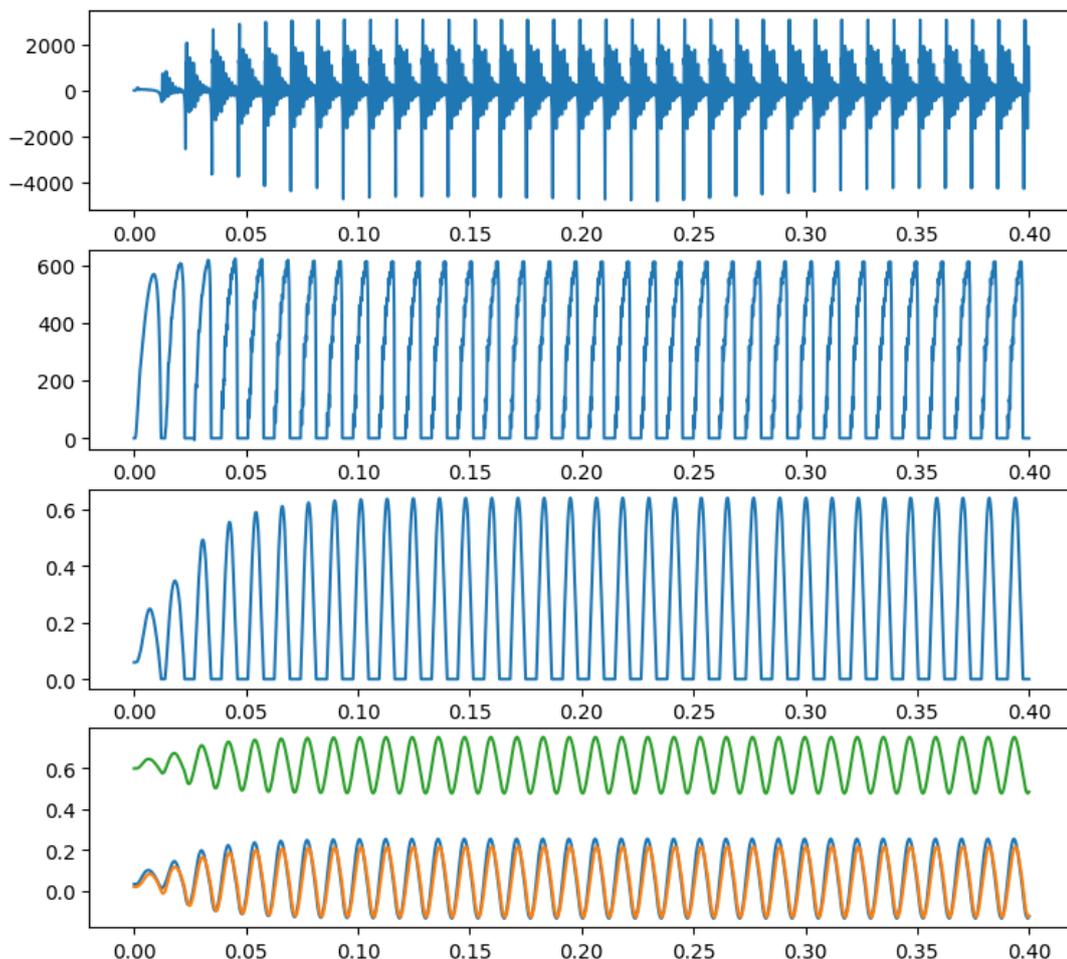


Figura 7.8 – Exemplo de sinal gerado pelo simulador. De cima para baixo: voz; fluxo glotal; área glotal; deslocamento das massas.

## CAPÍTULO

## 8

## Conclusão e Trabalhos Futuros

Neste trabalho foi conduzido um estudo sobre os mecanismos de produção da voz humana [7, 30, 32] e de como ocorre o processo de fonação [5, 6, 41]. Com base nisso, são apresentados diversos modelos matemático-computacionais das pregas vocais, estabelecidos na literatura, como ferramentas úteis para o estudo das pregas vocais mediante condições normais e alteradas.

Com base nos modelos estudados, foi proposto um modelo adaptado para as pregas vocais com a presença de um pólipos [18]. Através de vogais sintéticas geradas por esse modelo foi possível observar características do comportamento das pregas vocais com pólipos e levantar hipóteses sobre o uso de estimadores de dimensão intrínseca e entropia diferencial [22] na classificação de sinais de voz com pólipos. Um classificador simples forneceu resultados que corroboram a hipótese proposta, sugerindo que medidas de dimensão intrínseca e de entropia diferencial podem ser ferramentas úteis na detecção da presença de pólipos em sinais de voz.

Utilizando um modelo bem estabelecido na literatura [71, 28], que permite controlar os parâmetros que regem sua oscilação de acordo com a atividade de músculos intrínsecos da laringe, foi implementado um gerador computacional de vozes sintéticas. Regras para simular condições alteradas da voz humana foram adicionados a esse sintetizador [13, 15], fazendo dele uma ferramenta com potencial de aplicação na avaliação da qualidade vocal, como em métodos de análise-pela síntese [10, 108]. O sintetizador é potencialmente útil, também, para o estudo da dinâmica vocal.

Para futuros trabalhos, é necessária ainda uma averiguação mais detalhada do uso de entropia diferencial e dimensão intrínseca na detecção de pólipos, bem como aplicando técnicas de visualização de dados e projeção em dimensões mais baixas [132], sendo útil a condução de novas análises. Também pode ser estendido o estudo aqui conduzido a outros desvios de fonação, adaptando modelos para outros tipos e patologias na voz, como doença de Parkinson [110], tremores vocais [111] e paralisia unilateral [112], dentre outros.

Futuros avanços no sintetizador também podem ser conduzidos, acrescentando funcionalidades que forneçam mais flexibilidade e potencial de análise para os usuários.

Como exemplo, pode-se acrescentar formas de ajuste mais flexível do trato vocal, ao invés de formatos predefinidos, utilizando técnicas para parametrizar as áreas do trato vocal [133, 134]. Também é possível aperfeiçoar as curvas de ativação muscular, para serem mais adequadas à atividade real dos músculos [135] e permitir a simulação realista de patologias que afetem esses músculos, dentre outras aplicações. É necessário, entretanto, que o sintetizador seja testado e validado para a produção de amostras confiáveis na representação da voz humana com e sem alterações vocais.

