**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.
49 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

# Lista de Figuras

# Lista de Tabelas

# Capítulo 1

# Introdução

## 1.1 Motivação

# Capítulo 2

# Conceitos Básicos de Sintese de Voz

## 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 (7). 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 A Anatomia

A Figura 2.2 (7), 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.

### 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.

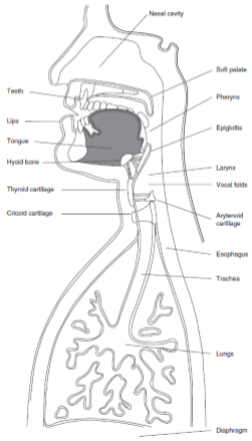


Figura 2.1: Aparelho Fonador

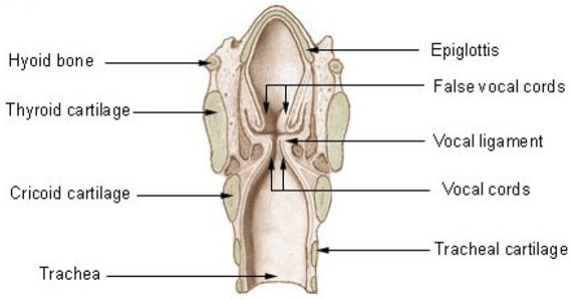



Figura 2.2: : Secção coronal da laringe e parte superior da traquéia. (7)

### Cartilagens da Laringe

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.

Os músculos na laringe podem ser divididos em dois grupos, os intrínsecos e os extrínsecos [1]. Os músculos intrínsecos interconectam as cartilagens da laringe, ao passo que, 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 2.3, 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.

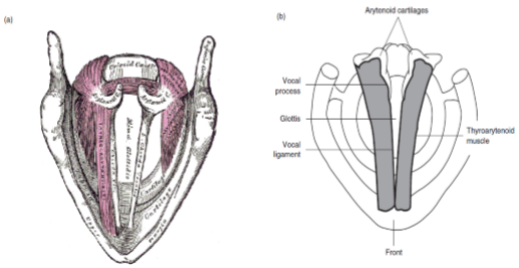



Figura 2.3: : Músculos Intrínsecos da Laringe. (1)

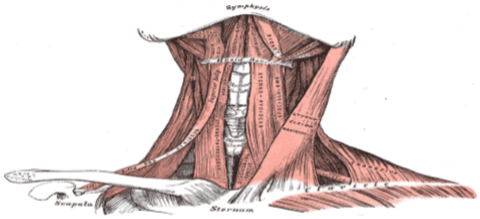


Figura 2.4: :Músculos Extrínsecos da Laringe. (1)

### 2.1.3 Pregas Vocais

As pregas vocais, como dito anteriormente, estão localizadas dentro da laringe, mais especicamente 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 2.5 mostra em mais detalhes a anatomia das pregas vocais, os componentes musculares e as cartilagens atuantes.

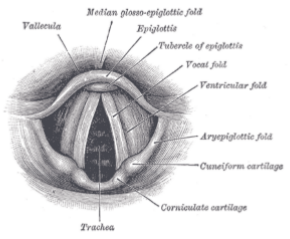



Figura 2.5: :Cordas Vocais e Componente (1)

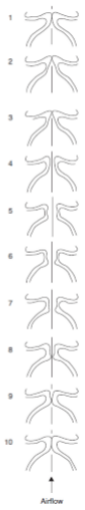


Figura 2.6: :Movimento das Cordas Vocais (7)

### 2.1.4 Overview: Geração de Som e Ressonadores

A produção do som da voz é composta de duas etapas importantes que ocorrem na região subglotal/glotal e supraglotal. A primeira é a transformação da energia aerodinâmica em energia sonora, pelo movimento e vibração das pregas vocais.

O segundo é a transformação do som primitivo gerado em voz, através da atuação dos ressonadores e formantes na região supraglotal.

