Síntese Articulatória: Uma Introdução

António Teixeira, Luís Nuno Silva, Francisco Vaz

***Resumo* – Neste artigo apresenta-se uma visão geral acerca da síntese articulatória, cobrindo: a história, abordagens habituais, os diferentes modelos existentes para os principais blocos que constituem um sintetizador deste tipo, e como obter os parâmetros necessários à utilização destes modelos. A título exemplificativo, os modelos adoptados no sintetizador articulatório em desenvolvimento pelos autores para aplicação à síntese em Português, baptizado de SAP (*Sintetizador Articulatório para o Português*), são apresentados com algum detalhe.**

***Abstract* – In this paper we present an overview of articulatory synthesis, covering: history, usual approaches, different models for the main blocks of an articulatory synthesizer, how to obtain parameters for the models. Models adopted for the articulatory synthesizer in development by the authors for Portuguese, named SAP, are described in some detail.**

I. Síntese Articulatória

> The next generation of text-to-speech will probably be based on vocal tract line analogues or a parallel formant synthesis designed for automatic and complete simulation of a line analogue.
>
> GUNNAR FANT [1, pág. 77]

A síntese articulatória gera o sinal de voz através da modelação das características físicas, anatómicas e fisiológicas do aparelho produtor de voz humano. A grande diferença para outros sistemas, como a síntese de formantes [2], é que nesta técnica se modela directamente o sistema em lugar de se modelar o sinal ou as suas características acústicas. Nas abordagens baseadas no sinal [1] o objectivo é reproduzir o sinal de voz natural o mais fielmente possível com poucas, ou nenhumas, preocupações para a forma como este é produzido. Por contraste, um modelo baseado no sistema produtor utiliza leis da física para descrever a propagação no tracto e modela os fenómenos de mecânica e física de fluidos para descrever a oscilação das cordas vocais.

Para implementar um sintetizador articulatório num computador digital precisa-se de um modelo matemático do sistema vocal. Geralmente os sintetizadores incluem dois subsistemas: um modelo anatómico-fisiológico das estruturas envolvidas na produção de voz e um modelo da produção e propagação do som nessas estruturas.

O primeiro modelo transforma as posições dos articuladores, como o maxilar, língua, e velo, na área de secção do tracto vocal. O segundo modelo consiste num conjunto de equações que descrevem as propriedades acústicas do sistema vocal. Geralmente é constituído por vários submodelos para simular diferentes fenómenos, como: a criação de uma fonte de excitação periódica por oscilação das cordas vocais; fontes de som causadas pelo fluxo turbulento no caso de existência de zonas de área bastante reduzida ao longo do tracto; propagação do som nas cavidades acima e abaixo das cordas vocais; radiação nos lábios e/ou narinas.

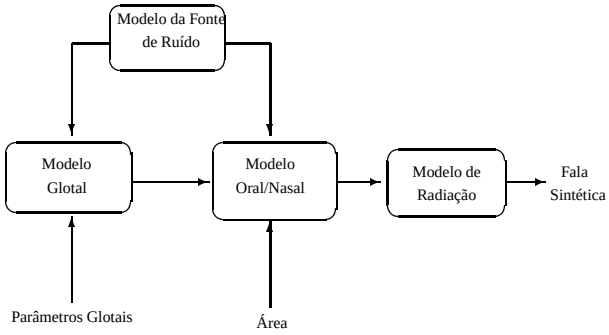


Figura 1 - Estrutura básica da síntese articulatória.

Os parâmetros para os modelos podem ter várias origens. Podem ser obtidos directamente de sinal de voz por um processo de inversão por optimização, serem definidos manualmente pelo investigador, ou serem a saída da parte de processamento linguístico de um sistema de conversão de texto para fala.

Estes sintetizadores ainda não atingiram o desenvolvimento necessário para serem uma alternativa aos métodos actualmente utilizados em sistemas de conversão de texto para fala. Isto deve-se a diversos factores: a dificuldade de obter informação acerca do tracto vocal e das cordas vocais durante a produção de voz em seres humanos; as técnicas de medição directa geralmente apenas nos darem valores para configurações estáticas, não sendo fácil obter informação acerca da dinâmica dos articuladores; não existe, ainda, um processo de análise para obtenção dos parâmetros articulatórios a partir de voz natural; os cálculos necessários são complexos e demorados.

Apesar das desvantagens, a síntese articulatória apresenta algumas vantagens importantes: os parâmetros do sintetizador estão directamente relacionados com os mecanismos articulatórios humanos, sendo portanto muito úteis em estudos de produção e percepção de voz [11]; como os parâmetros variam lentamente no tempo são bons candidatos ao uso em processos de codificação eficientes; este método pode produzir consoantes nasais e vogais nasais com elevada qualidade [12]; os parâmetros são mais fáceis de interpolar que os parâmetros LPC e os dos sintetizadores de formantes [13], pequenos erros nos sinais de controlo não provocam geralmente sons de baixa naturalidade, pelo facto dos valores interpolados serem fisicamente realizáveis; a in-

[1] A descrição destes métodos, mesmo sem grande profundidade, tornaria este artigo demasiado extenso. Descrições genéricas dos vários métodos podem ser encontradas em [3]-[5]. Diversas obras apresentam descrições detalhadas dos vários métodos existentes [6]-[10].

teracção entre fonte e o tracto, que é essencial para um som natural, pode ser convenientemente modelada [14].

II. Breve história da síntese de voz baseada no modelamento articulatório

> To understand a science it is necessary to know its history.
>
> Auguste Comte (1798–1857)

De uma forma resumida, apresentam-se de seguida alguns passos na síntese de voz, em especial os relacionados com a síntese articulatória.

Há muitos anos que o Homem demonstra curiosidade acerca da produção de voz. Essa curiosidade levou-o a investigar se seria ou não capaz de produzir voz artificial.

Os princípios da teoria acústica da produção de fala já eram conhecidos no século XVIII. Já nessa época a laringe era considerada como a principal fonte sonora utilizada na fala.

O Professor Kratzenstein, na Rússia, construiu em 1769 cinco tubos acústicos que excitados por palhetas produziam as vogais /a,e,i,o,u/ [15, pág. 8].

O primeiro sintetizador de voz deve-se ao barão Wolfgang von Kempelen, um nobre Austríaco. Em 1791 demonstrou, em Viena, a sua máquina mecânica falante que imitava vogais e algumas consoantes, incluindo nasais. O seu sintetizador, capaz de produzir cerca de 20 sons diferentes, era composto por um fole, uma caixa de ar comprimido, um ressoador de couro e apitos accionados por alavancas. Embora a qualidade deixasse certamente muito a desejar, estes eram suficientemente próximos dos sons da fala para poderem ser identificados como vogais e consoantes. As vogais eram produzidas alterando manualmente o volume do ressoador de couro. A produção de consoantes exigia um maior virtuosismo por parte do operador que tinha de accionar as alavancas para criar orifícios por onde passava o ar, ao mesmo tempo que, com os dedos, controlava o grau de fechamento e a forma desses orifícios. Apesar de rudimentar, esta máquina abriu caminho para futuras explorações. Mais detalhes podem ser encontrados em [15] e [16].

Steward [17] foi o primeiro a produzir vogais utilizando um dispositivo eléctrico.

Um dos primeiros sintetizadores eléctricos foi demonstrado em 1936 por Homer Dudley. O seu *Voder* (ou *Voice Operation Demonstrator*) conseguiu, pela primeira vez, sintetizar voz contínua usando circuitos eléctricos. Este dispositivo foi demonstrado na Feira Mundial de Nova Iorque, em 1939, onde operadores especialmente treinados produziram frases a pedido dos visitantes.

O *Pattern Playback* [18] que apareceu em 1950 nos Laboratórios Haskins é o primeiro exemplo de um sintetizador moderno, não articulatório. A evolução das formantes era desenhada numa placa de vidro, depois varrida (*scanned*) para produzir voz. Este dispositivo, conhecido como sintetizador opto-electrónico, produzia o som descrito pelo espectrograma. O uso extensivo desta ferramenta promoveu muito o estudo da produção e percepção de voz.

Chiba e Kajiyama [19] publicaram, em 1958, estudos da resposta do tracto utilizando integração numérica da equação de Webster.

Num trabalho percursor, Dunn [20] recorreu à teoria das linhas de transmissão eléctricas para desenvolver uma descrição quantitativa da acústica do tracto vocal. Construiu um modelo análogo eléctrico. É considerada a primeira simulação do tracto vocal. Este modelo consistia de 25 secções em T de $0.5\ cm$ de comprimento e área igual a $6\ cm^2$. Uma indutância variável podia ser inserida entre duas secções para simular a língua. Outra indutância variável representava a constrição nos lábios. A radiação era simulada medindo a tensão na saída aos terminais de uma pequena indutância. Para sons vozeados, o sintetizador era excitado com uma onda triangular de que se podia controlar a frequência fundamental. O espectro da fonte era ajustado de forma a ter um decréscimo de $12\ dB/oitava$. Para simular os sons surdos e murmurados, uma fonte de ruído era aplicada num ponto apropriado da linha.

Foi efectuado um modelo eléctrico melhorado por Stevens, Kasowski e Fant [21]. Mais tarde Rosen [22] construiu um modelo mais detalhado incluindo o tracto nasal. Para o estudo do sistema subglotal van den Berg [23] construiu outro modelo eléctrico. A variação contínua dos elementos da linha de transmissão por meios electrónicos permitiu a estes dispositivos sintetizar sons contínuos [22]. Outro exemplo de modelo eléctrico foi o sintetizador FLEA, desenvolvido por Fant [24].