## Referências

- 1 REDFORD, M. A. **The handbook of speech production**. Chichester: John Wiley & Sons, 2015. Citado na página 1.
- 2 CATALDO, E.; SAMPAIO, R.; NICOLATO, L. Uma discussão sobre modelos mecânicos de laringe para síntese de vogais. **Engevista**, v. 6, n. 1, 2004. Citado nas páginas 1, 24 e 42.
- 3 CHEN, G.; SHUE, Y.-L.; KREIMAN, J.; ALWAN, A. Estimating the voice source in noise. In: **Thirteenth Annual Conference of the International Speech Communication Association, INTERSPEECH**. Portland, OR, USA: International Speech Communication Association - ISCA, 2012. Citado na página 1.
- 4 ISSHIKI, N. **Phonosurgery: theory and practice**. Tokyo: Springer Japan, 1989. Citado na página 1.
- 5 VAN DEN BERG, J. Myoelastic-aerodynamic theory of voice production. **Journal of speech and hearing research**, ASHA, v. 1, n. 3, p. 227–244, 1958. Citado nas páginas 1, 3, 19, 20, 21 e 72.
- 6 TITZE, I. R. Comments on the myoelastic-aerodynamic theory of phonation. **Journal of Speech, Language, and Hearing Research**, ASHA, v. 23, n. 3, p. 495–510, 1980. Citado nas páginas 1, 19, 21 e 72.
- 7 FERRAND, C. T. **Speech science: an integrated approach to theory and clinical practice**. Needham, MA: Allyn & Bacon, 2001. Citado nas páginas 1, 4, 6, 7, 9, 12, 13, 16, 18, 19, 46, 47, 48 e 72.
- 8 ERATH, B. D.; ZAÑARTU, M.; STEWART, K. C.; PLESNIAK, M. W.; SOMMER, D. E.; PETERSON, S. D. A review of lumped-element models of voiced speech. **Speech Communication**, Elsevier, v. 55, n. 5, p. 667–690, 2013. Citado nas páginas 1, 20 e 24.
- 9 COLTON, R.; CASPER, J. **Understanding voice production: A physiological perspective for diagnosis and treatment**. 2. ed. Baltimore: Williams & Wilkins, 1996. Citado nas páginas 1 e 46.
- 10 GERRATT, B. R.; KREIMAN, J. Measuring vocal quality with speech synthesis. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 110, n. 5, p. 2560–2566, 2001. Citado nas páginas 1, 48, 49, 71 e 72.
- 11 KREIMAN, J.; GERRATT, B. R. Perceptual assessment of voice quality: past, present, and future. **Perspectives on Voice and Voice Disorders**, ASHA, v. 20, n. 2, p. 62–67, 2010. Citado nas páginas 1 e 48.

- 12 SAMLAN, R. A.; STORY, B. H.; BUNTON, K. Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. **Journal of Speech, Language, and Hearing Research**, ASHA, 2013. Citado na página 1.
- 13 LUCERO, J. C.; SCHOENTGEN, J.; BEHLAU, M. Physics-based synthesis of disordered voices. In: **Interspeech**. Lyon, France: International Speech Communication Association - ISCA, 2013. p. 587–591. Citado nas páginas 1, 52, 67, 68, 71 e 72.
- 14 KREIMAN, J.; GERRATT, B. R.; GARELLEK, M.; SAMLAN, R.; ZHANG, Z. Toward a unified theory of voice production and perception. **Loquens**, NIH Public Access, v. 1, n. 1, 2014. Citado na página 1.
- 15 ENGLERT, M.; MADAZIO, G.; GIELOW, I.; LUCERO, J.; BEHLAU, M. Perceptual error analysis of human and synthesized voices. **Journal of Voice**, Elsevier, v. 31, n. 4, p. 516–e5, 2017. Citado nas páginas 1, 52, 67, 68, 71 e 72.
- 16 ROSA, M. d. O. **Laringe digital**. Tese (Doutorado) — Universidade de São Paulo, 2002. Citado nas páginas 2, 24, 41 e 50.
- 17 ZHANG, Y.; JIANG, J. J. Chaotic vibrations of a vocal fold model with a unilateral polyp. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 115, n. 3, p. 1266–1269, 2004. Citado nas páginas 2, 29, 30, 41, 50, 52, 58 e 62.
- 18 SANTOS, J.; MONTALVÃO, J.; SANTOS, I. Improved model for vocal folds with a polyp with potential application. In: **Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH**. Shanghai, China: International Speech Communication Association - ISCA, 2020. p. 1386–1390. Citado nas páginas 2, 40, 41, 52, 54, 58 e 72.
- 19 HIRANO, M.; GOULD, W. J.; LAMBIASE, A.; KAKITA, Y. Vibratory behavior of the vocal folds in a case with a unilateral polyp. **Folia Phoniatica et Logopaedica**, Karger Publishers, v. 33, n. 5, p. 275–284, 1981. Citado nas páginas 2 e 58.
- 20 VASCONCELOS, D. de; GOMES, A. d. O. C.; ARAÚJO, C. M. T. de. Vocal fold polyps: literature review. **International archives of otorhinolaryngology**, Thieme Revinter Publicações Ltda, v. 23, n. 01, p. 116–124, 2019. Citado nas páginas 2 e 58.
- 21 BEHRMAN, A. Global and local dimensions of vocal dynamics. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 105, n. 1, p. 432–443, 1999. Citado na página 2.
- 22 MONTALVÃO, J. Noise variance estimation through joint analysis of intrinsic dimension and differential entropy. **IEEE Signal Processing Letters**, IEEE, v. 26, n. 9, p. 1330–1333, 2019. Citado nas páginas 2, 3, 54, 55, 56, 57, 64 e 72.
- 23 FLANAGAN, J.; LANDGRAF, L. Self-oscillating source for vocal-tract synthesizers. **IEEE Transactions on Audio and Electroacoustics**, IEEE, v. 16, n. 1, p. 57–64, 1968. Citado nas páginas 3, 24, 25, 26, 27, 35 e 42.
- 24 ISHIZAKA, K.; FLANAGAN, J. L. Synthesis of voiced sounds from a two-mass model of the vocal cords. **Bell system technical journal**, Wiley Online Library, v. 51, n. 6, p. 1233–1268, 1972. Citado nas páginas 3, 26, 27, 29, 35, 42, 43, 50, 51, 61 e 86.

