



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.
53 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	Conceitos Básicos de Síntese de Voz	2
2.1	Anatomia da Voz	2
2.1.1	A Anatomia	2
2.1.2	Músculos e Cartilagens	2
2.1.3	Orgãos	4
2.1.4	Aparelho Fonador	4
2.2	Propriedades Físicas	5
2.2.1	Lei Bernouli	5
2.2.2	Crítérios para Oscilação	5
2.2.3	Tipos de Oscilação	5
2.2.4	Tensao	5
2.2.5	Curva Força e Alongamento	6
2.2.6	Viscosidade	6
2.2.7	Reflexao de Som	6
2.3	Sintetizadores de Voz	6
2.3.1	Modelo Massa-Mola Auto Sustentável	6
2.3.2	Synpath	7
2.3.3	HMMs	7
2.3.4	MHRMM	7
2.3.5	FrameWorks Sintetisador de Voz	8
2.3.6	Envoltoria F0	9
2.3.7	Síntese de Voz em Mandarim	9
2.4	Sistema Auditivo	10
2.4.1	Introdução	10
2.4.2	Intensidade	10
2.5	Voz e Propriedades Linguísticas	10
2.5.1	Vogais	10
2.5.2	Consoantes	10
3	Síntese de Voz	12
3.0.1	Modelo Massa-Mola Auto Sustentável	12
3.0.2	Synpath	12
3.0.3	HMMs	14

3.0.4	MHRMM	14
3.0.5	FrameWorks Sintetizador de Voz	15
3.0.6	Envoltoria F0	16
3.0.7	Sintese de Voz em Mandarim	16
Referências		17

Lista de Figuras

2.1	Aparelho Fonador	3
2.2	: Secção coronal da laringe e parte superior da traquéia. (9)	3
2.3	: Músculos Intrínsecos da Laringe. (1)	4
2.4	:Músculos Extrínsecos da Laringe. (1)	4

Lista de Tabelas

Capítulo 1

Introdução

1.1 Motivação

Capítulo 2

Conceitos Básicos de Síntese 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 (9). 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 A Anatomia

A Figura 2.2 (9), mostra os órgãos associados com a produção da voz. Dentro das condições normais, a voz é produzida quando um fluxo 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 fluxo de ar que por elas passam. Sendo assim, é fácil notar que há uma necessidade de mobilidade para toda a estrutura 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.

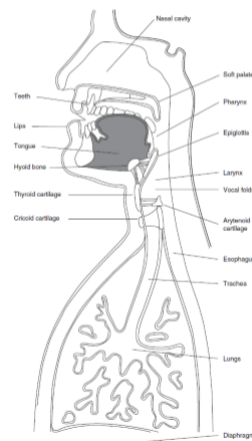


Figura 2.1: Aparelho Fonador

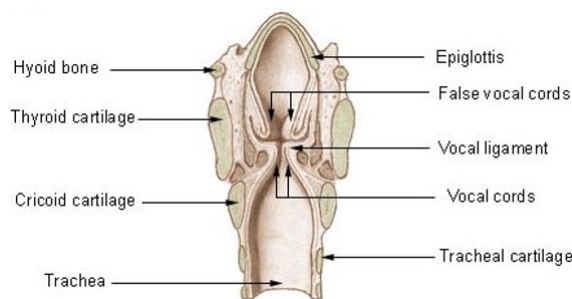


Figura 2.2: : Secção coronal da laringe e parte superior da traquéia. (9)

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 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 2.3, 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.