Todos os sintetizadores iniciais usando linhas de transmissão utilizaram redes analógicas na sua implementação. No entanto as técnicas digitais, tornadas possíveis com o desenvolvimento do computador, oferecem vantagens em termos de estabilidade e precisão. Um dos primeiros sintetizadores digitais utilizou os coeficientes de reflexão nas junções dos elementos cilíndricos [25].

Outra implementação em computador simulou as propriedades das linhas de transmissão usando equações diferença equivalentes. Com esta formulação foi possível estudar a interacção acústica entre o tracto vocal e as cordas vocais. Esta técnica foi usada num sintetizador completo para sons surdos e sonoros por Flanagan e colaboradores [26], [27].

Também na década de sessenta tiveram lugar as primeiras tentativas de obtenção da configuração do tracto com base no sinal acústico. As primeiras abordagens basearam-se na relação entre as áreas dos diversos tubos que podem ser usados para aproximar o tracto e os coeficientes de reflexão. Estes coeficientes são facilmente derivados dos coeficientes de predição linear (LPC) [28], [29]. Outra técnica utilizada baseou-se na medição da resposta impulsional nos lábios [30].

Os primeiros modelos representando a cavidade oral no plano sagital são apresentados no final da década de sessenta [31], [32, são dois exemplos]. Um dos modelos mais utilizados, ainda hoje, foi proposto por Mermelstein em 1973 [33].

Na década de oitenta os modelos foram sendo melhorados [34], [35] e é proposto o modelo híbrido por Sondhi e Schroeter [13].

Com a melhoria das técnicas computacionais e de obtenção de dados acerca do processo de produção tem-se assistido, nos últimos anos, ao desenvolvimento de modelos tridimensionais do tracto e a utilização de novos métodos de

simulação dos fenómenos acústicos. Em relação à inversão, o poder de cálculo permitiu: a utilização de métodos baseados em optimização, utilizando, por exemplo, algoritmos genéticos; a utilização de redes neuronais [36]; e melhorar os processos baseados na procura em tabelas.

III. Modelamento das cavidades

O primeiro aspecto do processo de produção de voz que é necessário modelar é a geometria dos tractos oral e nasal. O tracto nasal é essencialmente constante. O tracto oral, no entanto, varia continuamente a sua forma. Devido às suas características específicas, um e outro são modelados de forma diferente.

A. Modelos para o tracto vocal

A geometria do tracto vocal pode ser convenientemente descrita em termos da posição dos articuladores: a língua, lábios, glote, maxilar, etc. Modelos baseados neste tipo de descrição são designados por modelos articulatórios [37, pág. 233].

Um grande número de modelos articulatórios pode ser encontrado na literatura. Podem ser classificados em dois tipos principais: modelos paramétricos da área e modelos sagitais.

A.1 Modelos paramétricos da área

Os modelos paramétricos da área não representam as posições dos articuladores directamente, concentram-se no modelamento da área ao longo do tracto vocal. Um grande número de modelos deste tipo foi utilizado [38], [24], [39]-[43]. A sua característica comum é especificarem a área, A_c, e a posição, X_c, de máxima constrição. A área é geralmente representada por funções contínuas como hipérboles, parábolas ou sinusóides [41].

Parte destes modelos é baseada nas características acústicas, como o modelo DRM proposto por Mrayati em 1988 [44, pág. 224].

Este tipo de modelos, modelando directamente a área, contemplou inicialmente apenas os sons vocálicos, só mais recentemente foi feita a sua extensão para configurações consonânticas [45, por exemplo].

A.2 Modelos sagitais

Os modelos sagitais são baseados numa representação no plano sagital como o de uma imagem de raios X. Descrevem o movimento dos órgãos empregues na produção de voz num plano sagital. Todos os modelos deste tipo incluem as limitações do tracto vocal. Por exemplo, a língua não pode passar através do palato. A visualização e a interpretação do estado dos articuladores são as principais vantagens destes modelos. Estes modelos podem dividir-se em estáticos ou dinâmicos, descritivos ou funcionais [43]. Outra classificação, utilizada em [46], divide-os em: geométricos, estáticos, estatísticos e fisiológicos.

Um exemplo de um modelo dinâmico funcional é o de Henke [32]. É controlado por gestos (*gesture*) ou alvos (*targets*) articulatórios que são controlados por equações do movimento dos articuladores. Outros exemplos são o modelo de Perkell [47] e o desenvolvido nos Laboratórios Haskins [48, por exemplo].

Modelos articulatórios estatísticos, baseados na extracção de componentes principais de imagens de raios X e medições da abertura dos lábios, foram propostos por Kiritani e Maeda [49]. O modelo de Maeda é descrito em detalhe em [46].

Os modelos de mais fácil compreensão são os modelos descritivos estáticos como os desenvolvidos por Mermelstein [33] e Coker [31], [50].

Os modelos deste tipo apenas representam a configuração do tracto no plano sagital médio. Para os modelos acústicos é necessária informação tridimensional.

Antes da passagem de duas a três dimensões, o plano sagital é decomposto em várias secções para as quais se determina o comprimento e a distância entre os contornos superior e inferior no plano sagital. Utiliza-se, na decomposição, uma grelha onde cada secção corresponde à zona do tracto compreendida entre dois segmentos de recta que definem a secção. Utilizam-se diversos tipos de grelhas, sendo no entanto as mais utilizadas baseadas no sistema de coordenadas proposto por Heinz e Stevens [51]. Este sistema de coordenadas divide o tracto em três zonas: a primeira entre a glote e a parte superior da faringe, consistindo de linhas paralelas horizontais; a segunda, entre a faringe e a parte média da cavidade bucal, usando linhas radiais convergindo no ponto de origem das coordenadas; e a última representando as zonas restantes do tracto até aos lábios, usando linhas paralelas verticais.

Diversos autores estudaram a obtenção da função de área (área e comprimento das várias secções ao longo do tracto) com base nas distâncias sagitais [52]-[54, por exemplo]. Geralmente a conversão entre a distância sagital e a área de secção é efectuada usando uma formula do tipo

$$\text{Área} = a \times (\text{largura no plano sagital})^b \ ,$$

em que os coeficientes a e b são determinados empiricamente de medições do tracto, usando métodos directos como raios-X ou imagens de ressonância magnética (MRI). A relação não é no entanto simples, pois os coeficientes são bastante variáveis ao longo da laringe [55] e os coeficientes variam de estudo para estudo.

A.3 Modelo articulatório utilizado no SAP

O modelo, apresentado na Figura 2, é constituído por 3 partes distintas: uma parte fixa, uma parte ajustável e a parte variável definida pela posição dos articuladores. Manteremos na descrição a denominação original, proveniente do Inglês, dos pontos e dos parâmetros articulatórios.

Constituem a parte fixa: o ponto fixo F sobre o qual roda o maxilar; a parede posterior da faringe (pontos G, G1, G2 e W); a parte do palato duro (entre N e M) e incisivos superiores (ponto U). O contorno posterior-superior é fixo, excepto para a zona do palato mole, representada pelo arco M-V'. É também fixo o ponto mais elevado do velo, fechando a passagem para o tracto nasal (ponto V) e a inclinação da recta ao longo da qual se desloca a extremidade do velo. A distância entre o maxilar e o ponto fixo F, designada por sj,

é também mantida fixa, assim como o raio do arco de circunferência utilizada na representação do corpo da língua.

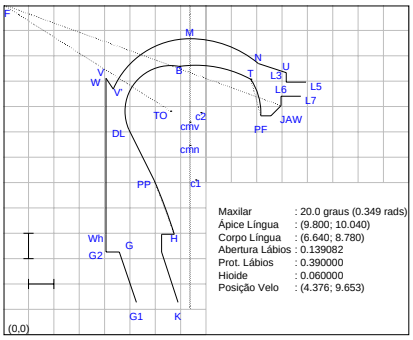


Figura 2 - Modelo Articulatório implementado. Os quadrados têm 1 cm de lado.

A parte inferior da faringe pode ter as suas dimensões alteradas variando-se 3 parâmetros:

- A distância horizontal **wh**, do ponto H, intersecção da parte anterior da epiglote com a parte superior do osso hióide, à parede posterior da faringe;
- A distância vertical **hk1**, do mesmo ponto H à posição da glote;
- A distância na horizontal **g1k**, entre os pontos K e G1. O ponto K representa uma estimativa da posição da extremidade anterior da laringe.

O resto do modelo depende das posições dos articuladores, definidas pelos parâmetros articulatórios. Os articuladores contribuem para a definição de pontos utilizados no modelo da seguinte forma:

- O maxilar é representado, em coordenadas polares, por (sj,thetaj). Como já referido a distância sj é mantida constante. O parâmetro articulatório **jaw** é igual ao ângulo thetaj. A zona junto ao ponto que define o maxilar é aproximada pelos segmentos de recta PF-PS-JAW-L6;
- O corpo da língua é representado pelo arco de circunferência DL-B com o centro móvel e raio fixo. As coordenadas rectangulares do centro (**tbodx**, **tbody**) constituem parâmetros articulatórios;
- O ápice da língua é representado pelas coordenadas rectangulares (**ttx**, **tty**), do ponto T. Os arcos B-T e T-PF representam o contorno. Como o ponto B varia com a posição do centro da língua (tbodc) e o ângulo do maxilar (jaw), a zona definida pela ponta da língua é afectada pela posição destes dois outros parâmetros articulatórios;
- Os lábios são representados pelos pontos L5 (lábio superior) e L7 (lábio inferior). Relativamente ao ponto **jaw**, as coordenadas do lábio inferior são representadas por (**lipp**,**lipo**) que representam, respectivamente, a protrusão e abertura dos lábios. A utilização destes dois parâmetros como variáveis separadas permite ter lábios fechados, lábios separados ou configuração arredondada. O lábio superior, representado por L5, tem as mesmas coordenadas mas em relação ao ponto U;
- A posição do hióide é definida pelo parâmetro **hyoid**, representando a distância entre o ponto PP e o segmento de recta H-DL. O ponto PP encontra-se na perpendicular ao segmento H-DL que passa pelo ponto médio deste;
- O estado do véu palatino é representado pela posição do ponto V', representando a ponta da úvula que se move ao longo do segmento de recta V-V'. A abertura velar é proporcional à distância entre o ponto V e a posição mais elevada do véu palatino. No modelo, esta distância é especificada pelo parâmetro **velum**. O arco M-V', com centro na linha vertical que passa pelo ponto M, é afectado pela posição do véu palatino.