- 25 STORY, B. H.; TITZE, I. R. Voice simulation with a body-cover model of the vocal folds. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 97, n. 2, p. 1249–1260, 1995. Citado nas páginas 3, 30, 32, 35, 58, 60 e 61.
- 26 KELLY, J.; LOCHBAUM, C. Speech synthesis. In: **Proceedings of the Fourth International Congress on Acoustics**. Copenhagen: [s.n.], 1962. Paper G42: p. 1–4. Citado nas páginas 3, 35, 41, 43, 44, 51 e 61.
- 27 FLANAGAN, J. L. **Speech analysis synthesis and perception**. New York: Academy Press, 1965. Citado nas páginas 3, 25, 29, 35, 41 e 42.
- 28 TITZE, I. R.; STORY, B. H. Rules for controlling low-dimensional vocal fold models with muscle activation. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 112, n. 3, p. 1064–1076, 2002. Citado nas páginas 3, 35, 36, 37, 39, 40, 41, 61, 62, 67 e 72.
- 29 HOIT, J. D. Influence of body position on breathing and its implications for the evaluation and treatment of speech and voice disorders. **Journal of Voice**, Elsevier, v. 9, n. 4, p. 341–347, 1995. Citado na página 4.
- 30 BEHLAU, M. **VOZ: O livro do especialista**. Rio de Janeiro - RJ: Livraria e Editora Revinter Ltda., 2001. I. 348 p. ISBN 85-7309-525-3. Citado nas páginas 5, 7, 11, 12, 14 e 72.
- 31 BEHLAU, M.; AZEVEDO, R.; MADAZIO, G. Anatomia da laringe e fisiologia da produção vocal. In: BEHLAU, M. (Ed.). **VOZ: O livro do especialista**. Rio de Janeiro - RJ: Livraria e Editora Revinter Ltda., 2001. I, cap. 1, p. 1–52. Citado nas páginas 4, 5, 6, 7, 8, 12 e 13.
- 32 RAPHAEL, L. J.; BORDEN, G. J.; HARRIS, K. S. **Speech Science Primer: Physiology, acoustics and perception of speech**. Baltimore-Philadelphia: Lippincott Williams & Wilkins, 2007. Citado nas páginas 5, 6, 15 e 72.
- 33 PINHO, S. M. R.; PONTES, P. **Músculos intrínsecos da laringe e dinâmica vocal**. Rio de Janeiro - RJ: Revinter, Rio de Janeiro, 2008. Citado nas páginas 8, 13 e 19.
- 34 PEREIRA, J. **Laryngeal source excitation modelling**. Tese (Doutorado) — University of Southampton, 1987. Citado nas páginas 9 e 29.
- 35 HIRANO, M. Morphological structure of the vocal cord as a vibrator and its variations. **Folia Phoniatria et Logopaedica**, Karger Publishers, v. 26, n. 2, p. 89–94, 1974. Citado nas páginas 10, 11, 30 e 35.
- 36 GRAY, S. Basement membrane zone injury in vocal nodules. In: GAUFFIN, J.; HAMMARBERG, B. (Ed.). **Vocal fold physiology: Acoustic, perceptual, and physiological aspects of voice mechanisms**. San Diego, California: Singular Publishing Group, 1991. p. 21–27. Citado na página 11.
- 37 GRAY, S.; HIRANO, M.; SATO, K. Molecular and cellular structure of vocal fold tissue. In: TITZE, I. R. (Ed.). **Vocal fold physiology: frontiers of basic science**. San Diego, Calif: Singular Publishing Group, 1993. p. 1–34. Citado na página 11.
- 38 KOYAMA, T.; HARVEY, J. E.; OGURA, J. H. Mechanics of voice production. ii. regulation of pitch. **The Laryngoscope**, Wiley Online Library, v. 81, n. 1, p. 47–65, 1971. Citado na página 13.

- 39 RABINER, L. R.; SCHAFER, R. W. **Digital processing of speech signals**. Englewood Cliffs, New Jersey: Prentice-Hall, 1978. 1–512 p. ISBN 0-13-213603-1. Citado nas páginas 18, 41 e 42.
- 40 MÜLLER, J. **The physiology of the senses, voice, and muscular motion, with the mental faculties**. Tradução William Bally. London: Taylor, Walton & Maberly, 1848. Citado na página 19.
- 41 ŠVEC, J. G.; SCHUTTE, H. K.; CHEN, C. J.; TITZE, I. R. Integrative insights into the myoelastic-aerodynamic theory and acoustics of phonation. scientific tribute to donald g. miller. **Journal of Voice**, Elsevier, 2021. Citado nas páginas 19, 22, 23 e 72.
- 42 ELEMANS, C.; RASMUSSEN, J. H.; HERBST, C. T.; DÜRING, D. N.; ZOLLINGER, S. A.; BRUMM, H.; SRIVASTAVA, K.; SVANE, N.; DING, M.; LARSEN, O. N. et al. Universal mechanisms of sound production and control in birds and mammals. **Nature communications**, Nature Publishing Group, v. 6, n. 1, p. 1–13, 2015. Citado na página 19.
- 43 VAN DEN BERG, J.; ZANTEMA, J.; DOORNENBAL JR, P. On the air resistance and the bernoulli effect of the human larynx. **The journal of the acoustical society of America**, Acoustical Society of America, v. 29, n. 5, p. 626–631, 1957. Citado nas páginas 20 e 25.
- 44 KITZING, P.; LÖFQVIST, A. Subglottal and oral air pressures during phonation—preliminary investigation using a miniature transducer system. **Medical and biological engineering**, Springer, v. 13, n. 5, p. 644–648, 1975. Citado na página 21.
- 45 KOIKE, Y. Sub-and supraglottal pressure variation during phonation. In: STEVENS, K. N.; HIRANO, M. (Ed.). **Vocal fold physiology**. Tokyo: University of Tokyo Press, 1981. p. 181–192. Citado na página 21.
- 46 MILLER, D.; SCHUTTE, H. Characteristic patterns of sub-and supraglottal pressure variations within the glottal cycle. In: **Transcr. XIIIth Symp. Care Prof. Voice**. New York, NY: The Voice Foundation, 1984. p. 70–75. Citado na página 21.
- 47 CRANEN, B.; BOVES, L. A set-up for testing the validity of the two mass model of the vocal folds. In: TITZE, I. R.; SCHERER, R. C. (Ed.). **Vocal Fold Physiology: Biomechanics, acoustics and phonatory control**. Denver, CO: The Denver Center for the Performing Arts, Inc., 1983. p. 500–513. Citado na página 21.
- 48 \_\_\_\_\_. Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 77, n. 4, p. 1543–1551, 1985. Citado na página 21.
- 49 SCHUTTE, H.; MILLER, D. Resonanzspieie der gesangsstimme in ihren beziehungen zu supra-und subglottalen druckverlufen: Konscquenzen fr die stimmbildungstheorie. **Folia phoniati**, v. 40, p. 65–73, 1988. Citado na página 21.
- 50 CHEN, C. J.; MILLER, D. A. Pitch-synchronous analysis of human voice. **Journal of voice**, Elsevier, v. 34, n. 4, p. 494–502, 2020. ISSN 0892-1997. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S089219971830448X>>. Citado nas páginas 21 e 22.
- 51 CRANEN, B.; BOVES, L. On subglottal formant analysis. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 81, n. 3, p. 734–746, 1987. Citado na página 22.