A vibração das pregas vocais é extremamente complexa e, diversos músculos em união com a pressão exercida pelo ar atuam para tornar esse movimento possível. De maneira sintetizada, as pregas vocais vibram do topo ao fundo de maneira que não vibram como se fosse um bloco, mas sim de forma ondulatória, conforme mostrada na Figura 2.6. Essa vibração é responsável por mudanças depressão entre a região subglotal e a região glotal, causando uma diferença de pressão(condensação e rarefação) nessa área durante o movimento das pregas vocais, o que ocasiona a geração de som.

Entretanto, o som produzido é um som primitivo, conforme dito anteriormente, e para se transformar na voz característica humana, ele deve ser ltrado e moldado pelos ressonadores(formantes) no trato vocal(Figura 2.7). Todas as cavidades mostradas no trato vocal atuam como ressonadores para a onda sonora produzida pelas pregas vocais. Um ressonador pode entrar em estado de vibração através de uma força aplicada ao mesmo em inércia ou por interação com algo que já esteja em estado de vibração.

Neste segundo caso, as vibrações produzidas pelo ressonador serão amplicações ou atenuações dependendo de quão próximas ou distantes, em termos de frequência, são as vibrações da onda sonora em contato com o ressonador. Caso a onda sonora possua vibrações cujas frequências se assemelhem às frequências do ressonador,estas então serão amplicadas pelo ressonador.

Entretanto,caso as frequências do som gerado vibrem em uma frequência distante da frequência natural do ressonador, então estas serão abafadas. A voz passa por esse processo ao ser formada. Um som primitivo advindo das pregas vocais entra em contato com os ressonadores no trato vocal, estes por sua vez em conjunto

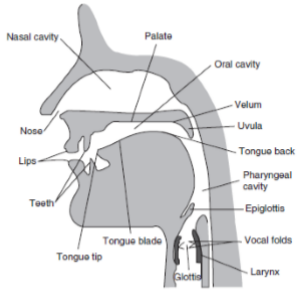



Figura 2.7: :Trato Vocal (7)

Os conceitos e propriedades descritos acima são extremamente importantes para se entender a manutenção da produção do som. 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 bras 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 (13) 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 (13) 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 Figura **??** retirada do estudo feito por Titze (13) demonstra uma curva hipotética de 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.

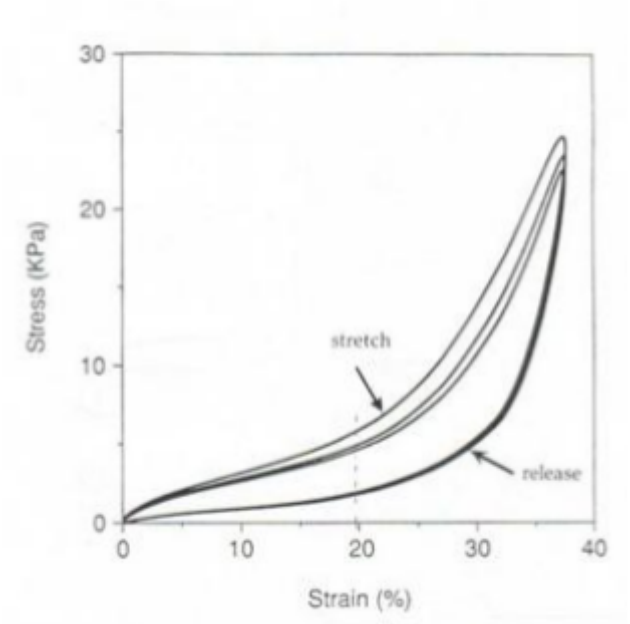



Figura 2.8: :Curva Hipotética Tensão-Deformação das Cordas vocais Humanas (13)

## 2.2 Propriedades Físicas

### 2.2.1 Lei Bernouli

Energia potencial e energica cinética em fluídos se mantém a mesma porém em proporções diferentes (12):

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

### 2.2.2 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.. . .

### 2.2.3 Tipos de Oscilação

De acordo com Titze (13), 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.