A.4 Modelos tridimensionais

Os modelos descritos, até agora, são bidimensionais. O tipo de dados disponíveis na altura em que foram desenvolvidos não permitia a inclusão da terceira dimensão. Mais recentemente, técnicas melhoradas usando imagens de ressonância magnética (MRI) contribuíram para conhecimento da geometria tridimensional [53], [56], tendo aparecido modelos tridimensionais como o desenvolvido por Engwall [57]. Este modelo consiste de uma malha tridimensional de polígonos repartidos por cinco áreas, representando as paredes do tracto oral e nasal, lábios, dentes e língua. A malha tem 750 vértices e aproximadamente 1000 polígonos. Para reduzir a complexidade foi assumido que existe simetria em relação ao plano sagital médio. Os parâmetros articulatórios utilizados neste modelo seguem, em larga medida, os do modelo de Mermelstein [33], modificados para o caso tridimensional. Os parâmetros permitem controlar a altura da laringe, abertura do maxilar, protrusão dos lábios, arredondamento dos lábios, posição do velo, e os movimentos da língua. O modelo considera a língua como um todo. Os movimentos do ápice e dorso são sobrepostos ao modelo base. Apesar de não atingir a sofisticação de modelos da língua como os propostos por Wilhelms-Tricarico [58] é um modelo bastante detalhado.

B. Modelos das cavidades subglotais

Não são geralmente modeladas directamente as dimensões destas cavidades, optando-se por modelar usando equivalentes acústicos, descritos mais adiante. Uma excepção é o modelamento efectuado por Boersma [59]. Este investigador utilizou uma sequência de 29 tubos com comprimento fixo e área dependente da região subglotal a modelar [59, pág. 46, para mais detalhes].

C. Modelos do tracto nasal

Ao longo dos anos, vários modelos do tracto nasal foram sendo usados em síntese articulatória. Os primeiros usaram dados provenientes de cadáveres [60] e de moldes do tracto nasal [24]. Estes primeiros modelos apenas modelavam as cavidades nasais não incluindo os seios paranasais e juntavam as duas passagens laterais, não considerando as assimetrias. Foi sugerido por Fujimura e Ludqvist [61] que as cavidades paranasais seriam necessárias para explicar o espectro de vogais naturais. Um dos primeiros a incluir no seu modelo o efeito dos seios foi Maeda [12] que obteve as suas dimensões por um processo de análise-síntese. Maeda con-

siderou apenas uma cavidade. Outros investigadores estudaram estas cavidades, como Masuda, em 1992, dissecando mais de 20 crânios e estudando as consequências acústicas de obstrução da passagem (ostia) (citado em [56]). Recentemente foi efectuado um estudo detalhado usando imagens de ressonância magnética (MRI) [56]. Este estudo obteve informação tridimensional da área do tracto nasal e dimensões dos seios. Os valores da área diferem consideravelmente dos anteriormente publicados [60], [24], em especial na zona média.

Apresentam-se, de seguida, resumidamente, alguns destes modelos.

C.1 Modelo de House e Stevens (1956) [60]

As dimensões deste modelo foram baseadas largamente em atlas anatómicos, crânios, e imagens de raios-X laterais. O modelo analógico nasal era acoplado ao modelo do tracto vocal $8\ cm$ acima da glote. Não era feito qualquer ajuste à área oral ao fazer variar a área de acoplamento nasal.

C.2 Modelo DANA

Este modelo foi desenvolvido por Hecker [62], [63] para ser integrado no sintetizador eléctrico desenvolvido no MIT por Rosen [22]. O nome de DANA adveio-lhe da denominação inglesa *Dynamic Analog of the NAsal cavities*.

Consistia em 9 secções com um comprimento total, fixo, de $12.5\ cm$. As secções 1 e 2, representando a nasofaringe, operam em conjunto e constituem um secção de $3\ cm$ com área variável electronicamente (de aproximadamente 0.05 a $5.0\ cm^2$). A área da secção 3 era manualmente variável ($2.0, 4.0, 6.0, 8.0$ e $10.0\ cm^2$). As secções 4 a 7 representavam uma região de área aproximadamente constante ($2.6\ cm^2$), e a secção 8 oferecia controlo manual ($0.4, 0.8, 1.2, 1.6$ e $2.0\ cm^2$). A secção 9 tinha área igual a $0.42\ cm^2$. Para um adulto do sexo masculino, as cavidades nasais eram acopladas aproximadamente $8\ cm$ acima da glote.

C.3 Modelo de Maeda, 1982

Neste modelo o tracto nasal tem um comprimento de $11\ cm$ e é representado por 11 secções de $1\ cm$ de comprimento com as áreas apresentadas na Figura 3. As primeiras 3 secções têm área variável, sendo a área da primeira secção a área de acoplamento nasal e a área das secções 2 e 3 obtida por interpolação linear entre a área da primeira e da quarta secção.

Os seios foram representados por uma única cavidade, os seios maxilares, com um volume de $20.8\ cm^3$ acoplado ao tracto nasal por um tubo de $0.5\ cm$ de comprimento e $0.1\ cm^2$ de secção, a uma distância de $7\ cm$ do véu palatino. Esta cavidade foi modelada por uma concatenação de várias secções [12]. O efeito do seio na resposta do tracto nasal, considerando a abertura na zona de acoplamento nula, pode ver-se na Figura 4.

Diversos investigadores usaram os dados de Maeda para a área do tracto nasal [64]. Este modelo, com adaptações, foi utilizado por Sondhi e Schroeter [13] no seu modelo híbrido. Estes investigadores modelaram os seios paranasais usando um circuito ressonante RLC com impedância

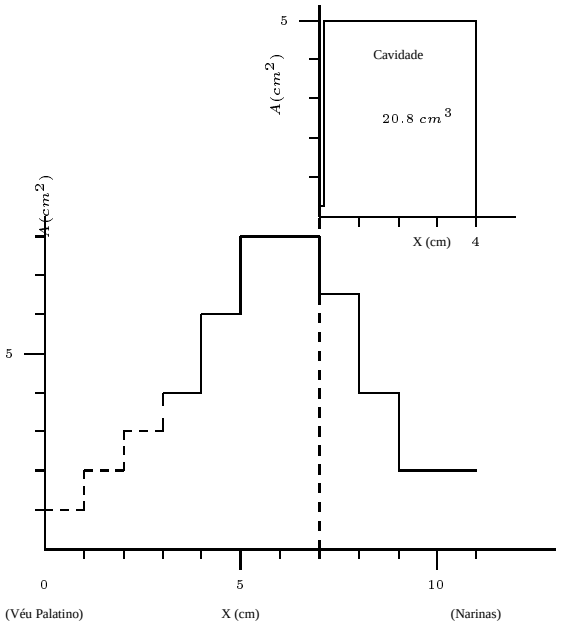


Figura 3 - A função de área do tracto nasal segundo Maeda, 1982. A tracejado indicam-se as secções com área variável.

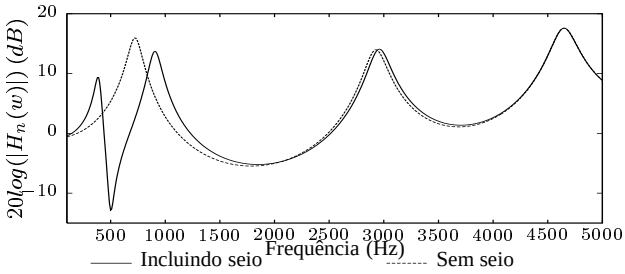


Figura 4 - Resposta do tracto nasal usando os dados de Maeda, 1982. Foi calculada a resposta incluindo ou não os seios maxilares. O factor S para as perdas usado foi de 2, modelo de radiação de Flanagan, 1972 [12].

$Z_{seio} = R_{seio} + j\omega L_{seio} + \frac{1}{j\omega C_{seio}}$, representanto o ressoador de Helmholtz constituído pela cavidade dos seios e a ligação, de área reduzida, destes com as cavidades nasais [65, pág. 15].

C.4 Modelos com área de radiação reduzida

Diversos autores propuseram, e utilizaram, modelos das cavidades nasais em que a área de radiação, isto é, a área das narinas é mais reduzida do que a medida em seres humanos. Como base para esta escolha encontra-se a necessidade de ter modelos com características acústicas equivalentes às do tracto nasal humano.

Os primeiros a utilizar este tipo de modelos foram House e Stevens [60], que utilizaram uma área de $0.23\ cm^2$.