- 52 FANT, G. **Acoustic theory of speech production**: with calculations based on x-ray studies of russian articulations. The Hague: Mouton & Co. N.V., 1960. v. 2. (Description and Analysis of Contemporary Standard Russian, v. 2). Citado nas páginas 22 e 49.
- 53 STORY, B. H. Mechanisms of voice production. In: REDFORD, M. A. (Ed.). **The handbook of speech production**. Chichester: John Wiley & Sons, 2015. p. 34–58. Citado nas páginas 22 e 23.
- 54 ROTHENBERG, M. Acoustic interaction between the glottal source and the vocal tract. In: STEVENS, K. N.; HIRANO, M. (Ed.). **Vocal fold physiology**. Tokyo: University of Tokyo Press, 1981. p. 305–323. Citado nas páginas 22 e 23.
- 55 TITZE, I. R. The physics of small-amplitude oscillation of the vocal folds. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 83, n. 4, p. 1536–1552, 1988. Citado nas páginas 22, 23, 26 e 52.
- 56 TITZE, I. R.; JIANG, J. J.; HSIAO, T.-Y. Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. **Annals of Otology, Rhinology & Laryngology**, SAGE Publications Sage CA: Los Angeles, CA, v. 102, n. 1, p. 58–63, 1993. Citado nas páginas 22, 23 e 39.
- 57 GUPTA, V.; WILSON, T.; BEAVERS, G. A model for vocal cord excitation. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 54, n. 6, p. 1607–1617, 1973. Citado na página 26.
- 58 LILJENCRAFTS, J. A translating and rotating mass model of the vocal folds. **STL-QPSR**, Royal Institute of Technology in Stockholm - KTH, v. 32, n. 1, p. 1–18, 1991. Citado nas páginas 26, 35 e 87.
- 59 DRIOLI, C. A flow waveform adaptive mechanical glottal model. **TMH-QPSR**, Royal Institute of Technology in Stockholm - KTH, v. 43, p. 69–79, 2002. Citado na página 26.
- 60 KOIZUMI, T.; TANIGUCHI, S.; HIROMITSU, S. Two-mass models of the vocal cords for natural sounding voice synthesis. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 82, n. 4, p. 1179–1192, 1987. Citado na página 29.
- 61 MILLER, J.; PEREIRA, J.; THOMAS, D. Fluid flow through the larynx channel. **Journal of sound and vibration**, Elsevier, v. 121, n. 2, p. 277–290, 1988. Citado na página 29.
- 62 PEREIRA, J. C. Simlar: um programa para a simulacao da laringe durante a fonacao. **Research on Biomedical Engineering**, v. 7, n. 1, p. 486–498, 1990. Citado na página 29.
- 63 PELORSON, X.; HIRSCHBERG, A.; HASSEL, R. V.; WIJNANDS, A.; AUREGAN, Y. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. application to a modified two-mass model. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 96, n. 6, p. 3416–3431, 1994. Citado nas páginas 29, 35 e 87.
- 64 STEINECKE, I.; HERZEL, H. Bifurcations in an asymmetric vocal-fold model. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 97, n. 3, p. 1874–1884, 1995. Citado nas páginas 29, 50 e 60.

- 65 LOUS, N.; HOFMANS, G.; VELDHUIS, R.; HIRSCHBERG, A. A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design. **Acta Acustica united with Acustica**, S. Hirzel Verlag, v. 84, n. 6, p. 1135–1150, 1998. Citado na página 29.
- 66 BIRKHOLZ, P.; KRÖGER, B. J.; NEUSCHAEFER-RUBE, C. Articulatory synthesis of words in six voice qualities using a modified two-mass model of the vocal folds. In: **First International Workshop on Performative Speech and Singing Synthesis**. Vancouver, BC, Canada: [s.n.], 2011. v. 370. Citado nas páginas 29, 30 e 41.
- 67 \_\_\_\_\_. Synthesis of breathy, normal, and pressed phonation using a two-mass model with a triangular glottis. In: **Twelfth Annual Conference of the International Speech Communication Association, INTERSPEECH**. Florence, Italy: International Speech Communication Association - ISCA, 2011. Citado nas páginas 29, 30 e 41.
- 68 ISHIZAKA, K.; MATSUDAIRA, M. Fluid mechanical considerations of vocal cord vibration. **SCRL-Monograph Speech Communication Research Laboratory, Santa Barbara, Calif**, 1972. Citado na página 33.
- 69 TITZE, I. R. Parameterization of the glottal area, glottal flow, and vocal fold contact area. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 75, n. 2, p. 570–580, 1984. Citado nas páginas 34, 51 e 61.
- 70 SCHERER, R. C.; TITZE, I. R.; CURTIS, J. F. Pressure-flow relationships in two models of the larynx having rectangular glottal shapes. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 73, n. 2, p. 668–676, 1983. Citado na página 34.
- 71 TITZE, I. R. Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 111, n. 1, p. 367–376, 2002. Citado nas páginas 34, 35, 36, 41, 60, 61, 72, 86 e 87.
- 72 LILJENCRAANTS, J. **Speech synthesis with a reflection-type line analog**. Tese (Doutorado) — Dept. Speech Commun. and Music Acoust., Royal Inst. of Tech., Stockholm, Sweden, 1985. Citado nas páginas 35 e 45.
- 73 TITZE, I. R.; JIANG, J.; DRUCKER, D. G. Preliminaries to the body-cover theory of pitch control. **Journal of Voice**, Elsevier, v. 1, n. 4, p. 314–319, 1988. Citado na página 37.
- 74 YUMOTO, E.; KADOTA, Y.; KUROKAWA, H. Tracheal view of vocal fold vibration in excised canine larynxes. **Archives of Otolaryngology–Head & Neck Surgery**, American Medical Association, v. 119, n. 1, p. 73–78, 1993. Citado na página 38.
- 75 BAER, T. **Investigation of phonation using excised larynxes**. Tese (Doutorado) — Massachusetts Institute of Technology, 1975. Citado na página 39.
- 76 CHAN, R. W.; TITZE, I. R. **Dynamic shear modulus of vocal-fold tissues and phonosurgical biomaterials**. Tese (Doutorado) — Acoustical Society of America, 1997. Citado na página 39.
- 77 CHAN, R. W.; TITZE, I. R. Viscoelastic shear properties of human vocal fold mucosa: measurement methodology and empirical results. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 106, n. 4, p. 2008–2021, 1999. Citado na página 39.

