



Universidade de Brasília

**Instituto de Ciências Exatas
Departamento de Ciência da Computação**

Síntese 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.

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

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

1. Síntese, 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 Prof. Dr. Professor II
CIC/UnB 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
2.3	Propriedades Físicas	5
2.3.1	Lei Bernouli	5
2.3.2	Critérios para Oscilação	6
2.3.3	Tipos de Oscilação	6
2.3.4	Tensao	6
2.3.5	Curva Força e Alongamento	6
2.3.6	Viscosidade	7
2.3.7	Reflexao de Som	7
2.3.8	Fluxo de Ar na Glote	7
2.4	Sistema Auditivo	8
2.4.1	Introdução	8
2.4.2	Intensidade	8
2.5	Voz e Propriedades Linguísticas	8
2.5.1	Vogais	9
2.5.2	Consoantes	9
3	Síntese de Voz	10
3.0.1	Modelo Massa-Mola Auto Sustentável	10
3.0.2	Synpath	10
3.0.3	HMMs	11
3.0.4	MHRSM	11
3.0.5	FrameWorks Sintetizador de Voz	12
3.0.6	Envoltoria F0	13
3.0.7	Síntese de Voz em Mandarim	13
	Referências	14

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

Lista de Tabelas

Capítulo 1

Introdução

1.1 Motivação

A UnB possui atuação na área de simulação de voz a ser considerada. Essa atuação considero ser principalmente ser por esforço e trabalho do Doutor Jorge Carlos Lúcerio 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 influenciado 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 deglutição. 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.

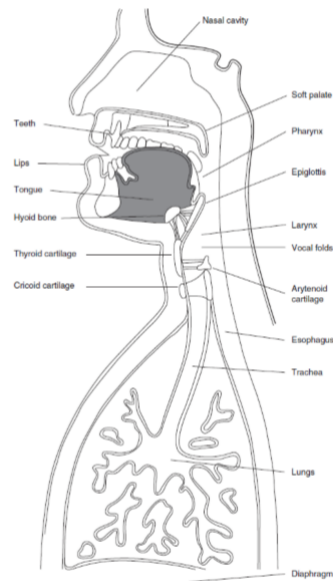


Figura 2.1: Aparelho Fonador

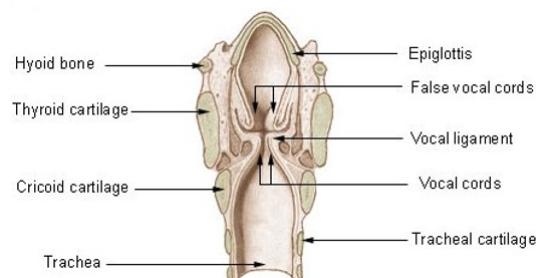


Figura 2.2: Secção coronal da laringe e parte superior da traquéia

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 extrínsecos⁶. Os músculos intrínsecos interconectam as cartilagens da laringe, ao passo que,

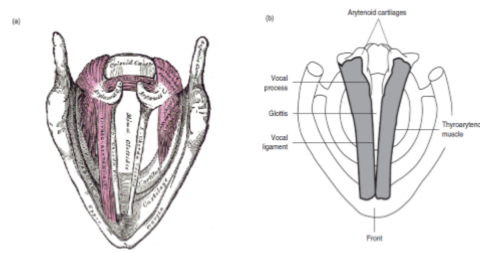


Figura 2.3: Músculos Intrínsecos da Laringe

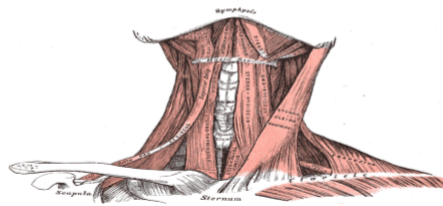


Figura 2.4: Músculos Extrínsecos da Laringe

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 influência direta em algumas características da voz. Por exemplo, o músculo cricoti-roideo é o músculo primário utilizado no controle do tom da voz. Por sua vez, o músculo cricoaritenóideo posterior atua na abdução das pregas vocais, ao passo que o músculo interaritenóideo 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 especificamente na parte superior da traqueia. Elas estão posteriormente ligadas às cartilagens aritenóides, e anteriormente ligadas à cartilagem tireóide. 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.