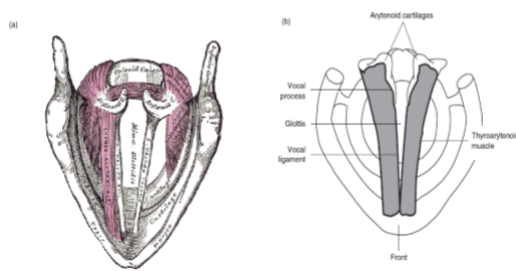


Figura 2.3: : Músculos Intrínsecos da Laringe. (1)

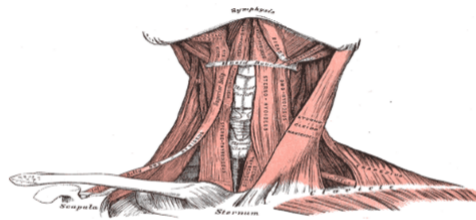


Figura 2.4: :Músculos Extrínsecos da Laringe. (1)

2.1.3 Órgãos

Órgãos envolvidos com a síntese da voz:

- Diafragma
- Pulmões
- Traqueos
- Esôfago
- laringe
- epiglote
- faringe
- palato , lingua cavidade nasal, dente lábios(8)

2.1.4 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.

2.2 Propriedades Físicas

2.2.1 Lei Bernouli

Energia potencial e energia cinética em fluídos se mantêm a mesma porém em proporções diferentes (14):

$$P + \frac{\rho * v^2}{2} = \text{Constante}$$

Sendo : ρ = Densidade do fluido P = Pressão do fluido onde o fluido se encontra v = velocidade

2.2.2 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...

2.2.3 Tipos de Oscilação

De acordo com TITZE (15), 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 Tensão

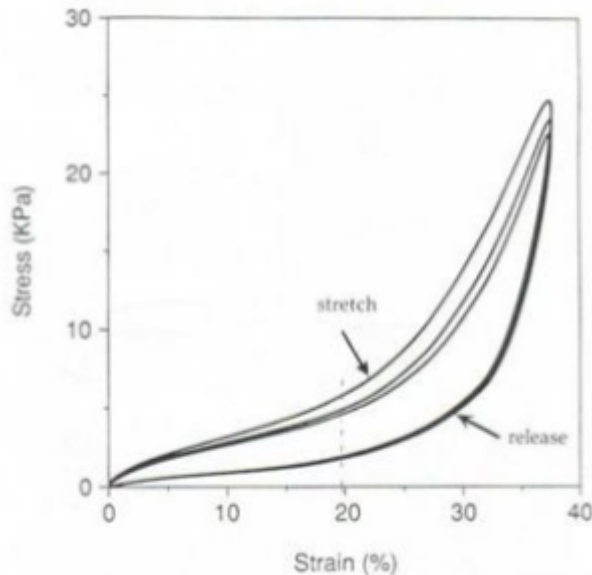
Conceito de Tensão Força por unidade de área

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

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 materiais 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 Reflexão de Som

Um fenômeno ligado a rigidez e amortecimento entre um meio e outro.(8) Ondas quando tentam penetrar em um segundo meio, sendo o segundo meio rígido, as partículas do primeiro meio se aglomeram tentando passar porém falham, seu acúmulo de partículas gera pressão que acabam criando uma outra onda no primeiro meio decorrente da primeira onda.(5)

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

2.3 Sintetizadores de Voz

2.3.1 Modelo Massa-Mola Auto Sustentável

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

$$P = (1 - \frac{a_2}{a_1}) * (P_s - P_i) + P_i$$

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 $a_1 = a_2$. 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.

2.3.2 Synpath

Sintetizador computacional, criado por Lucero. Sintetizador concebido por Fraj (4) 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 cilindricos os quais se propagam uma onda acustica. Superglótica não leva em conta o trato vocal. Perdas e viscosidade são considerada.

2.3.3 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.(6). Após isso ocorre a fase de síntesis, usa-se outra partitura para ser convertida em "labels" dependentes de contexto e estima-se quais "labels" pré-processadas são correspondentes.(11)