- 78 TITZE, I. R. The human vocal cords: a mathematical model - part I. **Phonetica**, Karger Publishers, v. 28, n. 3-4, p. 129–170, 1973. Citado na página 40.
- 79 \_\_\_\_\_. The human vocal cords: a mathematical model - part II. **Phonetica**, Karger Publishers, v. 29, n. 1-2, p. 1–21, 1974. Citado na página 40.
- 80 TOKUDA, I. T.; ZEMKE, M.; KOB, M.; HERZEL, H. Biomechanical modeling of register transitions and the role of vocal tract resonators. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 127, n. 3, p. 1528–1536, 2010. Citado na página 40.
- 81 ZAÑARTU, M.; GALINDO, G. E.; ERATH, B. D.; PETERSON, S. D.; WODICKA, G. R.; HILLMAN, R. E. Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 136, n. 6, p. 3262–3271, 2014. Citado na página 40.
- 82 GALINDO, G. E.; PETERSON, S. D.; ERATH, B. D.; CASTRO, C.; HILLMAN, R. E.; ZAÑARTU, M. Modeling the pathophysiology of phonotraumatic vocal hyperfunction with a triangular glottal model of the vocal folds. **Journal of Speech, Language, and Hearing Research**, ASHA, v. 60, n. 9, p. 2452–2471, 2017. Citado nas páginas 40 e 41.
- 83 ALIPOUR-HAGHIGHI, F.; TITZE, I. Simulation of particle trajectories of vocal fold tissue during phonation. In: TITZE, I. R.; SCHERER, R. C. (Ed.). **Vocal Fold Physiology: Biomechanics, acoustics and phonatory control**. Denver, CO: The Denver Center for the Performing Arts, Inc., 1983. p. 183–190. Citado na página 41.
- 84 ALIPOUR, F.; BERRY, D. A.; TITZE, I. R. A finite-element model of vocal-fold vibration. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 108, n. 6, p. 3003–3012, 2000. Citado na página 41.
- 85 ROSA, M. d. O.; PEREIRA, J. C.; GRELLET, M.; ALWAN, A. A contribution to simulating a three-dimensional larynx model using the finite element method. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 114, n. 5, p. 2893–2905, 2003. Citado na página 41.
- 86 XUE, Q.; MITTAL, R.; ZHENG, X.; BIELAMOWICZ, S. Computational modeling of phonatory dynamics in a tubular three-dimensional model of the human larynx. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 132, n. 3, p. 1602–1613, 2012. Citado na página 41.
- 87 XUE, Q.; ZHENG, X.; MITTAL, R.; BIELAMOWICZ, S. Computational study of effects of tension imbalance on phonation in a three-dimensional tubular larynx model. **Journal of Voice**, Elsevier, v. 28, n. 4, p. 411–419, 2014. Citado na página 41.
- 88 GREISS, R.; ROCHA, J.; MATIDA, E. Validation of a finite element code for a continuum model of vocal fold vibration under the influence of a sessile polyp. **Canadian Acoustics**, v. 43, n. 1, p. 13–23, 2015. Citado na página 41.
- 89 \_\_\_\_\_. Modal analysis of a parameterized model of pathological vocal fold vibration. In: IEEE. **2016 IEEE EMBS International Student Conference (ISC)**. Ottawa, ON, Canada, 2016. p. 1–4. Citado na página 41.
- 90 STORY, B. H. **Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract**. Tese (Doutorado) — University of Iowa, 1995. Citado nas páginas 41, 44, 45 e 61.