G. Feng [66] fez um estudo exaustivo, concluindo pela necessidade de utilização deste tipo de modelos. Apresenta, também, uma possível explicação anatómica para este tipo de modelos. Segundo este autor a utilização de uma área reduzida das narinas justifica-se pela existência de uma zona de passagem relativamente estreita, um pouco antes das narinas, designada por *limen nasi*. Nos seus trabalhos de simulação utilizaram um valor de $0.6\ cm^2$ [66], [67].

M. Chen [68], baseando-se em dados de [69] e [56], tam-

bém utilizou uma área de radiação reduzida, $0.5\ cm^2$.

Båvegård e colegas [70] utilizaram os dados anatómicos de [24] reduzindo para metade a área das secções, cobrindo os primeiros $4\ cm$ a contar das narinas. Chamaram a este modelo "nariz estreito" (do Inglês *narrow nose*).

C.5 Modelos assimétricos

A utilização de ressonância magnética permitiu a obtenção de dados anatómicos mais detalhados das cavidades nasais. Tornou possível a medição em condições mais próximas das reais (sem aplicação de soluções destinadas a diminuir a cobertura mucosa), a obtenção de dados acerca das duas passagens laterais e, ainda, dados acerca das dimensões das cavidades paranasais e suas ligações às cavidades nasais.

Um dos estudos mais relevantes foi o efectuado por Dang e Honda [56]. Foram medidas áreas de secção, e perímetros das passagens nasais para 4 indivíduos. Foram também obtidos dados relativamente aos seios paranasais maxilares e esfenoidais. As principais conclusões dos estudos efectuados por estes investigadores foram [56]:

1. as diferenças entre indivíduos são maiores para o volume do que para o comprimento da cavidade nasal;
2. os valores de área diferem grandemente dos anteriormente publicados [60], [24], em particular na parte média do tracto nasal;
3. é necessário implementar as cavidades paranasais num modelo do tracto nasal para descrever adequadamente as suas propriedades acústicas;
4. o tracto nasal tende a ser assimétrico, sendo necessário estudar os efeitos dessa assimetria em sons nasais.

Experiências com um modelo assimétrico, baseado nos dados de Dang e Honda foram efectuadas por Lin em 1994 [71].

Os valores obtidos por Story [72], usando também imagens de ressonância magnética, confirmam a assimetria das passagens nasais esquerda e direita.

C.6 Modelamento do tracto nasal no SAP

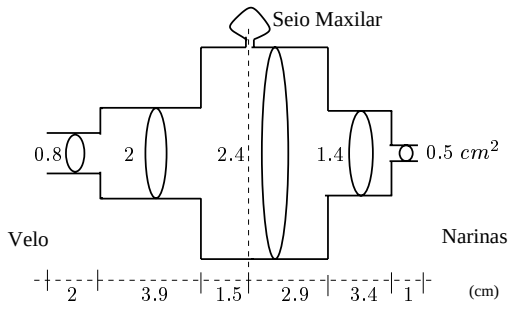


Figura 5 - Modelo nasal utilizado.

O modelo por nós adoptado foi o apresentado por Chen [68] baseado nos estudos de Dang, Honda [56] e Stevens [69]. As dimensões encontram-se representadas na Figura 5. Note-se a área de radiação pelas narinas igual a $0.5\ cm^2$ e a inclusão de um seio paranasal, o maxilar. O modelo utiliza apenas um tubo, assumindo simetria das duas passagens nasais.

IV. Modelos acústicos

O modelo acústico do sistema vocal humano engloba vários submodelos. Os modelos do tracto oral e nasal simulam a propagação nesses tractos. A fonte de excitação glotal representa e gera a onda de excitação glotal. O fluxo turbulento de ar numa constrição nas fricativas e oclusivas é produzido pelo modelo da fonte de ruído. O modelo de radiação simula a radiação da energia acústica dos lábios e narinas.

O tracto vocal é um tubo acústico tridimensional curvo com forma lentamente variável ao longo do tempo. As paredes do tubo são flexíveis, existem perdas provocadas por atrito e condução de calor. As condições de fronteira, nos pontos de radiação e glote, são também variáveis. Existe a possibilidade de acoplamento de um tubo adicional representando o tracto nasal. O acoplamento é feito na parte superior da faringe. O tracto nasal tem dimensões fixas, mas o acoplamento é variável.

Investigações preliminares demonstraram que a equação de Navier-Stokes para fluxo de fluidos pode caracterizar as não linearidades envolvidas na produção de som pelas cordas vocais; a produção de fricativas surdas por fluxo turbulento em constrições; e efeitos de radiação condicionados por propagação num tubo não-uniforme, com perdas e de paredes flexíveis [73], [74]. No entanto, os resultados têm sido limitados pelas exigências computacionais necessárias à resolução da equação de Navier-Stokes numa grelha tempo-espaço realística. Estas limitações levaram os investigadores à procura de modelos simplificados.

A. Simplificações

> Truth is much too complicated to allow anything but approximations.
>
> JOHN VON NEUMAN

A primeira simplificação que é geralmente efectuada no modelamento acústico do tracto consiste em "esticar" o tracto vocal. Segundo estudos de Sondhi [75] a variação nas formantes provocada por esta simplificação situa-se no intervalo 2 a $4\ \%$, para frequências inferiores a $4\ kHz$. O tracto vocal pode ser representado por um tubo direito sem grande perda de precisão.

A segunda aproximação é considerar a propagação das ondas como sendo planar ao longo do tubo. Existem duas razões que justificam esta aproximação: o tecido ao longo do tracto contraria a propagação radial; e as dimensões laterais médias na ordem dos $2.0\ cm$ levam a que outros modos de propagação só ocorram para frequências próximas ou acima do limite superior das frequências [2] com informação do sinal de voz. Em [76], pág. 175, deduzem-se as expressões para as frequências a partir das quais existem outros modos de propagação, para o caso de um tubo cilíndrico infinito. Os primeiros três modos de ordem superior têm frequências angulares de corte iguais a $1.84c/a$, $3.05c/a$ e $3.80c/a$, sendo a o raio do tubo e c a velocidade do som. Para uma área, relativamente elevada [3], de

[2] Geralmente 4 a $5\ kHz$ para sons não fricativos e $8\ kHz$ para fricativas.

[3] O valor máximo da área para as vogais americanas, segundo os dados de Story [72], não atinge os $8\ cm^2$.

$15\ cm^2$, a frequência de corte, do primeiro modo, é de cerca de $4700\ Hz$. Por esta razão os algoritmos, baseados na aproximação planar da propagação, são considerados válidos até $4000 - 5000\ Hz$ [72, pág. 30]. Felizmente, a maior parte da informação do sinal de voz encontra-se abaixo dos $4000\ Hz$.

Mesmo desprezando as perdas por fricção, condução e as resultantes das paredes flexíveis, as equações daí resultantes, em geral, apenas podem ser resolvidas numericamente. Precisamos, pois, de mais uma aproximação. Uma abordagem habitual é dividir o tracto num conjunto de secções cilíndricas contíguas. Faz-se, portanto, uma discretização espacial do tubo. Se o número de secções for elevado, estes elementos de comprimento reduzido constituem uma boa aproximação da função de área contínua. As frequências de ressonância do conjunto de tubos são muito próximas das obtidas no caso contínuo. O tubo cilíndrico uniforme torna-se de muito mais fácil análise. Mesmo assim, as primeiras análises não incluíram as perdas.

B. Equação de onda

O tracto vocal constitui um tubo acústico com forma variável. Considerando, numa primeira aproximação, as paredes rígidas, a teoria acústica linear [76]-[78] descreve a propagação do som através das equações de continuidade e conservação do momento [79, pág. 7]

$$\frac{\partial p}{\partial t} + \rho c^2 \frac{\partial v_i}{\partial x_i} = 0 \qquad \text{e} \qquad \rho \frac{\partial p}{\partial t} + \frac{\partial p}{\partial x_i} = 0$$

Nestas equações ρ é a densidade do meio, c a velocidade de propagação do som, v_i a velocidade da partícula na direcção x_i, e p representa a pressão.

Assumindo propagação planar, apenas é necessário considerar formas das equações utilizando uma dimensão. Reescrevendo as equações, com substituição das velocidades pelo fluxo, obtém-se

$$\frac{\partial p}{\partial t} + \rho c^2 \frac{\partial}{\partial x}\frac{u}{A(x)} = 0 \qquad \text{e} \qquad \rho \frac{\partial}{\partial t}\frac{u}{A(x)} + \frac{\partial p}{\partial x} = 0$$

em que, u é o fluxo, e $A(x)$ é a área de secção que é função de x. Estas duas equações combinadas resultam na conhecida equação de Webster [80]

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{1}{A(x)} \frac{\partial}{\partial x}\left[A(x)\frac{\partial p}{\partial x}\right] .$$

Esta equação não inclui perdas. A falta de uma solução analítica desta equação para geometrias arbitrárias levou, nos primeiros modelos de sintetizadores articulatórios desenvolvidos, à utilização de analogias com linhas de transmissão, assunto da próxima secção.

C. Modelo para um tubo usando a analogia com uma linha de transmissão

Dunn [20] propôs um modelo em que o tracto é aproximado por uma série de tubos com área constante. Cada tubo foi modelado recorrendo à analogia com linhas de transmissão, relacionando a resistência acústica, inertância e complacência com a resistência, indutância e capacidade eléctricas.