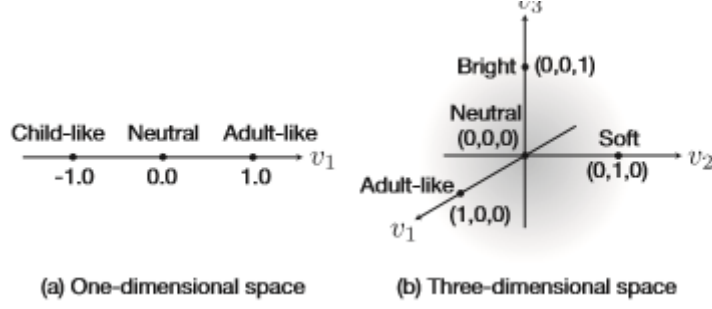
2.3.4 MHRSMM

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

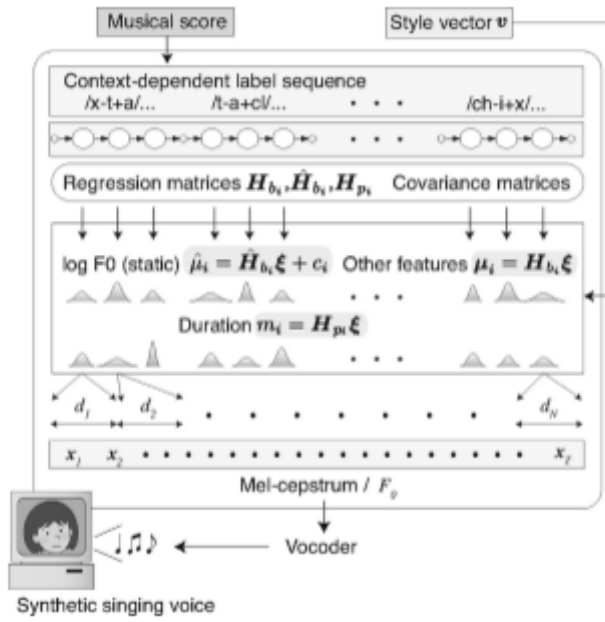
$$\begin{aligned}\mu_i &= H_{bi}\xi \\ m_i &= H_{pi}\xi \\ \xi &= [1, v_1, v_2, \dots, v_L]^T \\ \xi &= [1, v^T]^T\end{aligned}$$

Onde L é a dimensão do vetor de estilo e v_i é a intensidade do enésimo 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



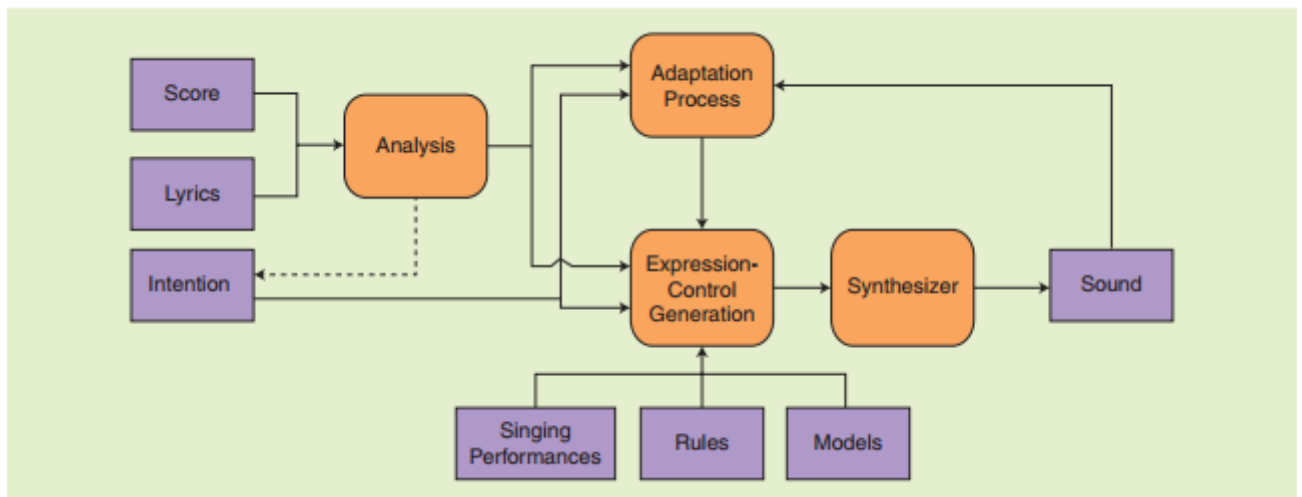
Durante a fase de síntese o usuário do programa adiciona vetores de estilos de acorodo com a intenção e a expressividade pretendida. Parametros de output como duração são gerados pelos vetores de estilos dados e matrizes de regressões treinadas usando MRHSMMs