- 91 STORY, B. H.; TITZE, I. R.; HOFFMAN, E. A. Vocal tract area functions from magnetic resonance imaging. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 100, n. 1, p. 537–554, 1996. Citado nas páginas 41, 61 e 69.
- 92 TITZE, I. R. **Principles of Voice Production**. Englewood Cliffs, NJ: Prentice-Hall, 1994. Citado na página 41.
- 93 STRUBE, H. W. The meaning of the kelly-lochbaum acoustic-tube model. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 108, n. 4, p. 1850–1855, 2000. Citado na página 44.
- 94 MATHUR, S.; STORY, B. H. Vocal tract modeling: Implementation of continuous length variations in a half-sample delay kelly-lochbaum model. In: IEEE. **Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No. 03EX795)**. Darmstadt, Germany, 2003. p. 753–756. Citado na página 44.
- 95 CHILDERS, D. G.; LEE, C. K. Vocal quality factors: Analysis, synthesis, and perception. **The Journal of the Acoustical Society of America**, v. 90, p. 2394–410, 12 1991. Citado na página 46.
- 96 BEHLAU, M.; MADAZIO, G.; FEIJÓ, D.; PONTES, P. Avaliação de voz. In: BEHLAU, M. (Ed.). **VOZ: O livro do especialista**. Rio de Janeiro - RJ: Livraria e Editora Revinter Ltda., 2001. I, cap. 3, p. 85–246. Citado nas páginas 46 e 48.
- 97 HIRANO, M. Psycho-acoustic evaluation of voice. **Clinical examination of voice**, Springer-Verlag, p. 81–84, 1981. Citado nas páginas 47, 48 e 51.
- 98 KEMPSTER, G. B.; GERRATT, B. R.; Verdolini Abbott, K.; BARKMEIER-KRAEMER, J.; HILLMAN, R. E. Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol. **American Journal of Speech-Language Pathology**, v. 18, n. 2, p. 124, may 2009. ISSN 1058-0360. Disponível em: <[http://ajslp.pubs.asha.org/article.aspx?doi=10.1044/1058-0360\(2008/08-0017\)](http://ajslp.pubs.asha.org/article.aspx?doi=10.1044/1058-0360(2008/08-0017))>. Citado nas páginas 47 e 48.
- 99 BEHLAU, M.; AZEVEDO, R.; PONTES, P. Conceito de voz normal e classificação das disfonias. In: BEHLAU, M. (Ed.). **VOZ: O livro do especialista**. Rio de Janeiro - RJ: Livraria e Editora Revinter Ltda., 2001. I, cap. 2, p. 53–84. Citado na página 47.
- 100 BEHLAU, M.; PONTES, P. **Avaliação e Tratamento das Disfonias**. São Paulo: Lovise, 1978. 312 p. Citado na página 47.
- 101 SILBERGLEIT, A. K.; JOHNSON, A. F.; JACOBSON, B. H. Acoustic analysis of voice in individuals with amyotrophic lateral sclerosis and perceptually normal vocal quality. **Journal of Voice**, Elsevier, v. 11, n. 2, p. 222–231, 1997. Citado na página 48.
- 102 NEMR, K.; SIMÕES-ZENARI, M.; CORDEIRO, G. F.; TSUJI, D.; OGAWA, A. I.; UBRIG, M. T.; MENEZES, M. H. M. GRBAS and cape-V scales: High reliability and consensus when applied at different times. **Journal of Voice**, v. 26, n. 6, p. 17–22, 2012. ISSN 08921997. Citado na página 48.
- 103 ENGLERT, M.; LOPES, L.; VIEIRA, V.; BEHLAU, M. Accuracy of acoustic voice quality index and its isolated acoustic measures to discriminate the severity of voice disorders. **Journal of Voice**, Elsevier, 2020. Citado na página 48.

- 104 KLATT, D. H.; KLATT, L. C. Analysis, synthesis, and perception of voice quality variations among female and male talkers. **the Journal of the Acoustical Society of America**, Acoustical Society of America, v. 87, n. 2, p. 820–857, 1990. Citado na página 49.
- 105 LALWANI, A.; CHILDERS, D. Modeling vocal disorders via formant synthesis. In: **IEEE. Proceedings of ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing**. Toronto, Ont., Canada, 1991. p. 505–508. Citado na página 49.
- 106 FANT, G.; LILJENCRANTS, J.; LIN, Q.-g. A four-parameter model of glottal flow. **STL-QPSR**, Royal Institute of Technology in Stockholm - KTH, v. 4, n. 1985, p. 1–13, 1985. Citado na página 49.
- 107 BANGAYAN, P.; LONG, C.; ALWAN, A. A.; KREIMAN, J.; GERRATT, B. R. Analysis by synthesis of pathological voices using the klatt synthesizer. **Speech communication**, Elsevier, v. 22, n. 4, p. 343–368, 1997. Citado na página 49.
- 108 KREIMAN, J.; GERRATT, B. R. Perception of aperiodicity in pathological voice. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 117, n. 4, p. 2201–2211, 2005. Citado nas páginas 49 e 72.
- 109 ISHIZAKA, K.; ISSHIKI, N. Computer simulation of pathological vocal-cord vibration. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 60, n. 5, p. 1193–1198, 1976. Citado nas páginas 50 e 52.
- 110 ZHANG, Y.; JIANG, J.; RAHN III, D. A. Studying vocal fold vibrations in parkinson's disease with a nonlinear model. **Chaos: An Interdisciplinary Journal of Nonlinear Science**, American Institute of Physics, v. 15, n. 3, p. 033903, 2005. Citado nas páginas 50 e 72.
- 111 ZHANG, Y.; JIANG, J. J. Nonlinear dynamic mechanism of vocal tremor from voice analysis and model simulations. **Journal of sound and vibration**, Elsevier, v. 316, n. 1-5, p. 248–262, 2008. Citado nas páginas 51 e 72.
- 112 XUE, Q.; MITTAL, R.; ZHENG, X.; BIELAMOWICZ, S. A computational study of the effect of vocal-fold asymmetry on phonation. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 128, n. 2, p. 818–827, 2010. Citado nas páginas 51 e 72.
- 113 FRAILE, R.; KOB, M.; GODINO-LLORENTE, J. I.; SÁENZ-LECHÓN, N.; OSMA-RUIZ, V. J.; GUTIÉRREZ-ARRIOLA, J. M. Physical simulation of laryngeal disorders using a multiple-mass vocal fold model. **Biomedical Signal Processing and Control**, Elsevier, v. 7, n. 1, p. 65–78, 2012. Citado na página 51.
- 114 FRAJ, S.; SCHOENTGEN, J.; GRENEZ, F. Development and perceptual assessment of a synthesizer of disordered voices. **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 132, n. 4, p. 2603–2615, 2012. Citado nas páginas 51 e 52.
- 115 MONTALVÃO, J.; CANUTO, J.; MIRANDA, L. **Bias-corrected estimator for intrinsic dimension and differential entropy—a visual multiscale approach**. 2020. ArXiv:1210.1594. Citado na página 54.
- 116 BENNETT, R. S. **Representation and analysis of signals part xxi. the intrinsic dimensionality of signal collections**. Baltimore, 1965. Citado na página 54.

