Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Sintese de Voz

Leandro Ramalho Motta Ferreira

Monografia apresentada como requisito parcial
para conclusão do Curso de Computação — Licenciatura

Orientador
Prof. Dr. Jorge Carlos Lucero

Brasília
2016

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Curso de Computação — Licenciatura

Coordenador: Prof. Dr. Coordenador

Banca examinadora composta por:

Prof. Dr. Jorge Carlos Lucero (Orientador) — CIC/UnB
Prof. Dr. Professor I — CIC/UnB
Prof. Dr. Professor II — CIC/UnB

CIP — Catalogação Internacional na Publicação

Ferreira, Leandro Ramalho Motta.

Sintese de Voz / Leandro Ramalho Motta Ferreira. Brasília : UnB, 2016.
57 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2016.

1. Sintese, 2. Voz, 3. Saúde

CDU 004.4

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília–DF — Brasil

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Sintese de Voz

Leandro Ramalho Motta Ferreira

Monografia apresentada como requisito parcial
para conclusão do Curso de Computação — Licenciatura

Prof. Dr. Jorge Carlos Lucero (Orientador)
CIC/UnB

Prof. Dr. Professor I
CIC/UnB

Prof. Dr. Professor II
CIC/UnB

Prof. Dr. Coordenador
Coordenador do Curso de Computação — Licenciatura

Brasília, 10 de maio de 2016

Dedicatória

Dedico a....

Agradecimentos

Agradeço a....

Resumo

AINDA Não tem

Palavras-chave: Sintese, Voz, Saúde

Abstract

Still there isn't.

Keywords: Synthesis, Voice, Health

Sumário

1 Introdução **1**
1.1 Motivação 1

2 Aparelho Fonador **2**
2.1 Anatomia da Voz 2
2.1.1 Aparelho Fonador 2
2.1.2 Músculos e Cartilagens 3
2.1.3 Pregas Vocais 4
2.2 Fundamentos Biofísicos para a Produção da Voz 5
2.2.1 A Biomecanica da Laringe 5
2.2.2 Reflexão de Som 7
2.2.3 Fluxo de Ar na Glote 8
2.2.4 O Sistema Pulmonar 10
2.2.5 Leis de Conservação para Fluxo de Dutos 11
2.3 Oscilação das Pregas Vocais 12
2.3.1 Lei Bernouli 13
2.3.2 Critérios para Oscilação 13
2.3.3 Tipos de Oscilação 13
2.4 Sistema Auditivo 13
2.4.1 Introdução 13
2.4.2 Intensidade 14
2.5 Voz e Propriedades Linguísticas 14
2.5.1 Vogais 14
2.5.2 Consoantes 14

3 Síntese de Voz **15**
3.0.1 Modelo Massa-Mola Auto Sustentável 15
3.0.2 Synpath 15
3.0.3 HMMs 17
3.0.4 MHRSMM 17
3.0.5 FrameWorks Sintetisador de Voz 18
3.0.6 Envoltoria F0 19
3.0.7 Sintese de Voz em Mandarím 19

Referências **20**

Lista de Figuras

2.1 Aparelho Fonador 3
2.2 Secção coronal da laringe e parte superior da traquéia 3
2.3 Músculos Intrínsecos da Laringe 4
2.4 Músculos Extrínsecos da Laringe 4
2.5 Cordas Vocais e Componentes 5
2.6 Curva Hipotética Tensão-Deformação das Cordas vocais Humanas (17) . . 7
2.7 Pressões Atuantes no Torso (17) 9
2.8 Sistema Pulmonar (17) 10
2.9 Dinâmica do Fluído em um Duto com Mudança de Área Transversal . . . 11

Lista de Tabelas

Capítulo 1

Introdução

1.1 Motivação

A UnB possúi atuação na área de simulação de voz a ser considerada. Essa atuação considero ser principalmente ser por esforço e trabalho do Doutor Jorge Carlos Lúcero do Departamento de Ciência da Computação. Seu trabalho de simulação de voz e simulação de voz com patologias na área da saúde precisa ter um apoio. Este Trabalho de conclusão de curso é feito com intuito de seguir uma nova direção na área de simulação de voz mas mantendo os avanços já feitos por outros trabalhos, sendo essa direção nova a do entretenimento.

Mercadológicamente a música é um campo muito interessante economicamente, onde produtos da área de síntese de voz estão cada vez mais avançados, por exemplo o sucesso das Vocaloids no japão. Por outro lado sintetizar vozes e a área de música são atualmente meu sonho de projeto.

Capítulo 2

Aparelho Fonador

2.1 Anatomia da Voz

Para estudar a produção e a síntese da voz, é necessário ter um conhecimento acerca da anatomia e do funcionamento físico da voz 11. Sendo assim, as subseções seguintes descreverão brevemente detalhes da anatomia do sistema fonador humano e como o som é produzido, moldado e inuenciado por este sistema.

2.1.1 Aparelho Fonador

O estudo do aparelho fonador começa-se por suas estruturas e componentes importantes. Após um estudo detalhado dos fenômenos físicos e como se comportam é essencial também.

A Figura 2.1 11, mostra os órgãos associados com a produção da voz.

Dentro das condições normais, a voz é produzida quando um uxo de ar vindo dos pulmões é convertido em energia acústica através da vibração das pregas vocais, localizadas na laringe. Os padrões de vibrações resultantes são moldados acusticamente quando o som passa pelo trato vocal acima da laringe. O sistema respiratório serve como uma fonte de potência para a produção do som, sendo responsável por movimentar o ar através do trato vocal. A laringe atua como um oscilador convertendo a potência aerodinâmica produzida em energia sonora, sendo frequentemente retratada como a fonte da voz. No entanto, a mais importante função da laringe não é a produção de som, e sim, vedar as vias aéreas aos pulmões completamente, protegendo-as de objetos estranhos ou líquidos, principalmente durante a deglutinação. De maneira análoga, a laringe serve como uma válvula de acesso às vias respiratórias e por essa característica, atua também no controle do uxo de ar que por elas passam. Sendo assim,é fácil notar que há uma necessidade de mobilidade para toda estruturada laringe, logo é de se esperar que sua estrutura seja formada em sua maioria por cartilagens. De fato o é, com exceção de um osso chamado de Hioide, a laringe é basicamente formada por cartilagens e músculos. A seguir, analisaremos brevemente a dinâmica dos músculos e cartilagens da laringe.