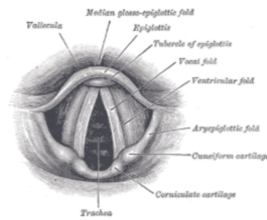


Figura 2.5: Cordas Vocais e Componentes

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 análise, 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 conseqüentemente, se contrapõe à mecânica de partículas. Dentro da mecânica de meios contínuos, mais especificamente, 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:

2.3 Propriedades Físicas

2.3.1 Lei Bernouli

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

$$P + \frac{\rho * v^2}{2} = \text{Constante Sendo : } \rho = \text{Densidade do fluido } P = \text{Pressão do fluido onde o fluido se encontra } v = v$$

2.3.2 Critérios para Oscilação

Alguns critérios devem ser atendidos para que um determinado 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.

2.3.4 Tensão

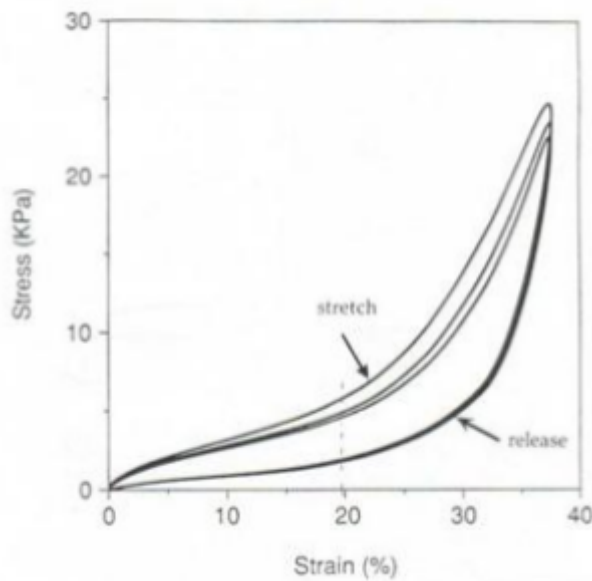
Conceito de Tensão For por unidade de ar

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

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

2.3.5 Curva Força e Alongamento

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



2.3.6 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: ?

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

2.3.7 Reflexao 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 partículas 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 partículas 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).

2.3.8 Fluxo de Ar na Glote

Fluxo de ar na Glote:

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

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

$A^* =$

$\rho =$ Densidade do Ar

$c =$ Velocidade do som

$P_s e P_i =$ Pressão de incidência na entrada e saída da glote

Uma vez descoberto o fluxo, as pressões de reflexão $P_e^s P_i^+$ podem ser encontradas com a seguinte equação :

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

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

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

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 características 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 cérebro.(11)

2.5 Voz e Propriedades Linguísticas

Uma divisão importante de acordo com Flanagan, 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 uma articulação normal, mantém-se relativamente estável. Há uma opção de contribuição das cavidades nasais, uma cobertura porém é negligenciável. Baseado nessas características é divididas todas as consoantes. A tabela abaixo explica:

2.5.2 Consoantes

Sons produzidos com constrictões 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 constrictão 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 subglótica
- 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

Sintetizador computacional, criado por Lucero. Sintetizador concebido por Fraj (5) Foi escolhido não usar o modelo multi-massa por variações não suaves e é tematicamente muito complexo apesar instabilidades numéricas. Synpath utiliza como base de representação das pregas vocais o modelo mono massa. O trato vocal é representado por uma concatenação de tubos cilíndricos os quais se propagam uma onda acústica. Superglótica não leva em conta o trato vocal. Perdas e viscosidade são considerada.