- 117 COSTA, J. A.; HERO, A. O. Determining intrinsic dimension and entropy of high-dimensional shape spaces. In: KRIM, H.; YEZZI, A. (Ed.). **Statistics and Analysis of Shapes**. Basel-Boston-Berlin: Birkhäuser Verlag, 2006. p. 231–252. Citado na página 54.
- 118 SHANNON, C. E. A mathematical theory of communication. **The Bell system technical journal**, Nokia Bell Labs, v. 27, n. 3, p. 379–423, 1948. Citado na página 54.
- 119 GRASSBERGER, P.; PROCACCIA, I. Characterization of strange attractors. **Physical Review Letters**, American Physical Society (APS), v. 50, n. 5, p. 346–349, jan. 1983. Disponível em: <<https://doi.org/10.1103/physrevlett.50.346>>. Citado na página 54.
- 120 ROWEIS, S. T. Nonlinear dimensionality reduction by locally linear embedding. **Science**, American Association for the Advancement of Science (AAAS), v. 290, n. 5500, p. 2323–2326, dez. 2000. Disponível em: <<https://doi.org/10.1126/science.290.5500.2323>>. Citado na página 54.
- 121 TENENBAUM, J. B. A global geometric framework for nonlinear dimensionality reduction. **Science**, American Association for the Advancement of Science (AAAS), v. 290, n. 5500, p. 2319–2323, dez. 2000. Disponível em: <<https://doi.org/10.1126/science.290.5500.2319>>. Citado na página 54.
- 122 MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. **Journal of machine learning research**, v. 9, n. Nov, p. 2579–2605, 2008. Citado na página 54.
- 123 MA, S.-k. Calculation of entropy from data of motion. **Journal of Statistical Physics**, Springer, v. 26, n. 2, p. 221–240, 1981. Citado na página 55.
- 124 GRASSBERGER, P.; PROCACCIA, I. Measuring the strangeness of strange attractors. **Physica D Nonlinear Phenomena**, v. 9, n. 1-2, p. 189–208, 1983. Citado na página 55.
- 125 MONTALVÃO, J. 2020. Comunicação pessoal. Citado na página 55.
- 126 MONTALVÃO, J.; ATTUX, R.; SILVA, D. A pragmatic entropy and differential entropy estimator for small datasets. **Journal of Communication and Information Systems**, v. 29, n. 1, 2014. Citado na página 56.
- 127 JOHNS, M. M. Update on the etiology, diagnosis, and treatment of vocal fold nodules, polyps, and cysts. **Current opinion in otolaryngology & head and neck surgery**, LWW, v. 11, n. 6, p. 456–461, 2003. Citado na página 58.
- 128 BEHLAU, M.; MADAZIO, G.; PONTES, P. Disfonias organofuncionais. In: BEHLAU, M. (Ed.). **VOZ: O livro do especialista**. Rio de Janeiro - RJ: Livraria e Editora Revinter Ltda., 2001. I, cap. 5, p. 295–341. Citado na página 58.
- 129 SAMLAN, R. A.; STORY, B. H. Relation of structural and vibratory kinematics of the vocal folds to two acoustic measures of breathy voice based on computational modeling. **Journal of Speech, Language, and Hearing Research**, ASHA, 2011. Citado na página 61.
- 130 STORY, B. **LeTalker 1.22**. 2016. Disponível em: <<http://sal.arizona.edu/node/26>> (último acesso em 7 de março de 2020). Citado nas páginas 61, 67 e 69.
- 131 PÜTZER, M.; BARRY, W. J. **Saarbruecken voice database**. 2007. Available at: <http://stimmdb.coli.uni-saarland.de/>. Institut für Phonetik - Universität des Saarlandes. Citado na página 64.

- 132 MONTALVÃO, J.; MIRANDA, L.; DORIZZI, B. Straightforward working principles behind modern data visualization approaches. **IEEE Access**, IEEE, v. 9, p. 4242–4252, 2020. Citado na página 72.
- 133 LUCERO, J. C. Functional principal component analysis of vocal tract area functions. In: **INTERSPEECH**. Stockholm, Sweden: International Speech Communication Association - ISCA, 2017. p. 969–973. Citado na página 73.
- 134 \_\_\_\_\_. Parametrization of the vocal tract area function using a subset selection approach (I). **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 149, n. 6, p. 4036–4038, 2021. Citado na página 73.
- 135 MANRÍQUEZ, R.; PETERSON, S. D.; PRADO, P.; ORIO, P.; GALINDO, G. E.; ZAÑARTU, M. Neurophysiological muscle activation scheme for controlling vocal fold models. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, IEEE, v. 27, n. 5, p. 1043–1052, 2019. Citado na página 73.

# Apêndices

## APÊNDICE

# A

## Modelo *Body-Cover* com Massa Rotacional

As equações de movimento do modelo *body-cover* com massa rotacional proposto por Titze [71], como já apresentadas em 3.1.3.1 no Capítulo 3, são dadas como segue:

$$\begin{cases} I_c \ddot{\theta} + B_c \dot{\theta} + \kappa \theta = T_a, & \text{(A.1)} \\ m \ddot{\xi} + b(\dot{\xi} - \dot{\xi}_b) + k(\xi - \xi_b) = F_a, & \text{(A.2)} \\ M \ddot{\xi}_b + b(\dot{\xi}_b - \dot{\xi}) + k(\xi_b - \xi) + K \xi_b + B \dot{\xi}_b = 0. & \text{(A.3)} \end{cases}$$

A área glotal para o modelo, assumindo simetria entre as pregas é calculada da seguinte maneira:

$$a(z) = 2L [\xi_{0n} - (z - z_n) \tan(\theta_0 + \theta) + \xi] \quad \text{(A.4)}$$

em que  $L$  é o comprimento da glote e  $\xi_{0n}$  é abertura glotal pré-fonatória no ponto nodal  $z_n$ . Com base em (A.4), podemos encontrar as áreas no ponto nodal e nas porções inferior e superior e da glote:

$$a_n = 2L \text{Max} \{ \delta, [\xi_{0n} + \xi] \}, \quad \text{(A.5)}$$

$$a_1 = 2L \text{Max} \{ \delta, [\xi_{0n} - (0 - z_n) \tan(\theta_0 + \theta) + \xi] \}, \quad \text{(A.6)}$$

$$a_2 = 2L \text{Max} \{ \delta, [\xi_{0n} - (T - z_n) \tan(\theta_0 + \theta) + \xi] \}. \quad \text{(A.7)}$$

em que  $\delta$  é usado para restringir a um valor mínimo de área permitido. Normalmente,  $\delta = 0$ .