Para um tubo de área constante A, não incluindo perdas, as equações anteriores simplificam-se, obtendo-se,

$$\frac{\partial p}{\partial t} + \frac{\rho c^2}{A}\frac{\partial u}{\partial x} = 0 \qquad \text{e} \qquad \frac{\rho}{A}\frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0$$

Os leitores familiarizados com a teoria das linhas de transmissão recordarão que, para uma linha de transmissão uniforme sem perdas, a tensão v e a corrente i na linha satisfazem as equações

$$\frac{\partial v}{\partial x} + L\frac{\partial i}{\partial t} = 0 \qquad \text{e} \qquad \frac{\partial i}{\partial x} + C\frac{\partial v}{\partial t} = 0$$

onde L e C são a indutância e capacitância por unidade de comprimento, respectivamente. A teoria de linhas de transmissão aplica-se ao estudo da transmissão num tubo acústico, se usarmos as analogias apresentadas na Tabela I.

Grandeza acústica		Grandeza eléctrica análoga	
p	pressão	v	tensão
u	fluxo (*volume velocity*)	i	corrente
ρ/A	inertância (ou indutância acústica)	L	indutância
$A/(\rho c^2)$	complacência (ou capacitância acústica)	C	capacitância

Tabela I

ANALOGIAS ENTRE GRANDEZAS ACÚSTICAS E ELÉCTRICAS.

C.1 Inclusão de perdas no modelo

As perdas, desprezadas por Dunn [20], foram adicionadas ao modelo por Stevens, Kasowski e Fant [21]. É possível representar as perdas provocadas pelo fluxo laminar por uma resistência em série e outra em paralelo [15, pág. 43]. A resistência em série representa as perdas devidas à viscosidade, proporcionais ao quadrado do fluxo; a condutância em paralelo representa as perdas devidas à transmissão de calor, proporcionais ao quadrado da pressão.

O mecanismo que provoca as perdas por viscosidade é o atrito. Se uma camada de ar junto à parede do tubo se pode considerar estacionária e o ar no centro do tubo se move com uma velocidade v, então existe um gradiente radial de velocidade. A fricção pode considerar-se como ocorrendo entre anéis concêntricos de ar, cada um movendo-se a uma velocidade ligeiramente diferente dos seus vizinhos. A "resistência acústica" por unidade de comprimento é dada por [10] $R = \frac{S}{A^2}\sqrt{\frac{\omega\rho\mu}{2}}$. Atente-se que R depende da frequência angular não sendo portanto uma resistência no sentido usado em Electrotecnia. Também depende de S.

A condutância em paralelo introduz perdas proporcionais ao quadrado da tensão, representando as perdas no tubo acústico por condução de calor nas paredes. Este processo é difícil de visualizar porque se considera o tubo a uma temperatura uniforme. Isto é apenas verdade ao nível macroscópico. As variações rápidas e adiabáticas da pressão causam variações de temperatura. Flanagan [10] mostrou que a "condutância acústica" é $G = \frac{S(\eta-1)}{\rho c^2}\sqrt{\frac{\lambda\omega}{2\xi\rho}}$ sendo λ a condutividade térmica e ξ o calor específico do ar. Como R, também G depende de ω e S.

Um problema surge no que respeita à escolha do perímetro para calcular estas resistências. Geralmente o tubo acústico é considerado circular, o que implica $S = 2\sqrt{\pi A}$ para a circunferência. Fant [24] duplicou esse valor, o que corresponde a uma forma elíptica ou, no caso de uma forma circular, a um aumento do atrito. Este valor duplo foi adoptado por exemplo por Wakita e Fant [81]. Num modelo mais detalhado, uma conversão entre a área e o perímetro dependente da localização no tracto poderia ser usada. No entanto, são necessários mais dados anatómicos e acústicos para se poder fazer esse refinamento do modelo.

C.2 Paredes flexíveis

Até este momento consideraram-se as paredes do tubo rígidas. No tracto vocal esta aproximação não é válida. Não só as paredes vibram devido à onda de pressão, como também o volume do tubo é alterado com a variação da pressão. O efeito das paredes é primordial no caso das oclusivas sonoras. Flanagan, Ishizaka e Shipley [82] adicionaram novos elementos à analogia das linhas de transmissão para modelar as paredes flexíveis, assim como o som radiado, devido às vibrações das paredes do tracto.

As variações de pressão no interior do tracto submetem as paredes a uma força variável. Como as paredes são elásticas, a área do tubo irá variar. Assumindo que a reacção das paredes é local, o movimento normal à superfície, de um segmento das paredes, depende apenas da pressão acústica nesse segmento e é independente de qualquer outro segmento; a vibração da parede é simulada por um modelo mecânico com massa, viscosidade e complacência. O circuito equivalente é um circuito RLC série, com $L_p = \frac{m_p}{S^2}$, $C_p = \frac{S^2}{k_p}$ e $R_p = \frac{b_p}{S^2}$ [49], onde S representa, mais uma vez, o perímetro do tubo.

A impedância das paredes pode ser incluída em cada secção como um elemento distribuído [10], [83], [82], [40], [84], [49], [85] ou inserida como duas impedâncias discretas, uma na faringe e outra ao nível do queixo [81], [86], [41]. Pode ainda usar-se um factor de correcção [41]. O modelo discreto, que é independente da configuração do tracto, pode não dar resultados satisfatórios. O condensador foi eliminado em alguns estudos devido a ter um efeito muito reduzido [81]. Mais informação pode ser encontrada em [49], [41], [15], [85].

C.3 Modelo equivalente

Incluindo os vários componentes descritos anteriormente obtém-se o circuito equivalente para uma secção de tubo elementar, com área de secção constante, representado na Figura 6. Fazendo $z = R + j\omega L$, $y = G + j\omega C + 1/Z_p$ com $Z_p = R_p + j\omega L_p + \frac{1}{j\omega C_p}$, a constante de propagação γ é dada por $\gamma = \sqrt{zy}$ e a impedância característica Z por $Z = \sqrt{\frac{z}{y}}$. Os elementos Z_a e Z_b do circuito equivalente em T são,

$$Z_a = Z\tanh\left(\frac{\gamma l}{2}\right) \qquad Z_b = \frac{Z}{\sinh(\gamma l)}$$

Utilizando estes valores, as relações entre as correntes e tensões para o circuito em T podem ser facilmente derivadas, quer no domínio do tempo [87], quer no domínio da frequência [13].

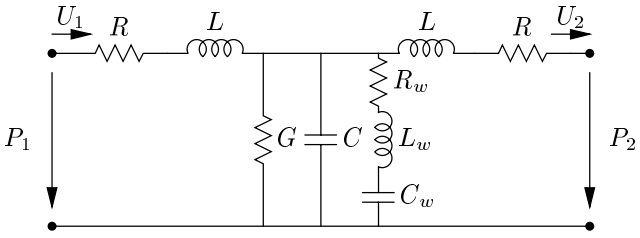


Figura 6 - Circuito equivalente de um tubo com perdas. Adaptado de [10]

D. Modelo alternativo proposto por Sondhi

Sondhi e Schroeter [13], com base em trabalho anterior do primeiro [88], derivaram um outro circuito equivalente. Algumas correcções foram propostas depois em [37]. O resultado final é semelhante ao obtido por Flanagan [10]. No entanto, segundo os autores, a derivação é mais rigorosa. A dedução entra em linha de conta com as perdas devidas a viscosidade, condutividade e paredes flexíveis, sem ter de aproximar o tracto por uma cascata de secções uniformes. Os parâmetros do modelo podem ser determinados a partir de medições acústicas.

Remetem-se os leitores interessados em mais detalhes para as referências citadas, apenas se referindo aqui os resultados. A relação entre as pressões e fluxos nos dois extremos de uma secção do tracto homogénea é dada por,

$$\begin{pmatrix} P_s \\ U_s \end{pmatrix} = \begin{pmatrix} A & C \\ B & D \end{pmatrix} \times \begin{pmatrix} P_e \\ U_e \end{pmatrix} = K \times \begin{pmatrix} P_e \\ U_e \end{pmatrix} ,$$

onde a entrada se encontra do lado glotal e a saída do lado dos lábios ou narinas. P_s representa a pressão na saída e U_s o fluxo na saída. P_e e U_e representam as mesmas grandezas, mas agora na entrada do tubo. Para um tubo de comprimento l e área A os elementos da matriz de transmissão K são dados pelas expressões [13, pág. 959],

$$A = \cosh(\sigma l/c) \qquad C = -\frac{A}{\rho c}\frac{\sinh(\sigma l/c)}{\lambda}$$
$$B = -\frac{\rho c}{A}\lambda\sinh(\sigma l/c) \qquad D = \cosh(\sigma l/c).$$

As variáveis complexas σ e λ definem-se como $\lambda = \sqrt{\frac{a+j\omega}{\beta+j\omega}}$, $\sigma = \lambda(\beta + j\omega)$ com $\alpha = \sqrt{j\omega c_1}$ e $\beta = \frac{j\omega\omega_0^2}{(j\omega+a)j\omega+b} + \alpha$. Valores para os parâmetros (a, b, c_1, ω_0) podem ser encontrados em [13], [37].

Comparando as expressões dos elementos da matriz de transmissão apresentados com o caso geral, em função de γ e Z, de um modelo de tubo com inclusão de perdas [89, por exemplo], em que $A = \cosh(\gamma l)$ e $B = -Z\sinh(\gamma l)$, facilmente se obtém a constante de propagação e impedância característica como $\gamma = \frac{\sigma}{c}$ e $Z = \frac{\rho c}{A}\lambda$. Caso se esteja interessado no circuito em T, Z_a e Z_b obtêm-se da mesma forma que no modelo anterior.

E. Modelos das cavidades subglotais

O sistema subglotal, que inclui a traqueia e os pulmões, é geralmente omitido nas simulações, pois o seu efeito nas características espectrais é considerado pequeno, excepto para sons surdos, em que a abertura da glote é grande [90].