Requisitos Funcionais

O Synpath é consistido também dos seguintes requisitos funcionais, os requisitos 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 gráfico inicial do trato vocal de acordo com os parâmetros do usuário.
- 3 - Plotar três gráficos referentes as propriedades da voz simuladas com os parâmetros fornecidos pelo usuário. – O primeiro gráfico 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 gráfico refere-se às características do som gerado pela simulação física do aparato fonador pelo programa. – O terceiro gráfico refere-se ao espectro de frequência do som gerado e do uxo da glótis.
- 4 - Gerar um arquivo de texto com as características de voz gerada, frequência, 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: Mat-PlotLib 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 síntese, usa-se outra partitura para ser convertida em "labels" dependentes de contexto e estima-se quais "labels" pré-processadas são correspondentes.(13)

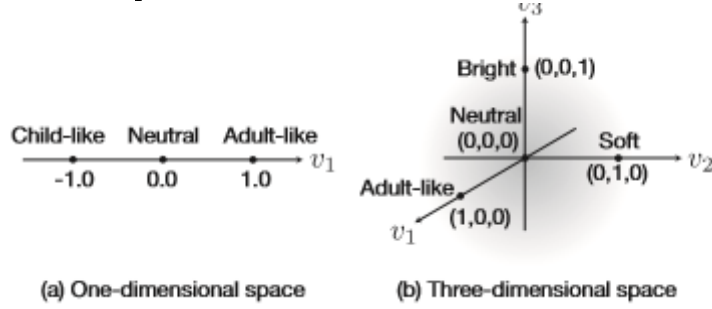
3.0.4 MHRSMM

Uma variação HSMM. Modelo de múltipla regressão HSMM. Parametros importantes são i e m_i dos outputs pdfs

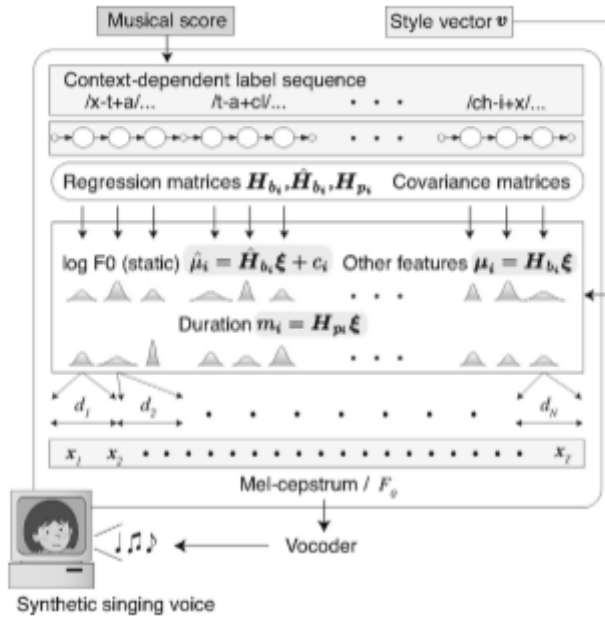
$$\begin{aligned}\mu_i &= H_{bi}\xi \\ m_i &= H_{pi}\xi \\ \xi &= [1, v_1, v_2, \dots, v_L]^T\end{aligned}$$

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

Onde L é a dimensão do vetor de estilo e v_i é a intensidade do i -ésimo estilo de canto. Um exemplo de um vetor de estilos de canto de tamanho $L = 2$ e $L = 3$.