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

2.3.5 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.(16)

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.(16)

2.3.6 Envoltória F0

Envoltórias F0 são usadas para expressar informação linguística, para-linguística e não-linguística.(12)

As Envoltórias F0 apresentam três (3) características importantes que fazem diferenciar uma voz falada de uma voz cantada.(13)

- 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

2.3.7 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.(10)

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.(9)

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.(9)

2.4.2 Intensidade

Como a frequência de um estímulo a intensidade dele é processado e codificado sub-cortical nos dois lados do cérebro em todos os níveis no cérebro.(9)

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 (7).

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 articulação normal, mantém-se relativamente estável. Há uma opção de contribuição das cavidades nasais, uma cobertura porém é negligenciável. Baseado nessas características é divididos 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 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 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

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 Capítulo 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 estendido 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 dificuldade de se realizar o supracitado computacionalmente. Mesmo com essa restrição, o modelo de Lucero [11] é o modelo utilizado para a representação do caráter oscilatório das pregas vocais no software SynPath, sendo que esta restrição não solucionada não tem forte influência no produto final.

Requisitos Funcionais

O Synpath é constituído também dos seguintes requisitos funcionais, os requisitos funcionais são as funcionalidades que o sistema executará(?)

- 1 - Validação dos 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 às 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 parâmetros 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: 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.(6). 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.(11)