O equacionamento de aerodinâmico é variável de acordo com as condições em que as pregas vocais se encontram, como ilustrado na Figura A.1. Sendo

$$P_{kd} = \frac{P_s - P_e}{1 - k_e} \quad \text{(A.8)}$$

a pressão cinética no ponto de separação do fluxo,

$$k_e = 2 \frac{a_d}{A_e} \left( 1 - \frac{a_d}{A_e} \right) \quad \text{(A.9)}$$

o coeficiente de recuperação de pressão proposto por Ishizaka e Flanagan [24],

$$z_d = \min \left\{ T, \max \left[ 0, -\frac{0.2}{\tan(\theta_0 + \theta)} [\xi_{0n} + z_n \tan(\theta_0 + \theta) + \xi] \right] \right\} \quad \text{(A.10)}$$

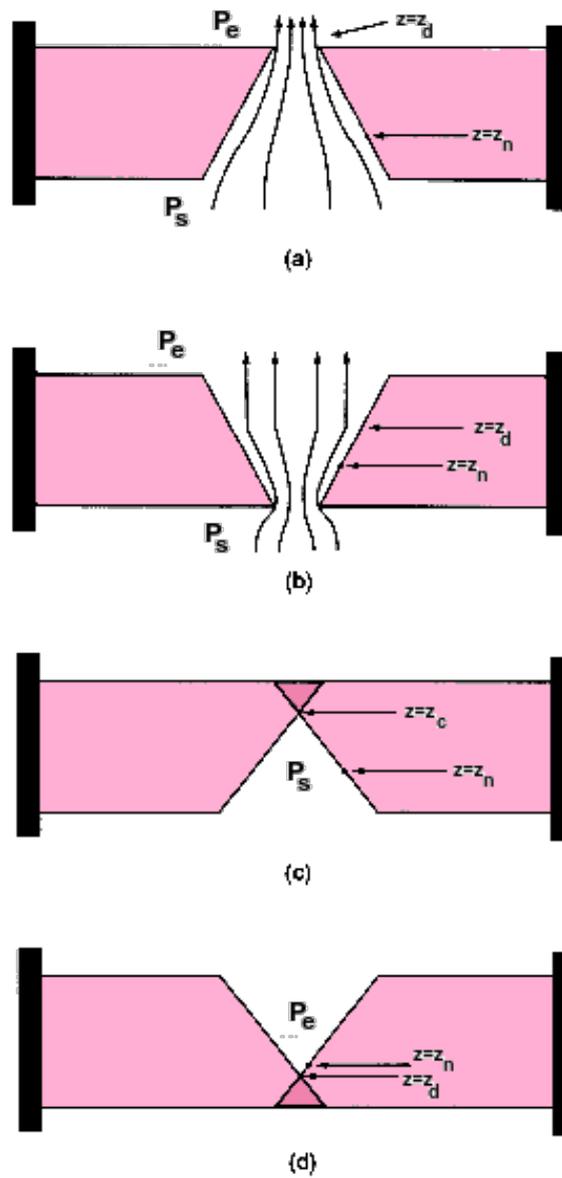
o ponto de separação dos fluxos (ilustrado na Figura A.1), obtido com base na formulação de Liljencrants [58] e no equacionamento de Pelorson et al. [63], e

$$z_c = \min \left\{ T, \max \left[ 0, z_n + \frac{\xi_{0n+\xi}}{\tan(\theta_0 + \theta)} \right] \right\} \quad (\text{A.11})$$

o ponto em que ocorre o contato nas pregas (ilustrado na Figura A.1), em caso de colisão, e

$$P_h = \frac{P_s + P_e}{2} \quad (\text{A.12})$$

a pressão hidrostática, temos que as forças agindo na região superior e inferior ( $F_u$  e  $F_l$ ,



(Adaptado de Titze [71])

Figura A.1 – Esboço das configurações utilizadas para o cálculo das pressões na glote: (a) glote convergente; (b) glote divergente; (c) glote com contato na porção superior; (d) glote com contato na porção inferior.

respectivamente) nas pregas podem ser calculadas, de acordo com as diferentes configurações e condições, da seguinte maneira:

- Para a glote convergente (condição (a) na Figura A.1), as forças são obtidas da seguinte maneira:

$$F_l = Lz_n \left( P_s - \frac{a_2^2}{a_n a_1} P_{kd} \right), \quad (\text{A.13})$$

$$F_u = L(T - z_n) \left( P_s - \frac{a_2}{a_n} P_{kd} \right). \quad (\text{A.14})$$

- Para a glote divergente (condição (b) na Figura A.1), as forças são dependentes das seguintes condições:

- Se  $z_d \leq z_n$ :

$$F_l = Lz_n P_s - L \left( z_n - z_d + \frac{a_d}{a_1} z_d \right) P_{kd}, \quad (\text{A.15})$$

$$F_u = L(T - z_n) (P_s - P_{kd}). \quad (\text{A.16})$$

- Se  $z_d > z_n$ :

$$F_l = Lz_n \left( P_s - \frac{a_d^2}{a_n a_1} P_{kd} \right), \quad (\text{A.17})$$

$$F_u = L(T - z_n) P_s - L \left[ (T - z_d) + \frac{a_d}{a_n} (z_d - z_n) \right] P_{kd}. \quad (\text{A.18})$$

- Para a glote com contato na porção superior (condição (c) na Figura A.1), as forças são dependentes das seguintes condições:

- Se  $z_c \geq z_n$ :

$$F_l = Lz_n P_s, \quad (\text{A.19})$$

$$F_u = L(z_c - z_n) P_s + L(T - z_c) P_h. \quad (\text{A.20})$$

- Se  $z_c < z_n$ :

$$F_l = Lz_c P_s + L(z_n - z_c) P_h, \quad (\text{A.21})$$

$$F_u = L(T - z_n) P_h. \quad (\text{A.22})$$

- Para a glote com contato na porção inferior (condição (d) na Figura A.1), as forças são dependentes das seguintes condições:

- Se  $z_c \leq z_n$ :

$$F_l = Lz_c P_h + L(z_n - z_c) P_e, \quad (\text{A.23})$$

$$F_u = L(T - z_n) P_e. \quad (\text{A.24})$$

– Se  $z_c > z_n$ :

$$F_l = Lz_n P_h, \quad (\text{A.25})$$

$$F_u = L(z_c - z_n) P_h + L(T - z_c) P_e. \quad (\text{A.26})$$

- Para a glote totalmente fechada, com contato na porção inferior e porção superior, as forças são calculadas como segue:

$$F_l = Lz_n P_h, \quad (\text{A.27})$$

$$F_u = L(T - z_n) P_h. \quad (\text{A.28})$$

A força  $F_a$  e o torque  $T_a$  aplicados nas pregas vocais, presentes nas Equações (A.1), (A.2) e (A.3), são calculadas de acordo com as seguintes expressões:

$$F_a = F_l + F_u, \quad (\text{A.29})$$

$$T_a = F_l \left( \frac{z_n}{2} \right) - F_u \left( \frac{T - z_n}{2} \right). \quad (\text{A.30})$$