### 2.2.4 Tensao

Conceito de Tensão For por unidade de ar

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

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

### 2.2.5 Curva Força e Alongamento

Utiliza-se para não ser dependente da geometria do material. Utilizamos nas cordas vocais(?) por serem materias biológicos. Cria uma figura ilustrando comportamento da deformação das pregas vocais

### 2.2.6 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: **?**

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

### 2.2.7 Reflexao de Som

Um fenômeno ligado a rigidez e amortecimento entre um meio e outro.(6) 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.(4)

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 (2). A propagação é sempre em direção oposta à fonte, no caso é na direção contrária à coluna de ar(meio 1).

# Capítulo 3

# Fundamentos Básicos para a Produção da Voz

Neste capitulo serão introduzidos princípios físicos e biológicos da produção da voz. Além dos conceitos introduzidos aqui, serão analisados os processos mecânicos da geração da voz por parte da laringe em conjunto com o trato vocal pois são de extrema importância para a concepção dos modelos computacionais e matemáticos utilizados para se desenvolver o sintetizador.

### 3.0.1 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 (13). Outra analogia interessante é sobre o instrumento e quem o utiliza. Um bom pianista por exemplo, sua musica é boa porque ele é habilidoso como 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 inuencia 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 (13).

### 3.0.2 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 uidos. 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:

### Tensão e Deformação

São características de forças atuantes em superfícies, como por exemplo a resistência do ar.

### Tensão

é quantidade de força por unidade de área (13), podemos escrever na forma da equação 3.0.2 :

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

(3.1)Onde é a tensão, f é a força aplicada e A a área de aplicação desta força.

### Deformação

é a medida de deformação de um meio após a aplicação de uma tensão (13) e pode ser escrito na forma da equação :

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

(3.2)

Onde 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 uido quando atuam forças de tensão no mesmo. Matematicamente pode ser expresso conforme a equação 3.0.2:

$$\epsilon = \eta \frac{d_e}{d_t}$$

(3.3)

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

### 3.0.3 Modelo Massa-Mola Auto Sustentável

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

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

### 3.0.4 Elasticidade

É uma propriedade do meio que determina quão completa será a restauração do meio após uma dada deformação.

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.

### 3.0.5 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(**?** )

- 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(**?** ).

- 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.

### 3.0.6 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.(5). 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.(9)