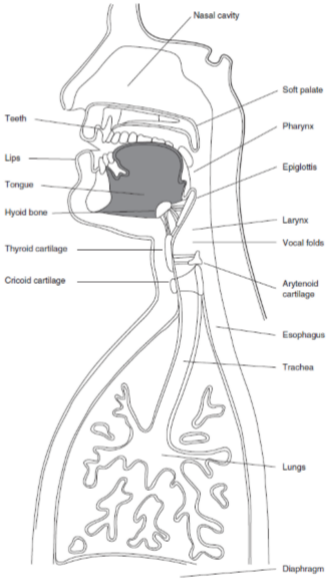


Figura 2.1: Aparelho Fonador

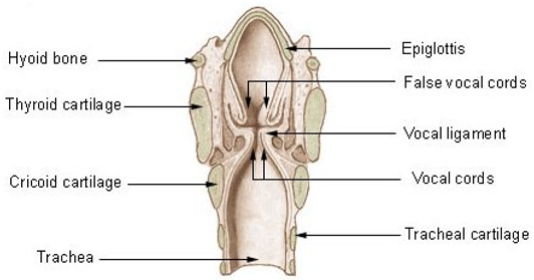


Figura 2.2: Secção coronal da laringe e parte superior da traquéia

2.1.2 Músculos e Cartilagens

Os músculos e cartilagens atuam diretamente no processo de abdução e adução das pregas vocais. Estas estão localizadas dentro da laringe e devido à dinâmica das cartilagens e dos músculos, podem executar os movimentos citados de forma a produzir som.

Cartilagens da Laringe

A Figura 2.2, mostra uma secção da laringe, detalhando as cartilagens presentes. De maneira sucinta, estas cartilagens servem como base de interconexão para os músculos intrínsecos ao redor da laringe. Dentre as cartilagens acima, a epiglote é responsável por vedar as vias respiratórias movimentando-se sobre a entrada das mesmas. O resto das cartilagens garantem a mobilidade da laringe em conjunto com outras estruturas como por exemplo o sternum.

Músculos da Laringe

Os músculos na laringe podem ser divididos em dois grupos, os intrínsecos e os extrínsecos 6. Os músculos intrínsecos interconectam as cartilagens da laringe, ao passo que,

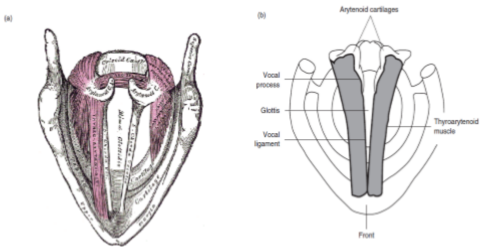


Figura 2.3: Músculos Intrínsecos da Laringe

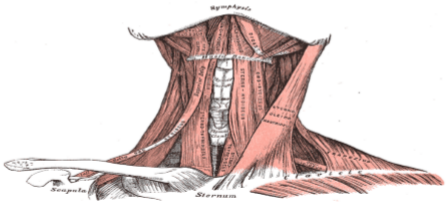


Figura 2.4: Músculos Extrínsecos da Laringe

os extrínsecos conectam a laringe à outras estruturas externas, como o osso hióide. A Figura 2.3 detalha alguns dos músculos intrínsecos da laringe. Alguns desses músculos têm inuência direta em algumas características da voz. Por exemplo, o músculo cricotiroideo é o músculo primário utilizado no controle do tom da voz. Por sua vez, o músculo cricoaritenoideo posterior atua na abdução das pregas vocais, ao passo que o músculo interaritenoideo atua como adutor das pregas vocais.

Os músculos extrínsecos, Figura 2.4, atuam basicamente no movimento da laringe, agindo como depressor e elevador da estrutura laríngea. Além disso também conectam estruturas do trato vocal à estrutura laríngea, como por exemplo a língua ao osso hioide.

2.1.3 Pregas Vocais

As pregas vocais, como dito anteriormente, estão localizadas dentro da laringe, mais especícamente na parte superior da traqueia. Elas estão posteriormente ligadas às cartilagens aritenoides, e anteriormente ligadas à cartilagem tireoide. As suas bordas exteriores estão ligadas a músculos na laringe, enquanto as suas bordas interiores são livres.

As bordas das pregas vocais são construídas de epitélio, sendo compostas também de algumas bras musculares. As pregas vocais são bandas triangulares planas de cor branca e acima de ambos os lados destas, se encontram as pregas vestibulares ou falsas pregas vocais.

O espaço entre as pregas vocais é chamado de glote, sendo que o que está acima da glote é denominado supraglotal e o que está abaixo é denominado subglotal. A Figura 1.5 mostra em mais detalhes a anatomia das pregas vocais, os componentes musculares e as cartilagens atuantes.

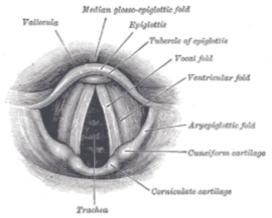


Figura 2.5: Cordas Vocais e Componentes

2.2 Fundamentos Biofísicos para a Produção da Voz

2.2.1 A Biomecanica da Laringe

Primeiramente, devemos ter em mente que o principal papel da laringe não é a produção de voz e sim a proteção das vias respiratórias. Dito isto, podemos fazer uma simples analise, visto que apesar de seu papel principal, a laringe também atua como um instrumento da fala humana. Se analisarmos os instrumentos criados pelo homem podemos notar que estes dependem basicamente de sua geometria, do material que o compõe e da interação de suas partes acústicas. Do mesmo modo, a laringe possui uma determinada geometria, é composta por tecido humano e em conjunto como trato vocal compõe a parte acústica do nosso corpo. Entretanto, nada é tão simples, a sua geometria e as propriedades do material humano envolvidos na produção do som são bastante irregulares(17).

Outra analogia interessante é sobre o instrumento e quem o utiliza. Um bom pianista por exemplo, sua musica é boa porque ele é habilidoso com o instrumento? Ou sua musica é boa por que o instrumento é bem feito e o som gerado por este é agradável? Ou os dois? Essas perguntas também podem ser feitas com respeito a voz. Para entendermos o que inuência na qualidade da síntese da voz é necessário analisar a biomecânica da voz, que nada mais é analisar o movimento do material vivo e as forças atuantes sobre ele(17).

Fatores Biológicos que Afetam a Produção de Som na Laringe

A parte da (bio)mecânica que se relaciona diretamente com a atuação da laringe na produção do som é a mecânica dos meios contínuos, que é a parte da mecânica que lida com a matéria distribuída sobre uma determinada região no espaço, e consequentemente, se contrapõe à mecânica de partículas. Dentro da mecânica de meios contínuos, mais especicamente, a parte que irá nos auxiliar no estudo do comportamento da laringe se chama mecânica de sólidos e uídos.