Foram efectuadas medições da impedância de entrada do sistema subglotal [90]. Ananthapadmanabha e Fant [91] usaram os dados destas medições e representaram o sistema subglotal por uma cascata de ressonâncias, representadas por circuitos paralelos RLC. Utilizaram apenas três circuitos para representar as três primeiras ressonâncias das cavidades subglotais. As formantes do sistema situam-se em 640, 1335, e $2110\ Hz$, com larguras de banda de 246, 155 e $140\ Hz$, respectivamente.

Outros valores foram propostos por Fant, Ishizaka, Lindqvist e Sundberg em 1972 (citados em [81, pág. 14]). As frequências de ressonância situam-se em 600, 1350 e $2160\ Hz$ com larguras de banda de 240, 180 e $190\ Hz$, respectivamente.

Os efeitos foram estudados por Fant em colaboração com Ananthapadmanabha [91], Badin [86] e Lin [41], concluindo que o efeito do sistema subglotal é pequeno, excepto para sons surdos, onde a abertura glotal é grande.

Além deste tipo de modelamento acústico simplificado, motivado pela escassez de dados acerca das configurações, foram também usadas técnicas semelhantes às utilizadas para as cavidades supraglotais [23], [59].

F. Modelos de radiação

A energia acústica abandona o tracto vocal pelos lábios. No caso dos sons nasais, parte da energia é também libertada pelas narinas. A pressões normais [4], a radiação das paredes da garganta é geralmente desprezável, excepto para oclusivas sonoras A radiação neste caso foi simulada por Flanagan, Ishizaka e Shipley [82], colocando uma impedância em cada secção do modelo do tracto vocal. Como neste trabalho não abordaremos as oclusivas, não consideramos este efeito.

Nos análogos que representam o tracto como uma linha de transmissão, os lábios e as narinas são representadas por impedâncias de radiação que carregam o tracto vocal e nasal. Estas impedâncias possuem uma parte resistiva e outra reactiva. A primeira, que é responsável pelo consumo de energia, aumenta por um factor superior a w^2 e é portanto um factor preponderante nas larguras de banda das formantes com frequências mais elevadas. A parte reactiva representa a massa efectiva posta em vibração em frente aos lábios e/ou narinas, tornando o comprimento efectivo do tracto superior às suas dimensões físicas.

Um tratamento matemático preciso desta impedância pode ser obtido considerando-a como um pistão vibrante (do Inglês *vibrating piston*), a abertura dos lábios ou narinas, numa esfera, a cabeça [10]. Este modelo é conhecido por PIS (do Inglês *Piston In Sphere*). No entanto este modelo não é computacionalmente eficiente envolvendo o cálculo de séries. Outro modelo mais simples é o de radiação de uma abertura circular num plano infinito. Este modelo é válido, pois a abertura de radiação, nos lábios ou narinas, é muito menor que a cabeça.

[4]Para voz de mergulhadores esta radiação é também importante devido à menor rigidez das paredes do tracto vocal.

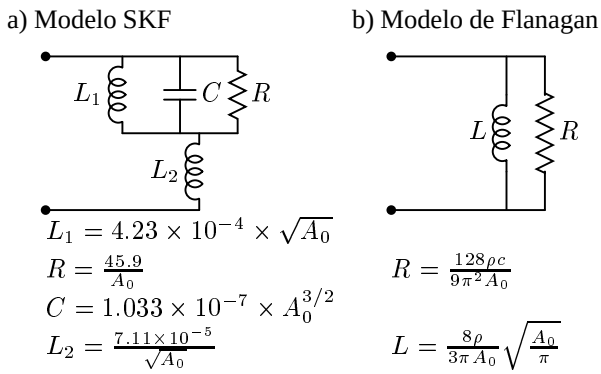


Figura 7 - Modelos de Radiação. (a) Modelo SKF [21]. (b) Modelo paralelo de Flanagan [10]

Várias simplificações foram propostas para aplicações práticas. Uma delas foi proposta por Stevens, Kasowski e Fant, em 1953, utilizando uma resistência e três outros componentes dependentes da frequência. Este modelo ficou conhecido pelas iniciais dos autores, modelo SKF. Outras propostas, usando elementos dependentes da frequência, foram feitas por Fant [24] e Wakita [81]. Flanagan [10] apresentou uma aproximação ao modelo de radiação num plano infinito, usando uma associação em paralelo de dois componentes independentes da frequência. A Figura 7 apresenta o modelo SKF e o modelo de Flanagan [10].

Como o ouvido humano é sensível às variações de pressão do ar, a pressão sonora a uma distância d dos pontos de radiação é o objectivo final dos cálculos. A pressão sonora a uma distância d, $p_r(t)$, está relacionada com o fluxo radiado, $u_r(t)$. A relação depende da forma da boca e narinas assim como da cabeça do locutor. Fant [24] usando o domínio da frequência propôs a seguinte relação:

$$\frac{P_r(w)}{U_r(w)} = \frac{\rho\omega}{4\pi d} K_T(w).$$

O factor $K_T(w)$ é um ênfase de cerca de $1.5\ dB$ por oitava de 312 a $5000\ Hz$. Representa dois efeitos: o do reflector (em Inglês *baffle*) e o aumento da resistência de radiação para além da sua proporcionalidade com a frequência. Devido à falta de verificação experimental, $K_T(w)$ é geralmente considerado unitário, sendo a relação, no tempo,

$$p_r(t) = \frac{\rho}{4\pi d} \frac{\partial u_r(t - \frac{d}{c})}{\partial t},$$

geralmente aproximada pela derivada de $u_r(t)$ [86], [69].

O leitor com interesse em mais detalhes sobre este assunto poderá consultar [92], [78], [10, pág. 36], [24], [15, pág. 48].

V. Métodos de resolução do modelo acústico

Tendo modelado a configuração dos tractos e os fenómenos acústicos de excitação, propagação e radiação, torna-se necessário obter a informação desejada que consiste, geralmente, no sinal radiado.

Em geral, são usadas três abordagens principais em sintetizadores articulatórios: filtros de onda digitais; resolução das equações diferenciais; método híbrido. Com a disponibilidade crescente de meios poderosos de cálculo, têm sido, em anos mais recentes, tentadas outras técnicas com menos limitações e menos simplificações.

A. Filtros de onda digitais (Wave Digital Filters)

A equação de Webster (da página 7) no caso de a área se manter constante reduz-se a:

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}$$

D'Alembert publicou uma solução geral para esta equação em 1747, com a forma [93],

$$p(x,t) = f(t - x/c) + g(t + x/c).$$

As funções $f(.)$ e $g(.)$ são completamente gerais e contínuas, podendo ser interpretadas como ondas de forma arbitrária, mas fixa, que se propagam em direcções opostas ao longo do eixo dos xx com velocidade c. A pressão $p(x,t)$ em qualquer ponto é dada pela soma de dois componentes, uma onda propagando-se para a frente $p^+(x,t)$ (sentido positivo do eixo), e outra para trás $p^-(x,t)$.

A equação apenas é válida para uma secção de área constante do tracto, sendo o tracto aproximado pela concatenação de vários tubos de área constante. Na junção de duas secções, com impedâncias diferentes, devido a áreas diferentes, cada onda sofre os efeitos da descontinuidade. Parte da onda $p^+(x,t)$ continua a propagação no mesmo sentido, a parte restante $rp^+(x,t)$ é reflectida, propagando-se no sentido contrário, somando-se a $p^-(x,t)$. O factor r é chamado coeficiente de reflexão, e é definido por:

$$r = \frac{Z_f - Z_t}{Z_f + Z_t},$$

onde Z_f é a impedância característica para a frente da junção, e Z_t é a impedância para trás da junção.

O mesmo tipo de considerações aplica-se à onda $p^-(x,t)$ excepto o cálculo de r, onde se tem de trocar os termos Z_f e Z_t devido à direcção de propagação.

A modelação acústica do tracto usando filtros de onda digitais baseia-se neste conceito. Mais detalhes podem ser obtidos em [94], [89], [15, por exemplo]. Refira-se que nesta técnica é utilizada directamente a função de área, não sendo necessário construir um modelo análogo.

Este modelo foi originalmente proposto por Kelly e Lochbaum [25], tendo sofrido, ao longo dos anos, diversos melhoramentos. Foi modificado para incluir efeitos da variação dinâmica da área por [95]. Rubin, Baer e Mermelstein [11] modificaram o modelo de Kelly e Lochbaum para representar uma terminação não ideal na glote, lábios e narinas. Calcularam os coeficientes de reflexão e a função de transferência no domínio z. Baseando-se na função de transferência, implementaram filtros digitais.

Uma apresentação mais elegante do modelo de Kelly e Lochbaum foi proposta por Fettweis e Meerkötter em 1975. Ficou conhecida por filtros de onda digitais (em Inglês *wave digital filters*) [96].

Em geral, esta abordagem é a mais rápida, sendo também adequada a implementações paralelas. Foi mesmo realizado um sistema completo em tempo real, usando *hardware* especial por Meyer, Wilhelms e Strube [97].

Os maiores problemas desta abordagem são: a dificuldade de modelar as perdas dependentes da frequência; dificuldade de inclusão da interação entre a fonte glotal e o tracto; a dificuldade em ter um comprimento do tracto arbitrário.

O problema do comprimento pode ser combatido utilizando variação da frequência de amostragem [98] ou *fractional delay wave digital filters* [99], [100].

No que respeita à inclusão das perdas, alguns progressos foram conseguidos nos últimos anos [72], [97], [72], mas ainda é necessária mais investigação.