3.0.4 MHRSMM

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

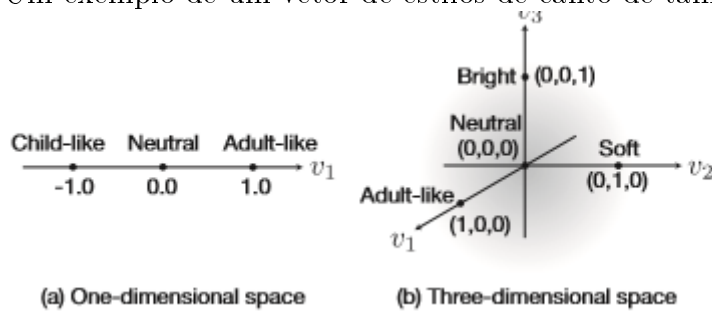
$$\mu_i = H_b i \xi$$

$$m_i = H_p i \xi$$

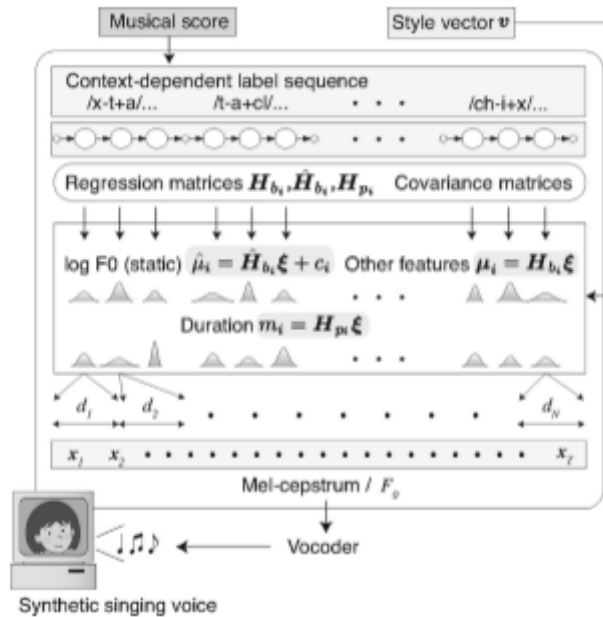
$$\xi = [1, v_1, v_2, \dots, v_L]^T$$

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

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



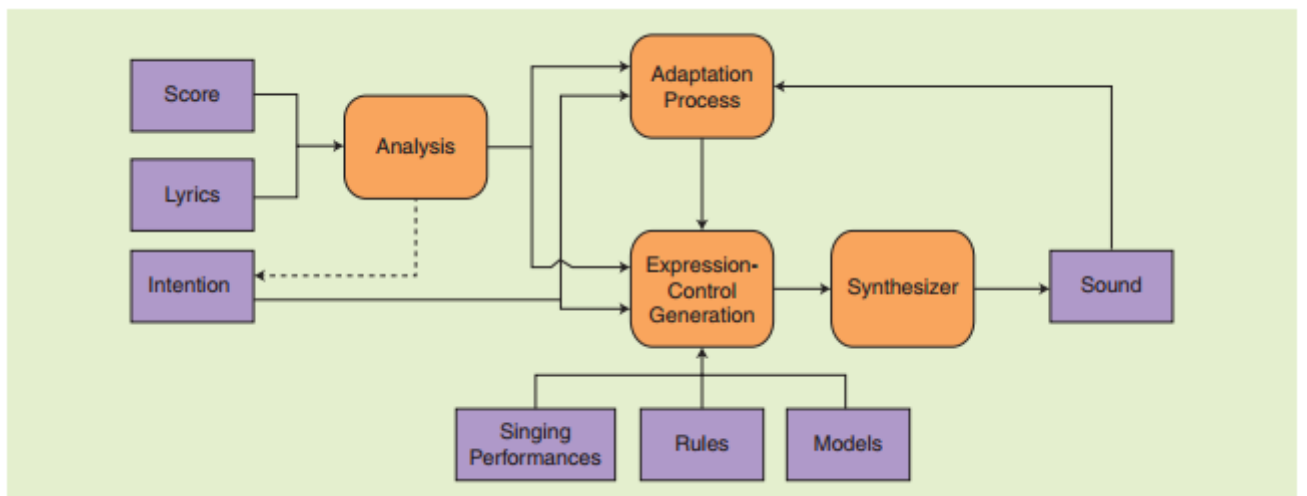
Durante a fase de síntese o usuário do programa adiciona vetores de estilos de acorodo com a intenção e a expressividade pretendida. Parametros de output como duração são gerados pelos vetores de estilos dados e matrizes de regressões treinadas usando MRHSMMs

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 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.(16)

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 tornam-se do timbre, vibrato, tremolo, timing fonético.(16)

3.0.6 Envoltória F0

Envoltórias F0 são usadas para expressar informação linguística, para-linguística e não-linguística.(12)

As Envoltórias F0 apresentam três (3) características importantes que fazem diferenciar uma voz falada a uma voz cantada.(13)

- 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.(10)

Referências

- [1] John & Charles Bell. The anatomy and physiology of the human body, 1826. [Down, pre-B, ED].
- [2] Ferdinand de Saussure. *Self-oscillating source for vocal-tract synthesis*. IEEE Tran. Audio Eletroacoust, Audio Eletroacoust., 1968. 6
- [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. 7, 14
- [4] Jean Schoentgen Samia Fraj and Francis Grenez. *Development and perceptual assesment of a synthesizer of disordered voices*. Journal of Acoustical Society of America, 2012. 7
- [5] Henry Gray. *Anatomy of the Human Body*. IEEE Tran. Audio Eletroacoust. 6
- [6] 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. 7, 14
- [7] James L. Flanagan Jont B. Allen Mark A. Hasegawa-Johnson. *Speech Analysis Synthesis and Perception*. 2008. 10
- [8] 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. 4, 6
- [9] Jody Kreiman and Diana Van Lancker Sidtis. Foundations of voice studies, 2011. vii, 2, 3, 10
- [10] 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. 9, 16
- [11] 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. 7, 14

- [12] 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. 9, 16
- [13] 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. 9, 16
- [14] BradH Story. *Tubetalker*. Dept. of Speech, Language, and Hearing, Tucson, AZ, 2010. 5
- [15] Ingo Titze. *Principles of Voice Production*. Prentice Hall, New Jersey 07632, 1994. 5
- [16] 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. 9, 16