Dito isto, analisaremos a seguir alguns conceitos físicos que tem forte ligação com os processos que ocorrem na região da laringe durante a produção do som:

Tensao

Tensão é quantidade de força por unidade de área (17) , podemos escrever na forma da equação

$$\sigma = \frac{F}{A}$$

Sendo: f : força aplicada.
A: área de aplicação desta força.

Deformação

é a medida de deformação de um meio após a aplicação de uma tensão17 e pode ser escrito na forma da equação abaixo:

$$\epsilon = (L - \frac{L_0}{L_0})$$

Onde ϵ é a medida de deformação, L é o comprimento após a tensão e L0 é o comprimento antes da tensão. Normalmente uma dada deformação em uma dimensão resulta em uma deformação oposta em outra dimensão em um dado meio.

Se uma deformação é uniforme por todo o corpo de um objeto, então chamamos de compressão, se o volume diminui por conta desta deformação, e expansão, se o volume aumenta.

Viscosidade

É a velocidade de deformação(consequentemente, de restauração) de um determinado fluido quando atuam forças de tensão no mesmo. Matematicamente pode ser expresso conforme a equação seguinte: **?**

$\sigma = \eta * \frac{d*\epsilon}{dt}$

Para η viscosidade e t tempo. Quanto maior a viscosidade, mais devagar será a deformação de um meio.

Elasticidade

É uma propriedade do meio que determina quão completa será a restauração do meio após uma dada deformação. Os conceitos e propriedades descritos acima são extremamente importantes para se entender a manutenção da produção do som. Como dito anteriormente no capitulo 1, as pregas vocais são músculos e músculos são compostos por bras, logo, as pregas vocais consistem de uma grande concentração de bras. Além disso, entre as fíbras que compõe as pregas vocais existem também uidos atuantes, o que caracteriza as pregas vocais como um material viscoelástico. Para se entender a capacidade de absorção e regeneração das pregas vocais, em detrimento das vibrações de alta frequência e as pressões do ar, deve-se primeiramente estudar as propriedades absorcivas do material que as compõe. Ou seja, em outras palavras, deve-se estudar as propriedades mecânicas do tecido viscoelástico, e uma ferramenta que facilita o entendimento é o estudo da curva força-alongamento de um material. Entretanto, construir uma curva de força-alongamento depende essencialmente da geometria da amostra do material (17) e, por se tratar de uma material biológico, é difícil obter uma geometria precisa pois as bras estão constantemente se reorientando em detrimento de lesões e cortes brosos. Para viabilizar este estudo, Titze (17) sugere normalizar as forças atuantes e as deformações resultantes para que não haja a dependência direta da geometria. Essa normalização se dá através da substituição da curva força-alongamento por um curva tensão-deformação. A Figura2.8 retirada do estudo feito por Titze (17) demonstra uma curva hipotética de

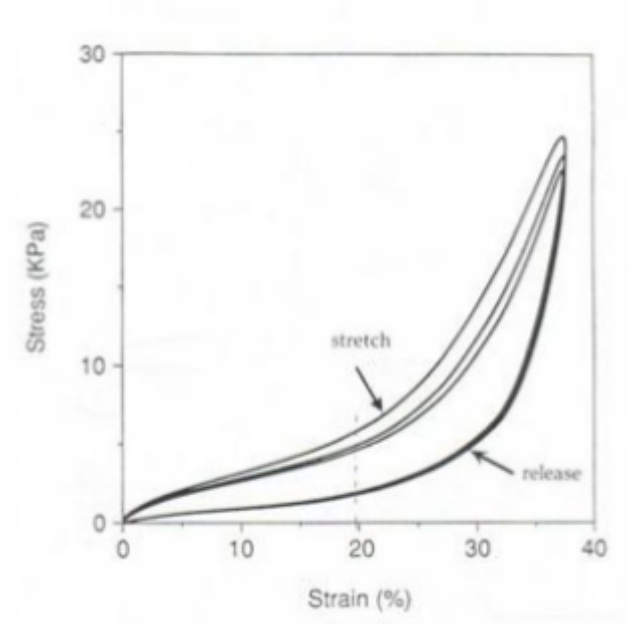


Figura 2.6: Curva Hipotética Tensão-Deformação das Cordas vocais Humanas (17)

tensão-deformação para os tecidos que compõe as pregas vocais humanas. Esta Figura ilustra o comportamento das bras das pregas vocais através da relação entre uma força atuante e a deformação gerada por esta. A importância desta analises e deve ao fato de que é possível estabelecer uma relação direta entre nódulos vocais e uma fonação prolongada, alta(em termos de frequência) e intensa. A partir da análise desta curva é então possível estabelecer um precedente para a formação de nódulos vocais: a frequência e amplitude da vibração estão diretamente ligadas ao surgimento de um nódulo vocal e consequentemente o de uma fenda pois a força de impacto entre as pregas vocais é proporcional à altura tonal quando acima do tom natural e à intensidade durante a fonação.

2.2.2 Reflexão de Som

Um fenômeno ligado a rigidez e amortecimento entre um meio e outro.(10) Ondas quando tentam penetrar em um segundo meio, sendo o segundo meio rígido, as partículas primeiro meio se aglomeram tentando passar porém falham, seu acúmulo de particulas geram pressão que acabam criando uma outra onda no primeiro meio decorrente da primeira onda.(7)

O mesmo ocorre com o meio 2 sendo totalmente não rígido e o primeiro meio sendo bem rígido, Exaurindo excesso de particulas do meio 1 no meio 2 criando rarefação no meio 1, o que cria uma outra onda de pressão negativa (1). A propagação é sempre em direção oposta à fonte, no caso é na direção contrária à coluna de ar(meio 1).

Para conrmar o precedente supracitado, tomemos a segunda lei de Newton:

$$f = \frac{impeto}{periodo}$$

Isto pode ser traduzido em termos de proporção:

$$f \propto \frac{impeto}{periodo}$$

Onde m é a massa do tecido das pregas vocais, T é o período de vibração. A velocidade máxima(V_m)é proporcional à máxima amplitude e inversamente proporcional ao tempo de fechamento das pregas vocais. Ou seja:

$$V_m \propto \frac{A}{T}$$

2.2.3 Fluxo de Ar na Glote

Em seções posteriores serão apresentados modelos para representar a vibração das pregas vocais, entretanto para entender por completo esses modelos, devemos antes entender o papel do uxo de ar nesse processo e para isso devemos analisar primeiro como a pressão de ar é gerada e exercida. À primeira vista, imagine-se que o uxo de ar é gerado somente pelos pulmões, entretanto, se formos analisar como inspiramos e expiramos percebemos que existem outros fatores contribuintes para a criação e manutenção de um uxo de ar nas vias respiratórias. O pulmão, como é de se imaginar, é responsável pelo armazenamento do ar e pela troca de oxigênio entre os tecidos que o compõe. Aqui, porém, somente nos interessa esse papel de armazenamento que o pulmão exerce. Ao inspirar o ar, o pulmão aumenta a pressão interna gerando uma tensão aplicada ao tecido elástico que o compõe, a expansão desse tecido gera uma mudança de pressão devido ao aumento do volume pulmonar e à deformação do tecido pulmonar. Em conjunto com essa mudança de pressão, o diafragma, músculo localizado abaixo dos pulmões, é comprimido levando a uma outra mudança de pressão. O diafragma, por sua vez, é um músculo que pode ser controlado, e em virtude disso, é possível controlar a pressão exercida nos pulmões, algo que cantores utilizam para sustentar a técnica vocal. Assim como o diafragma, a caixa torácica também sofre mudanças, expandindo devido ao aumento de volume do pulmão o que gera uma pressão torácica sobre os pulmões. Devido ao fato de ser uma estrutura óssea e portanto ter uma maior rigidez e menor controle se comparado à parede abdominal, a pressão exercida nos pulmões pela caixa torácica é signicantemente menor do que a pressão exercida pela aplicação de uma força à parede abdominal. A Figura 2.2 ilustra essa transmissão de pressão por todo o torso citada acima:

Fluxo de ar na Glote:

Como descrito no artigo de Elias temos informações como descobrimos a pressão via fluxo de ar(2)

$$U_g = + - (\frac{-a_m}{A_*} + [(a_m)^2 + -(\frac{4K_t}{C^2\rho})](P_s^+ - P_i^-))$$

A* =

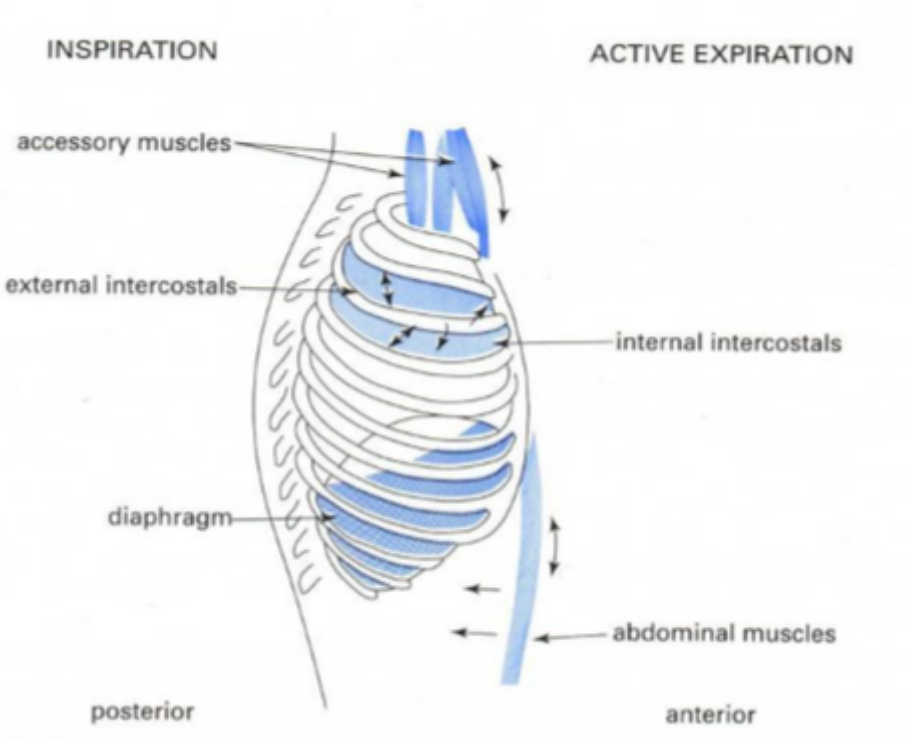


Figura 2.7: Pressões Atuantes no Torso (17)

$\rho =$ Densidade do Ar

c = Velocidade do som

$\mathrm{P}_s e P_i =$ Pressão de incidencia na entra e saída da glote

Uma vez descoberto o fluxo, as pressões de reflexão $\mathrm{P}_e^s P_i^+ podemserencontradascomaseguinteequao$:

$$P_s^- = P_s^+ - (\rho c/A_s)U_g$$

$$P_i^+ = P_i^- - (\rho c/A_i)U_g$$

Quando ocorre o fechamento da glote então alguns parametros assumem valores conhecidos. a(t) = 0 , $P_g = \frac{P_s - P_i}{2}$ e $U_g = 0$

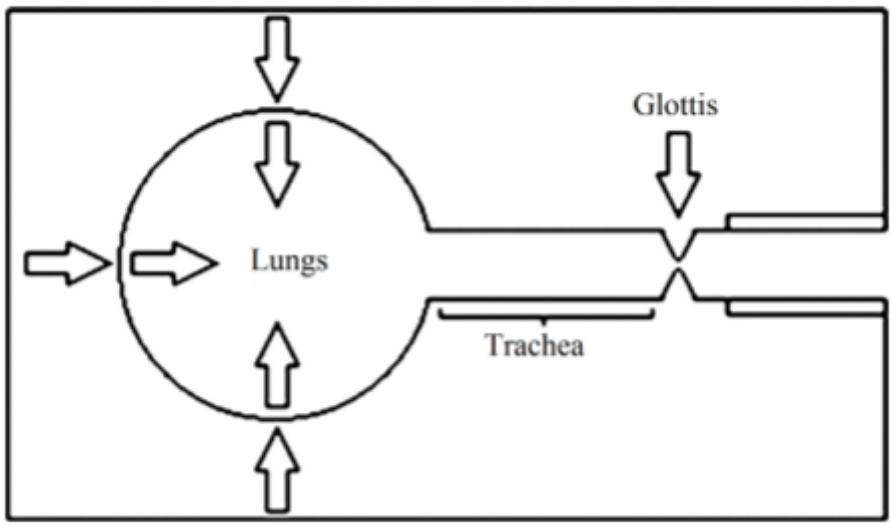


Figura 2.8: Sistema Pulmonar (17)

2.2.4 O Sistema Pulmonar

O sistema pulmonar consiste dos pulmões e vias respiratórias, compostas pela traqueia, glótis e trato vocal. Conforme descrito anteriormente, o tórax e o abdômen em conjunto com o pulmão atuam na geração do uxo de ar. A glótis, por sua vez, atua como reguladora do uxo de ar através de variações em seu fechamento que permitem mudanças na pressão envolvida com o movimento do uxo do ar pelas vias respiratórias. Isso signica que é possível dar constância ao uxo de ar(Figura 2.3)