Controle do Sintetizador de voz cantada baseado em MRHSMM



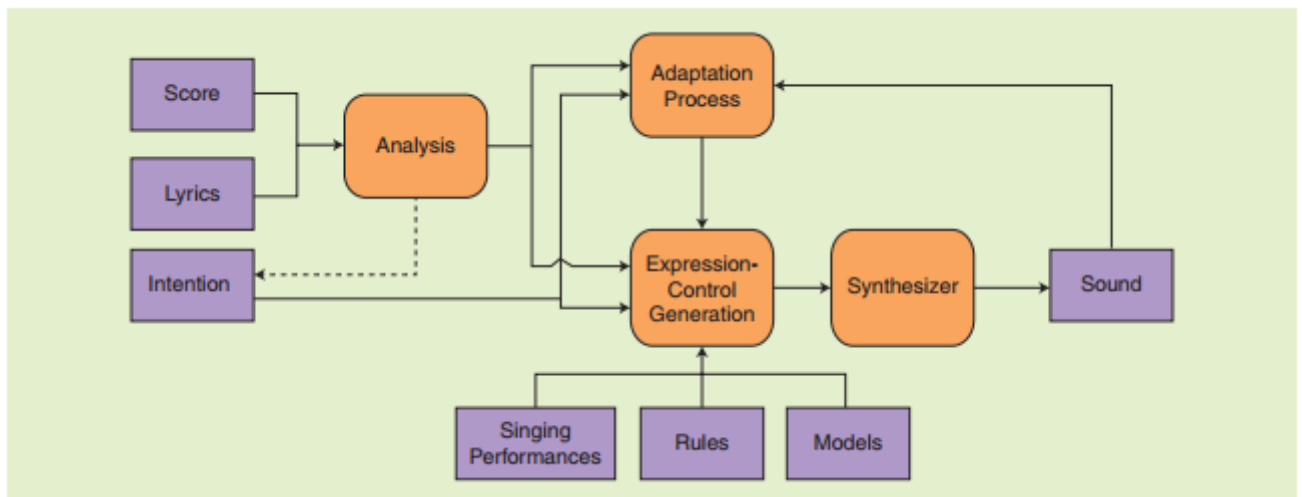
Durante a fase de síntese o usuário do programa adiciona vetores de estilos de acorodo com a intenção e a expressividade pretendida. Parametros de output como duração são gerados pelos vetores de estilos dados e matrizes de regressões treinadas usando MRHSMMs

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 Sintetizador de Voz

Frame Work de um Sistema Sintetizador 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 musical é um conceito intuitivo porém difícil de se definir. A expressão é chave na percepção da qualidade e naturalidade musical. No caso da voz cantada implica-se usar vários outros parâmetros além de frequência e amplitude. Psicologicamente contornos do timbre, vibrato, tremolo, timing fonético.(18)

3.0.6 Envoltória F0

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

As Envoltórias F0 apresentam três (3) características importantes que fazem diferenciar uma voz falada de 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 muitas flutuações f0 que são apenas observadas em apenas vozes cantadas

3.0.7 Síntese de Voz em Mandarim

Utiliza-se a técnica HNM para a síntese da voz cantada em mandarim. 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. 7
- [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. 11
- [4] Pierre Bourque & Richard E.(Dick) Farlay. *Guide to the Software Engineering Body of Knowledge Version 3.0*. IEEE Computer Society, 2004. 10, 11
- [5] Jean Schoentgen Samia Fraj and Francis Grenez. *Development and perceptual assesment of a synthesizer of disordered voices*. Journal of Acoustical Society of America, 2012. 10
- [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. 11
- [9] James L. Flanagan Jont B. Allen Mark A. Hasegawa-Johnson. *Speech Analysis Synthesis and Perception*. 2008. 8
- [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, 8
- [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. 13

- [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. 11
- [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 Generation International Conference on Speech Prosody. 13
- [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 Generation International Conference on Speech Prosody. 13
- [16] BradH Story. *Tubetalker*. Dept. of Speech, Language, and Hearing, Tucson, AZ, 2010. 5
- [17] Ingo Titze. *Principles of Voice Production*. Prentice Hall, New Jersey 07632, 1994. 5, 6
- [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. 13