### 3.0.7 MHRSMM

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

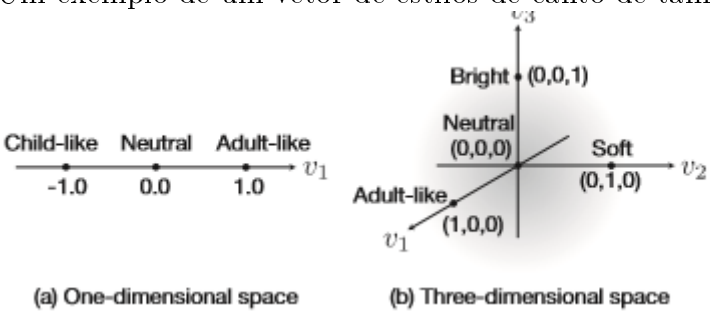
$$\mu_i = H_b i\xi$$

$$m_i = H_p i\xi$$

$$\xi = [1, v_1, v_2, ..., v_L]^T$$

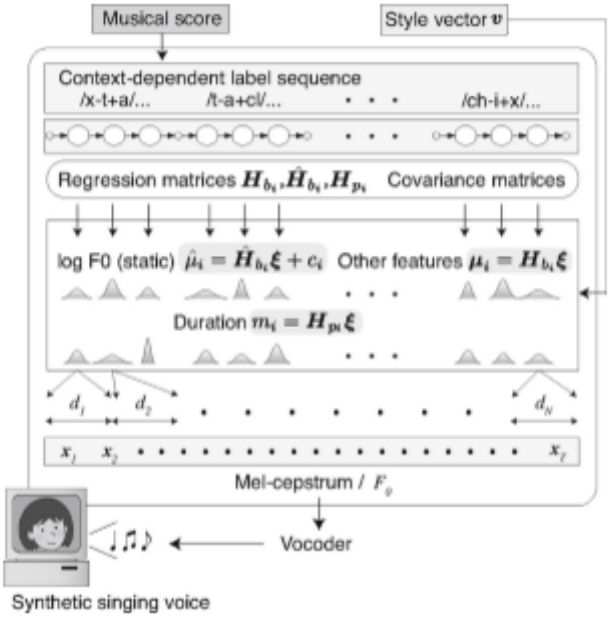
$$\xi = [1, v^T]^T$$

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**



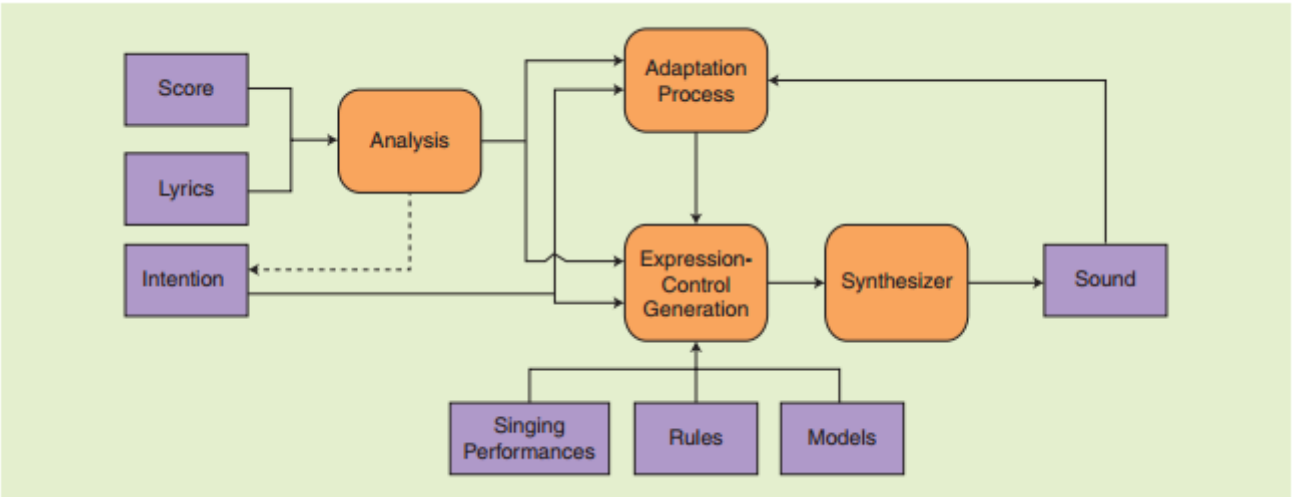

Durante a fase de síntese o usuário do programa adiciona vetores de estilos de acorodo com a intenção e a expressividade pretendida. Parametros de output como duração são gerados pelos vetores de estilos dados e matrizes de regressões treinadas usando MRHSMMs

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

### 3.0.8 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.(14)

**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.(14)

### 3.0.9 Envoltoria F0

Envoltórias F0 são usadas para expressar informação linguistica, para-linguística e não-linguistica.(10)

As Envoltória F0 apresentam três (3) características importantes que fazem diferenciar uma voz falada a uma voz cantada.(11)

- 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

### 3.0.10 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.(8)

# Referências


[1] John & Charles Bell. The anatomy and physiology of the human body, 1826. [Down, pre-B, ED]. vi, 4

[2] Ferdinand de Saussure. *Self-oscillating source for vocal-tract synthesis.* IEEE Tran. Audio Eletroacoust, Audio Eletroacoust., 1968. 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. 13

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

[5] 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. 13

[6] 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

[7] Jody Kreiman and Diana Van Lancker Sidtis. Foundations of voice studies, 2011. vi, 2, 3, 5, 6

[8] 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. 15

[9] 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. 13

[10] 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. 15

[11] 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. 15

[12] BradH Story. *Tubetalker*. Dept. of Speech, Language, and Hearing, Tucson, AZ, 2010. 6

[13] Ingo Titze. *Principles of Voice Production*. Prentice Hall, New Jersey 07632, 1994. 6, 9, 10

[14] 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. 15