O Processo Físico da Respiração

Uma importante lei que nos auxilia no entendimento da relação entre o volume pulmonar e a pressão pulmonar é a Lei de Boyle. A lei nos diz que em um ambiente cujas paredes não são rígidas e em uma temperatura constante, pressão e volume são inversamente proporcionais. Se aplicarmos a lógica dessa lei à respiração humana, ca fácil visualizar o processo físico que envolve a respiração. Quando inspiramos, o diafragma é contraído o que aumenta o volume do pulmão e diminui a pressão interna. Consequentemente, o pulmão se enche de ar. Relacionada à pressão atmosférica, a pressão pulmonar se torna menor, o que gera uma busca pelo equilíbrio das pressões no meio. Esse equilíbrio é alcançado ao se expelir o ar, diminuindo o volume pulmonar e então aumentando a pressão interna novamente. Na verdade é um pouco mais complexo do que isso, entretanto, esse conceito de equilíbrio de pressões é fundamental para o entendimento da abdução e adução das pregas vocais, o que permite a transformação da energia aerodinâmica em energia acústica.

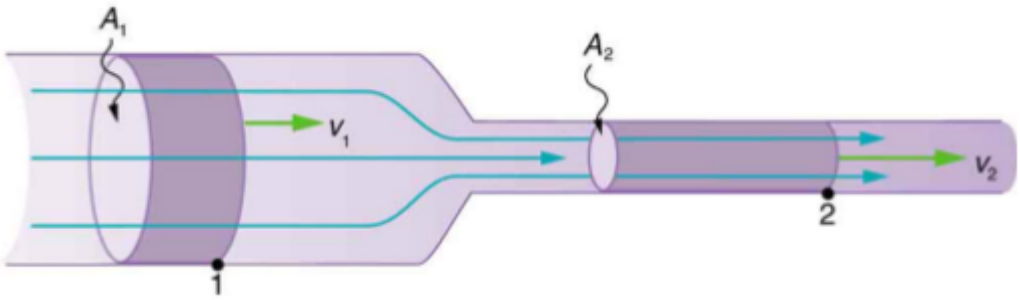


Figura 2.9: Dinâmica do Fluído em um Duto com Mudança de Área Transversal

2.2.5 Leis de Conservação para Fluxo de Dutos

Porém, antes de abordar a transformação de energia aerodinâmica em acústica, é importante se familiarizar com alguns conceitos que regem o comportamento de uidos em dutos. Isso porque trataremos o uxo de ar na síntese de voz como um uído e a modelagem da glote e do trato vocal como a concatenação de pequenos dutos.

Lei da Continuidade para um Fluxo Incompressível

Quando dizemos que um uido é incompressível signica dizer que sua densidade não se altera quando este é forçado a passar por uma constrição. Imagine agora um uido connado a um duto que possui uma mudança, ao longo de seu comprimento, em sua área transversal formando uma constrição(Figura 2.4). Se não for permitido ao uido vazar pelas paredes do duto, então todas as partículas do uido devem ser mantidas mesmo durante a mudança de área transversal ao longo do duto. Para que isso seja possível, as partículas devem acelerar durante a constrição, mantendo-se constante o número de partículas em movimento por unidade de área. Essa relação pode ser expressa pela equação 2.4.

$$v_1 * A_1 = v_2 * A_2 = constant = U$$

Onde v e A são respectivamente a velocidade da partícula e a área em que se encontra. U é o uxo. Matematicamente então, a Lei da Continuidade expressa que um uxo incompressível em um duto é constante, independente do que acontece com a área transversal ao longo do mesmo (17)

Lei de Bernoulli área conservação de Energia

Outra importante lei de aplicação para o entendimento da síntese de voz é a Lei de Bernoulli para Conservação de Energia. Essa lei foi desenvolvida a partir do reconhecimento de que a energia total atuante no uido em qualquer ponto durante o trajeto no duto tem duas componentes, a saber, uma energia potencial e uma energia cinética. A energia potencial está relacionada diretamente com a pressão no duto e a energia cinética.

é proporcional ao quadrado da velocidade da partícula. Portanto, podemos expressá-la conforme a equação 2.5.

$$P + \frac{1}{2} * \rho * v^2 = constant$$

Onde ρ é a densidade do uido, v a velocidade da partícula e P a pressão no duto. Segue disso o Princípio de Bernoulli: Se a energia em uma corrente de uido é constante, um aumento na velocidade da partícula deve ser acompanhado de uma queda na pressão (17).

Analisando a Figura 2.9, a pressão na área de constrição deve ser menor que a pressão na área maior do duto, considerando que não hajam perdas de energia no processo. Esse conceito é importante para se compreender o caráter oscilatório das pregas vocais.

Resistência Glotal

A resistência a um uxo é uma característica do sistema de transporte, ou seja, do meio, e pode ser descrita como a razão entre pressão e uxo(equação2.6)

$$R = \frac{P}{U}$$

Em sistemas de transporte, as constrições atuam como ponto de resistência a um uxo. No caso das vias respiratórias humanas, isso normalmente ocorre na glote ou em alguma parte do trato vocal com pouco espaço. Podemos então denir que a resistência glotal será a pressão na glote dividida pelo uxo que a atravessa. A resistência glotal está diretamente ligada a qualidade vocal pois é responsável por auxiliar no controle do uxo de ar.

2.3 Oscilação das Pregas Vocais

Nesta seção analisaremos os processos e princípios físicos que ocorrem na laringe para que as pregas vocais oscilem, tendo como objetivo destrinchar tais processos a m de entendermos a reprodução dos mesmos por modelos computacionais utilizados para simulação da voz humana. Os primeiros estudos sobre vibração das pregas vocais especulavam que as pregas vocais se ajuntavam por um efeito de pressão negativa(Princípio de Bernoulli) na glote. A esta descrição de vibração nas pregas vocais se deu o nome de Teoria Mioelástica Aerodinâmica da Vibração das Pregas Vocais, conhecida também pelo nome de seu criador como Teoria de van der Berg . A teoria de van der Berg serviu como marco para o desenvolvimento teórico do campo de estudos da voz e consequentemente o aparecimento dos primeiros modelos matemáticos de representação do funcionamento das pregas vocais. Entretanto, apesar disso, sua teoria é inadequada para explicar a vibração auto sustentável das pregas vocais (**?**). Isso porque as forças de Bernoulli por si só não são capazes de distinguir entre os movimentos interiores e posteriores das pregas vocais. Assim sendo, são necessários mecanismos para prover um aumento ou decréscimo às forças de Bernoulli durante a abertura e o fechamento das pregas vocais respectivamente, caso contrário, as oscilações serão amortecidas.