B. Métodos no tempo

Nestes métodos, começa-se por discretizar espacialmente o tracto, constrói-se de seguida um circuito equivalente (utilizando os modelos anteriormente descritos), sendo as equações diferenciais parciais que relacionam a pressão e o fluxo (ou os seus análogos eléctricos), discretizadas no tempo. O conjunto de equações diferença obtido é depois resolvido para cada instante de tempo, por forma a se obter a pressão e fluxo em cada ponto da linha de transmissão [27], [26], [82], [83], [40]. Os valores da pressão e fluxo, num instante no tempo, são usados para calcular parâmetros do circuito equivalente, a utilizar nos cálculos para o instante seguinte. Esta abordagem foi designada por resolução no tempo (em Inglês *time-domain*).

Nestes sintetizadores é usada uma frequência de amostragem bastante elevada, para evitar *frequency-warping* [81]. Os componentes dependentes da frequência são simulados a uma frequência fixa. Os efeitos desta aproximação foram estudados por Hsieh [85]. Apesar destas aproximações, o som obtido é natural.

O modelo inicial de Flanagan foi simplificado por Maeda [49], substituindo o modelo mecânico vibrante das cordas vogais por um modelo representando a área de abertura da glote; não incluindo as fontes de ruído e omitindo os efeitos dos seios nasais. Estas simplificações tornaram as simulações muito mais rápidas. Outras simplificações foram introduzidas por Bocchieri [101], reduzindo o número de fontes de ruído e usando um terminal gráfico para desenhar o contorno do tracto vocal. Baseado no trabalho de Maeda, foi desenvolvido um sintetizador por Childers e Ding [64], usando um circuito equivalente e convertendo as equações acústicas em equações algébricas lineares. Hsieh [85], trabalhando com Childers, rederivou as equações para incluir o sistema subglotal, a impedância glotal, o ruído de turbulência e os seios paranasais.

Um dos exemplos mais completos de aplicação desta técnica é o sintetizador desenvolvido por Boersma [59].

C. Métodos híbridos

Este método, proposto por Sondhi e Schroeter [13], e representado de forma muito resumida na Figura 8, difere dos dois anteriores, ao utilizar o domínio da frequência para

modelar as cavidades supraglotais. Enquanto a glote é modelada no domínio do tempo, devido à sua natureza altamente não linear, o tracto vocal e o tracto nasal são modelados na frequência, aproveitando o facto para modelar, de forma mais precisa, as perdas e a radiação, fenómenos que, como já foi referido, dependem da frequência. Os dois modelos, da fonte e tracto, são interligados, na proposta inicial, através da transformada inversa de Fourier e convolução. A utilização de informação acerca da impedância de entrada do tracto permite a realização de sistemas com interacção entre a fonte glotal e o tracto. A designação de método híbrido resulta da utilização simultânea do domínio tempo e domínio frequência. Informações complementares acerca deste método podem ser encontradas em [102].

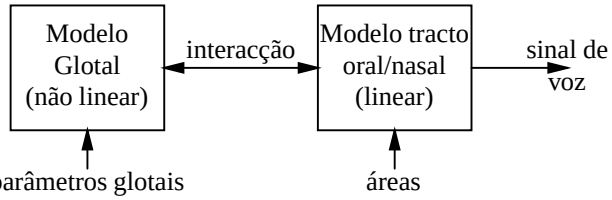


Figura 8 - Descrição geral do método híbrido de síntese articulatória, adaptado de [13].

Como a utilização de transformada inversa e convolução requerem um tempo considerável, foi proposto por Lin [41] a aproximação da resposta na frequência por um conjunto de filtros de segunda ordem, substituindo-se a convolução pelo processo de filtragem. Este método, apesar de mais rápido, necessita da obtenção dos filtros de segunda ordem, tarefa bastante complicada no caso de existência de antiformantes (zeros da resposta em frequência).

D. Outras técnicas

A aplicação de métodos variacionais à equação de Webster e às condições fronteira do tracto foi proposta por van Praag [103] e Jospa [104]. Esta técnica permite tratar configurações com diversos ramos, com aplicação na simulação de sons envolvendo uma cavidade nasal incluindo a assimetria entre as duas passagens que terminam nas narinas. Permite também a obtenção de pólos e zeros da função de transferência o que poderá ser útil para nasais.

Métodos geralmente aplicados na simulação de sistemas electromagnéticos como o *Transmission Line Matrix* ou *Transmission Line Model* (TLM) têm também sido aplicados aos modelos acústicos do tracto [105].

O acesso, recentemente, a poderosos meios computacionais tem permitido a utilização de métodos usando elementos finitos na resolução da equação de Navier Stokes [73], [106]. Este tipo de simulações é especialmente interessante para sons em que a propagação se torna turbulenta, como as fricativas. Servem também os resultados deste método para validação de modelos mais simplificados. De facto, resultados para vogais mostram como geralmente válidas as aproximações habitualmente utilizadas [107].

Os métodos usando elementos finitos e o TLM permitem a utilização de informação tridimensional disponibilizada por técnicas como a ressonância magnética .

E. Modelo acústico utilizado no SAP: Análise na frequência utilizando matrizes

Na exposição que se segue não se assume qualquer modelo para o tubo elementar. O método tanto é válido para o modelo de Sondhi e Schroeter utilizado como para qualquer outro em que tenhamos uma matriz ABCD.

E.1 Modelo de um tubo elementar

Cada secção do modelo acústico pode ser representada, na frequência, como uma função de transferência representada na forma matricial por uma matriz ABCD,

$$\begin{bmatrix} P_s(\omega) \\ U_s(\omega) \end{bmatrix} = \begin{bmatrix} A(\omega) & B(\omega) \\ C(\omega) & D(\omega) \end{bmatrix} \begin{bmatrix} P_e(\omega) \\ U_e(\omega) \end{bmatrix} = K(\omega) \begin{bmatrix} P_e(\omega) \\ U_e(\omega) \end{bmatrix} .$$

A matriz relaciona a pressão, $P_s(\omega)$, e a velocidade de volume, $U_s(\omega)$, à saída, com a pressão, $P_e(\omega)$, e velocidade de volume, $U_e(\omega)$, na entrada do tubo. Designemos esta matriz por $K(\omega)$.

Os elementos $A(\omega)$, $B(\omega)$, $C(\omega)$, $D(\omega)$, variam com a frequência, incluindo o efeito de vários tipos de perdas. São função do comprimento e área seccional do tubo. No resto da exposição do modelo não incluiremos a dependência dos elementos da matriz com a frequência para simplificar as expressões. Neste trabalho utilizamos o modelo proposto por Sondhi e Schroeter [13].

E.2 Modelo com vários tubos

Um modelo composto por várias secções, num total de N, pode ser representado pelo produto de N matrizes, cada uma representando uma secção,

$$K_{Nsecs} = \prod_{i=1}^{N} K_i = \begin{bmatrix} A_{Nsecs} & B_{Nsecs} \\ C_{Nsecs} & D_{Nsecs} \end{bmatrix} .$$

Com base na matriz K_{Nsecs} a função de transferência do conjunto, terminado por uma impedância de carga Z_c, pode ser obtida por:

$$H_{Nsecs} = \frac{U_s}{U_e} = \frac{A_{Nsecs} D_{Nsecs} - C_{Nsecs} B_{Nsecs}}{A_{Nsecs} - C_{Nsecs} Z_c},$$

e a impedância de entrada do conjunto é

$$Z_e = \frac{P_e}{U_e} = \frac{D_{Nsecs} Z_c - B_{Nsecs}}{A_{Nsecs} - C_{Nsecs} Z_c}.$$

Note-se a igualdade do denominador nas duas expressões anteriores. De uma forma similar podem obter-se as funções P_s/U_e, P_s/P_e e U_s/P_e.

E.3 Modelo completo do tracto

Para poder simular todas as cavidades supraglotais, incluindo as cavidades nasais, torna-se necessário decompor o tracto em várias regiões. Na Figura 9 estão representadas as regiões utilizadas. No seu estado de desenvolvimento o SAP apenas permite a produção de sons com excitação glotal. Para outros sons, como por exemplo as fricativas, teria de considerar-se ainda a decomposição da zona oral entre os lábios e a zona de acoplamento do tracto nasal [108].

O tracto vocal, terminado pela impedância de radiação dos lábios Z_l, divide-se em duas regiões:

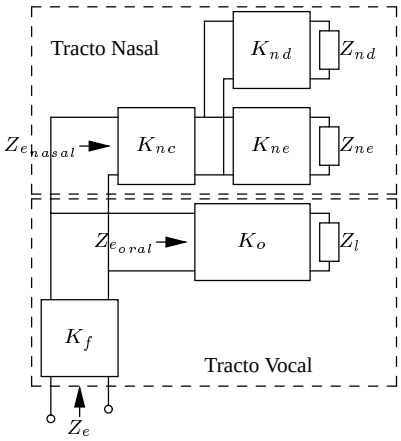


Figura 9 - Modelo acústico completo do SAP.

1. A região faríngea, entre a glote e a zona de acoplamento do tracto nasal, representada pela matriz K_f. Não são permitidas oclusões nesta região em Português;
2. A região oral, entre a zona de acoplamento do tracto nasal e os lábios, representada por K_o.

O modelo geral para o tracto nasal é constituído por três regiões:

1. A região nasal comum, que constitui a continuação da faringe, até à bifurcação do tracto nasal nas suas duas passagens, representada por K_{nc};
2. A passagem nasal esquerda, que termina na narina esquerda, representada por K_{ne};
3. A passagem nasal direita, representada por K_{nd}.

No modelo existem duas impedâncias de radiação, uma para cada narina, representadas por Z_{ne} e Z_{nd}.