2.3.1 Lei Bernouli

Energia potencial e energica cinética em fluídos se mantém a mesma porém em proporções diferentes (16):

$$P+\frac{\rho * v^2}{2} = ConstanteSendo : \rho = DensidadedofluidoP = Pressonodutoondeofludoseencontrav = v$$

2.3.2 Critérios para Oscilação

Alguns critérios devem ser atendidos para que um determinador padrão de movimento seja considerado como uma oscilação mecânica, a saber:

No sistema onde ocorre o movimento deve haver uma posição de equilíbrio estável, que é caracterizada por uma força restaurativa que sempre acelera o corpo em movimento de volta para a sua posição de repouso. Deve haver inércia(no caso do sistema mecânico, a massa atua como propriedade de inércia) no sistema para superar esta posição de equilíbrio. A perda, em excesso, de energia por ciclo de oscilação deve ser zero.. . .

2.3.3 Tipos de Oscilação

De acordo com Titze (17), os tipos de oscilação são:

- Oscilação Natural: Quando um sistema que se encaixa nos critérios anteriores se move sem interferência após um distúrbio inicial.
- Oscilação Natural: Quando um sistema que se encaixa nos critérios anteriores se move sem interferência após um distúrbio inicial.
- Oscilação Forçada: Requer uma fonte externa de condução que por si só é um oscilador. Dita grande parte do padrão de vibração do sistema.
- Oscilação Auto-Sustentável: Requer uma fonte de energia estável e uma interação não-linear entre os componentes internos ao sistema. As perdas de energia são compensadas, mantendo o padrão oscilatório.

2.4 Sistema Auditivo

2.4.1 Introdução

O Sistema auditivo consiste em componentes periféricos e centrais.Atualmente a maior parte do conhecimento do funcionamento dos sistemas auditivos deriva de estudos de animais não humanos.(11)

Sistema auditivo diferencia-se entre espécies em jeitos interessantes. Por exemplo algumas espécies tem caracteristicas diferentes relacionadas aos sinais vocais mais utilizados por ela mesma.(11)

2.4.2 Intensidade

Como a frequência de um estímulo a intensidade dele é processado e codificado subcortical nos dois lados do cérebro em todos os níveis no cerébro.(11)

2.5 Voz e Propriedades Linguísticas

Uma divisão importante de acordo com Flanagen, são as letras separados em classificações Vogais e Consoantes que se associam a um movimento do trato vocal correspondente (9).

Place of articulation	Voiced		Voiceless	
Labio-dental	/v/	vote	/f/	for
Dental	/ð/	then	/θ/	thin
Alveolar	/z/	zoo	/s/	see
Palatal	/ʒ/	azure	/ʃ/	she
Glottal			/h/	he

2.5.1 Vogais

O trato vocal ao produzir uma vocal, em um articação normal, mantém-se relativamente estável.Há uma opção de contribuição das cavidades nasais, uma cobertura porém é negligenciável.Baseado nessas caracteristicas é divididos todas as consonantes. A tabela abaixo explica:

2.5.2 Consoantes

Sons produzidos com constrições em algum ponto no trato vocal. Dividido em quatro (4) classes, baseados em duas funcionalidades binárias, sonorant e continuant.

Sonorant

Sonorant pode ser traduzido como "cantado". Consoantes Sonorant são sons que não aumentam a pressão do ar dentro do trato vocal pois a constrição não é muito justa ou o palato continua aberto, deixando ar escapar por ele.

Continuant

Uma consoante discontinuant é produzida por um fechamento completo em algum ponto no trato vocal.

Capítulo 3

Síntese de Voz

3.0.1 Modelo Massa-Mola Auto Sustentável

: A fechadura e abertura da glote num sistema massa mola de apenas um lado

$$P = (1 - \frac{a2}{a1}) * (Ps - Pi) + Pi$$

Versão simplificada da pressão massa mola

- P: Pressão resultante NA GLOTE
- a1: Areas de entrada da glote
- a2: Areas de saída da glote
- Ps: Pressão subglotica
- Pi: Pressão sobre o trato vocal(Pressão input)

No modelo mono massa a1 = a2.No caso em que Pressão na GLOTE,P, seja igual a pressão supraglotal, indica que a inércia da coluna de ar acima da glote altera a pressão.

3.0.2 Synpath

O SynPath é um sintetizador computacional, desenvolvido em linguagem Python [13, 20, 28], criado por Lucero (**?**). Este software é uma extensão do sintetizador concebido por Fraj[19], incorporando um modelo de vibração para as pregas vocais. O seu propósito é aumentar a delidade siológica do sintetizador de Fraj e permitir o controle direto dos sons sintetizados em termos de parametrização da laringe. Para se obter um simples controle sobre o sintetizador e facilitar o seu uso para aplicações práticas, é necessário que a representação das pregas vocais seja simples. Além disso, o modelo das pregas vocais deve garantir variações suaves no uxo gerado na glote. A falta de suavidade gera timbres não naturais e consequentemente a perda da delidade siológica buscada. Sendo assim, o modelo multi-massa para representação das pregas vocais não pode ser utilizado visto que produz variações não suaves e é um modelo matematicamente muito complexo levando a instabilidades numéricas, afetando o uso para aplicações práticas. O SynPath tomou como base para a representação das pregas vocais o modelo de onda mucosa desenvolvido por Titze [26]. Basicamente, o modelo é um oscilador mono-massa, conforme descrito no

Capitulo 3, incorporando a transferência de energia do uxo de ar para as pregas vocais. Entretanto, o modelo de Titze possui duas restrições, uma que foi solucionada e aplicada no desenvolvimento do SynPath e a outra que ainda não foi solucionada e portanto é também uma restrição do modelo computacional do SynPath. A primeira restrição é que o modelo de Titze foi concebido para o estudo em pequenas oscilações, sendo que para oscilações de maior amplitude, este não é apropriado. Entretanto, esse modelo foi extendido por Lucero [11] para abranger maiores oscilações utilizando um mecanismo limitador de amplitude durante as oscilações [14]. Mesmo com essa extensão para maiores amplitudes, o modelo ainda apresentava uma outra restrição, um atraso pequeno para o deslocamento da onda no canal glotal. A consequência desse atraso é que a pressão limite para que ocorra a vibração se torna independente da frequência de vibração [15]. Porém, sabe-se que um esforço maior é necessário para que tons mais agudos sejam vocalizados, ou seja, a pressão para que se ocorra vibração em frequências maiores(tons maiores) é maior. Essa restrição ainda não foi solucionada tendo em vista a diculdade de se realizar o supracitado computacionalmente. Mesmo com essa restrição, o modelo de Lucero [11] é o modelo utilizado para 26 a representação do caráter oscilatório das pregas vocais no software SynPath, sendo que esta restrição não solucionada não tem forte inuência no produto nal.

Requisitos Funcionais

O Synpath é consistido também dos seguintes requisitos funcionais, os requistos funcionais são as funcionalidades que o sistema executará(4)

- 1 - Validação do Parâmetros passados pelo Usuário, se condizem com restrições do programa.
- 2 - Plotar um gráco inicial do trato vocal de acordo com os parâmetros do usuário.
- 3 - Plotar três grácos referentes as propriedades da voz simuladas com os parâmetros fornecidos pelo usuário. – O primeiro gráco refere-se às posições adotadas pelas cordas vocais, a área da glótis, ao uxo de ar nessa área e às características desse uxo. – O segundo gráco refere-se às características do som gerado pela simulação física do aparato fonador pelo programa. – O terceiro gráco refere-se ao espectro de frequência do som gerado e do uxo da glótis.
- 4 - Gerar um arquivo de texo com as características de voz gerada, frequencia,amplitude, ruído da voz, entre outros
- 5 Gerar um arquivo de som de voz simulada.

Requisito Não-Funcionais

O Synpath consiste também dos seguintes requisitos não funcionais, requisitos não funcionais são requisitos são parametros de qualidade, requisitos que limitam as funcionalidades do sistema(4).

- 1 - O Sistema deve produzir os gráficos que os requisitos funcionais delimitaram em um intervalo de 1(um) minuto.

- 2 - Após gerar os gráficos e os exibi-los o arquivo texto e o arquivo de audio deverão ser exibidos
- 3 - Para que o sistema esteja funcional é necessário ter instalado os pacotes: MatPlotLib e NumPY
- 4 - O sistema deve ser executado em plataformas de um sistema operacionais como Windows, Linux ou MacOs, Versões recentes de acordo com a data desse documento.

3.0.3 HMMs

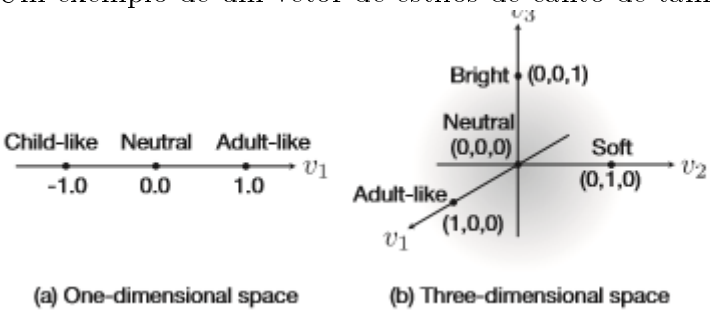
Minera-se de várias partituras musicais para treino. Os dados minerados dessas músicas são fonemas, altura, intensidade e os intervalos, isto é relação com outras notas.Esses dados são convertidos e mapeados em "labels"dependentes de contexto (3).Após isso as HMM's são treinados através dos dados de treinamento usando o algoritmo EM.(8). Após isso ocorre a fase de sintesis, usa-se outra partitura para ser convertida em "labels"dependentes de contexto e estima-se quais "labels"preprocessadas são correspondentes.(13)

3.0.4 MHRSMM

Uma variação HSMM. Modelo de multipla regressão HSMM. Parametros importantes são i e mi dos outputs pdfs

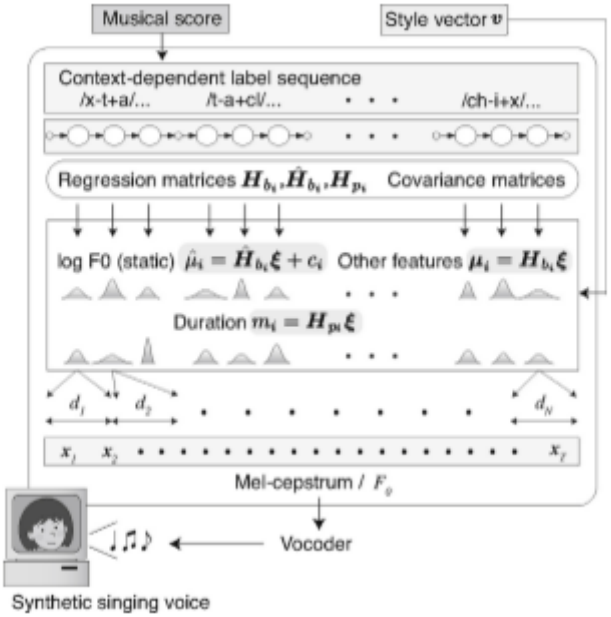
$$\mu_i = H_b i\xi$$
$$m_i = H_p i\xi$$
$$\xi = [1, v_1, v_2, ..., v_L]^T$$
$$\xi = [1, v^T]^T$$

Onde L é a dimensão do vetor de estilo e vi é a intensidade do enésimo estilo de canto. Um exemplo de um vetor de estilos de canto de tamanho L = 2 e L = 3.



(a) One-dimensional space (b) Three-dimensional space

Controle do Sintentisador de voz cantada baseado em MRHSMM



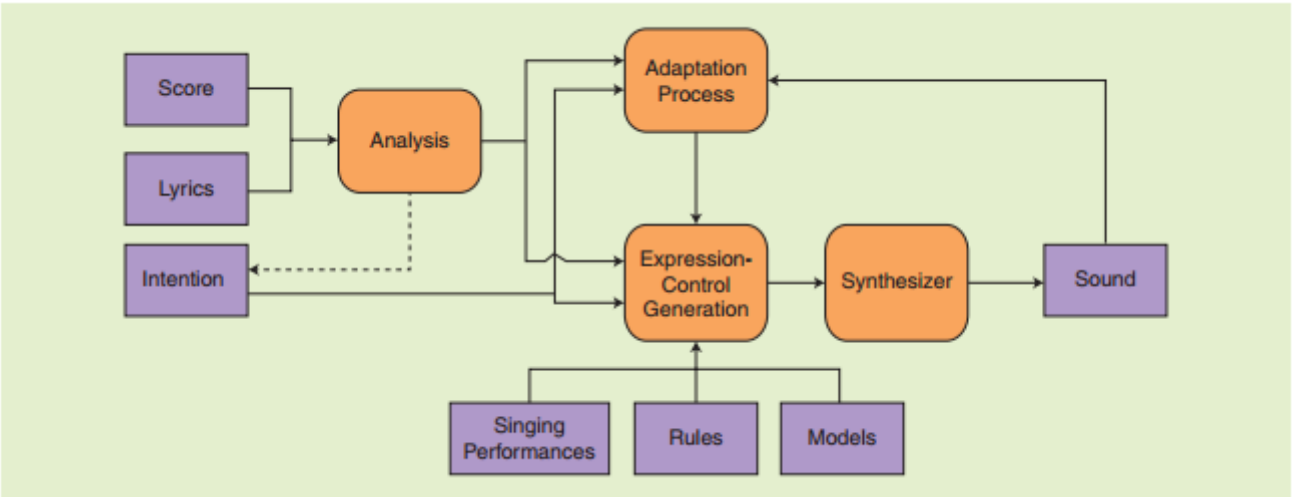
Durante a fase de síntese o usuário do programa adiciona vetores de estilos de acorodo com a intenção e a expressividade pretendida. Parametros de output como duração são gerados pelos vetores de estilos dados e matrizes de regressões treinadas usando MRHSMMs

Resultado de todo esse processo é um sequência HSMM usando parametros de geração de fala

MRHSMM possui uma dificuldade de gerar contorno F0 que acompanhe o contexto de mudança de altura das notas o author TAKASHI NOSE, propõe um treinamento de HSMM e HMM nos parametros

3.0.5 FrameWorks Sintetisador de Voz

Frame Work de um Sistema Sintetisador de Voz:



Input

Consiste da partitura, letra e emoção. o input é analisado e derivado em uma transcrição fonética, alinhamento com a performance alvo ou dados contextuais.(18)

Expressão

Expressão músical é um conceito intuitivo porém dificil de se definir. A expressão é chave na percepção da qualidade e naturalidade musical. No caso da voz cantada implica-se usar vários outros parametro além de frequencia e amplitude. Psicológicamente contorno do timbre, vibrato, tremolo, timing fonético.(18)

3.0.6 Envoltoria F0

Envoltórias F0 são usadas para expressar informação linguistica, para-linguística e não-linguistica.(14)

As Envoltória F0 apresentam três (3) características importantes que fazem diferenciar uma voz falada a uma voz cantada.(15)

- 1 - O alcance dinâmico de uma envoltória F0 é mais largo que o de uma voz falada
- 2 - A envoltória F0 corresponde e tende a se manter estável em uma nota. A mudança de nota de uma envoltória F0 corresponde a melodia da música
- 3 - Existem muita flutuações f0 que são apenas observadas em apenas vozes cantadas

3.0.7 Sintese de Voz em Mandarím

Utiliza-se a técnica HNM para a sintese da voz cantada em mandarím. HNM significa , "harmonic plus noise model". O modelo HNM divide o espectro de um sinal em dois(2) com larguras não iguais para modelagem melhor do espectro.(12)

Referências

[1] Ferdinand de Saussure. *Self-oscillating source for vocal-tract synthesis.* IEEE Tran. Audio Eletroacoust, Audio Eletroacoust., 1968. 7

[2] Elias Amadeu de Souza. Simulação computacional de uma fenda glotal, 2014. 8

[3] Gilles Degottex, Pierre Lanchantin, Axel Roebel, and Xavier Rodet. Mixed source model and its adapted vocal tract filter estimate for voice transformation and synthesis. *Speech Communication*, 55(2):278 – 294, 2013. 17

[4] Pierre Bourque & Richard E.(Dick) Farlay. *Guide to the Software Engineering Body of Knowledge Version 3.0.* IEEE Computer Society, 2004. 16

[5] Jean Schoentgen Samia Fraj and Francis Grenez. *Development and perceptual assessment of a synthesizer of disordered voices.* Journal of Acoustical Society of America, 2012.

[6] Antoine Giovanni, Maurice Ouaknine, and Renaud Garrel. Fisiología de la fonación. *{EMC} - Otorrinolaringología*, 33(1):1 – 17, 2004. 3

[7] Henry Gray. *Anatomy of the Human Body.* IEEE Tran. Audio Eletroacoust. 7

[8] Jon Gudnason, Mark R.P. Thomas, Daniel P.W. Ellis, and Patrick A. Naylor. Data-driven voice source waveform analysis and synthesis. *Speech Communication*, 54(2):199 – 211, 2012. 17

[9] James L. Flanagan Jont B. Allen Mark A. Hasegawa-Johnson. *Speech Analysis Synthesis and Perception.* 2008. 14

[10] Matias Zanartu Kelley C. Stewart Michael W. Plesniak David E. Sommer Sean D. Peterson Byron D. Erath. *A review of lumped-element models of voiced speech.* Speech Communication. 7

[11] Jody Kreiman and Diana Van Lancker Sidtis. Foundations of voice studies, 2011. 2, 13, 14

[12] Chyi-Yeu Lin, Li-Chieh Cheng, Chang-Kuo Tseng, Hung-Yan Gu, Kuo-Liang Chung, Chin-Shyurng Fahn, Kai-Jay Lu, and Chih-Cheng Chang. A face robot for autonomous simplified musical notation reading and singing. *Robotics and Autonomous Systems*, 59(11):943 – 953, 2011. 19

[13] Takashi Nose, Misa Kanemoto, Tomoki Koriyama, and Takao Kobayashi. Hmm-based expressive singing voice synthesis with singing style control and robust pitch modeling. *Computer Speech Language*, 34(1):308 – 322, 2015. 17

[14] Takeshi Saitou, Masashi Unoki, and Masato Akagi. Development of an {F0} control model based on {F0} dynamic characteristics for singing-voice synthesis. *Speech Communication*, 46(3–4):405 – 417, 2005. Quantitative Prosody Modelling for Natural Speech Description and GenerationInternational Conference on Speech Prosody. 19

[15] Takeshi Saitou, Masashi Unoki, and Masato Akagi. Development of an {F0} control model based on {F0} dynamic characteristics for singing-voice synthesis. *Speech Communication*, 46(3–4):405 – 417, 2005. Quantitative Prosody Modelling for Natural Speech Description and GenerationInternational Conference on Speech Prosody. 19

[16] BradH Story. *Tubetalker*. Dept. of Speech, Language, and Hearing, Tucson, AZ, 2010. 13

[17] Ingo Titze. *Principles of Voice Production*. Prentice Hall, New Jersey 07632, 1994. vi, 5, 6, 7, 9, 10, 11, 12, 13

[18] M. Umbert, J. Bonada, M. Goto, T. Nakano, and J. Sundberg. Expression control in singing voice synthesis: Features, approaches, evaluation, and challenges. *IEEE Signal Processing Magazine*, 32(6):55–73, Nov 2015. 19