No caso de se considerar o tracto nasal simétrico deixa de ser necessário considerar as três regiões. No modelo implementado, neste caso, apenas se considera a existência da zona nasal comum e uma única impedância de radiação nasal Z_n.

No modelo do tracto nasal as cavidades paranasais (seios) são representadas por circuitos ressonantes de Helmholtz, inseridos em paralelo. A impedância, $Z_{Helmholtz}$, representativa destes circuitos é incluída no cálculo das matrizes nasais utilizando a matriz

$$K_{seio} = \begin{bmatrix} 1 & 0 \\ -1/Z_{Helmholtz} & 1 \end{bmatrix}.$$

E.4 Função de transferência total e impedância de entrada

Tem-se, em geral, três pontos de radiação: as duas narinas e os lábios [5]. Desprezando os efeitos dos diferentes trajectos desde o ponto de radiação até ao ponto de medição da pressão sonora total, utilizaremos a soma das radiações nestes diferentes pontos como pressão total. Para obter o sinal radiado em cada um destes pontos, torna-se necessário obter as matrizes entre a glote e esse ponto.

[5]Não se incluem neste trabalho as radiações pelas paredes do tracto, devido á sua reduzida relevância para o caso das vogais.

A matriz entre a glote e os lábios, K_{gl}, é dada por

$$K_{gl} = K_o \times K_{an} \times K_f \ ,$$

onde K_{an} é a matriz referente ao acoplamento do tracto nasal,

$$K_{an} = \begin{bmatrix} 1 & 0 \\ -1/Z_{e_{nasal}} & 1 \end{bmatrix},$$

sendo $Z_{e_{nasal}}$ a impedância do tracto nasal vista do velo. Quando o velo se encontra posicionado de forma a fechar a passagem para o tracto nasal $Z_{e_{nasal}}$ é infinita e K_{an} torna-se a matriz identidade. Também se se pretender não incluir o efeito da carga nasal no cálculo de K_{gl} pode fazer-se esta matriz, K_{an}, igual á matriz identidade.

A matriz entre a glote e a narina esquerda, K_{gne}, é dada por

$$K_{gne} = K_{ne} \times K_{and} \times K_{nc} \times K_{ao} \times K_f \ ,$$

onde K_{ao} é a matriz de acoplamento representando a impedância de entrada da região oral na zona de acoplamento do tracto nasal, sendo K_{and} a matriz de acoplamento representativa da impedância de entrada da passagem nasal direita.

De uma forma similar obtém-se a matriz entre a glote e a narina direita como sendo

$$K_{gnd} = K_{nd} \times K_{ane} \times K_{nc} \times K_{ao} \times K_f .$$

A função de transferência completa é

$$H_{tot} = \frac{U_{ne} + U_{nd} + U_l}{U_g} = \frac{U_{ne}}{U_g} + \frac{U_{nd}}{U_g} + \frac{U_l}{U_g} \ ,$$

representando U_g o fluxo glotal. As respostas parciais obtêm-se de K_{gl}, K_{gnd}, e K_{gne}.

Como o ouvido humano é sensível às variações de pressão, o objectivo final dos cálculos é obter a pressão radiada. O efeito da radiação pode ser representado, de forma aproximada, pela derivada do fluxo radiado [24]. Na prática, é usual efectuar a derivada do fluxo à entrada do tracto em vez de a efectuar no fluxo radiado, técnica que adoptamos. Utilizando como excitação a derivada do fluxo glotal, a função de transferência, H_{tot}, permite obter directamente a pressão.

E.5 Obtenção da resposta impulsional

Na secção anterior, descrevemos como obter a resposta das cavidades supraglotais para uma frequência. Para sintetizar um som é necessário a resposta impulsional para efectuar a convolução com a onda de excitação glotal.

O processo utilizado é o seguinte: N amostras da função de transferência são obtidas entre 0 e metade da frequência de amostragem, com intervalo constante. Na implementação actual, em que a frequência de amostragem é de $10\ kHz$, $N = 256$, dando uma resolução de aproximadamente $19.5\ Hz$ e uma resposta impulsional com 512 amostras. A resposta em frequência é filtrada com o filtro utilizado por Schroeter e Sondhi [37] $H_f(z) = \frac{1+z^{-1}}{1+0.95z^{-1}}$.

À resposta, depois de filtrada, é aplicada uma Transformada Inversa de Fourier, utilizando-se uma implementação rápida desenvolvida por Frigo [109].

O mesmo procedimento é aplicado à impedância de entrada, $Z_e(w)$, para se obter $z_e(n)$ necessária para implementação de interacção entre a fonte glotal e o tracto, como veremos na secção seguinte.

VI. Fontes de excitação

A. Fonte de excitação glotal

Existem três tipos de modelos [110]: modelos glotais paramétricos não-interactivos, em que não existe interacção entre a fonte glotal e o tracto vocal; modelos glotais mecânicos e paramétricos interactivos, que incluem, implícita ou explicitamente, a interacção entre a fonte e o tracto vocal; e modelos glotais fisiológicos, baseados em teorias de comportamento fisiológico das cordas vocais.

A.1 Modelos paramétricos

São os mais simples pois assumem que a fonte e o tracto vocal são separáveis não existindo interacção entre os dois. Baseiam-se pois na teoria fonte-filtro, proposta por Fant [24]. São muito usados em codificação e em síntese acústica de voz.

O modelo trigonométrico [111], representado na Figura 10, é definido por:

$$u_g(t) = \begin{cases} \frac{\alpha}{2}\left(1-\cos\left(\frac{t\pi}{T_P}\right)\right) & \text{para } 0 \le t \le T_P \\ \alpha\cos\left(\frac{\pi}{2}\frac{t-T_P}{T_N}\right) & \text{para } T_P \le t \le T_P + T_N \end{cases} .$$

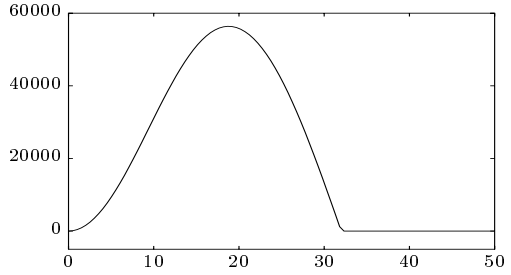


Figura 10 - Modelo do fluxo glotal trigonométrico de Rosemberg [111]. $\alpha = 56383.798$, $T_P = 18.784$ e $T_N = 13.216$, valores de [110, Fig. 5].

Como o efeito da radiação pode ser modelado usando a primeira derivada, pode modelar-se a derivada do fluxo glotal. O modelo mais utilizado é o modelo LF, proposto por Liljencrants e Fant [112]. A sua popularidade deve-se a diversos factores, dos quais ressalta a facilidade em obter os seus parâmetros. A forma do modelo é

$$u_g^{'}(t) = \begin{cases} E_0\, e^{\alpha t}\sin\left(\omega_g\,(t)\right) & ,0 \le t \le T_e \\ -\frac{E_e}{\epsilon T_a}\left(e^{-\epsilon(t-T_e)} - e^{-\epsilon(T_c-T_e)}\right) & ,T_e \le t \le T_P + T_c \end{cases}$$

A Figura 11 mostra um período do modelo LF. O modelo LF original tem 5 parâmetros: E_0, α, ω_g, T_a e F_0.

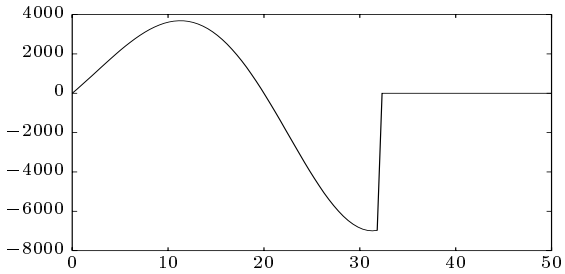


Figura 11 - Modelo LF do fluxo glotal [112]. $E_0 = 2622.799$, $\alpha = 0.032$, $f_g = 0.025$, $T_e = 32.0$, $\epsilon = 2.0$, valores de [110, Fig. 6].

A.2 Modelos glotais mecânicos e paramétricos interactivos

Exemplo de um modelo paramétrico interactivo é o modelo proposto por Allen e Strong [113]. Este modelo parametriza a área utilizando uma fórmula proposta por Titze: A área glotal, $A_g(\theta)$, é calculada segundo

$$A_g(\theta) = \begin{cases} A\left(\left(\frac{\theta}{\theta_m}\right)^{-\theta_m\,\cot\,\theta_m}\left(\frac{\sin\theta}{\sin\theta_m}\right)\right)^{\beta} & \text{para } \theta \le \pi \\ 0 & \text{para } \theta \ge \pi \end{cases} ,$$

onde $\theta = \frac{\pi t}{\gamma T}$; $\theta_m = \frac{\pi\delta}{(1+\delta)}$; A é a área glotal máxima; T o período; γ o quociente de velocidade; δ a simetria da forma de onda; e β o declive.

Exemplos de outros modelos interactivos são: o modelo de uma massa [26]; o modelo analítico desenvolvido para estudo da interacção fonte-tracto por Ananthapadmanabha e Fant [91]; o modelo utilizando parametrização da condutância de Rothemberg [114].

O modelo de duas massas, desenvolvido por Ishizaka e Flanagan [87], é o mais utilizado dos modelos interactivos mecânicos. O diagrama deste modelo encontra-se na Figura 12. O movimento das duas massas, m_1 e m_2, é controlado pelas forças aerodinâmicas e pelas forças mioelásticas, representadas por molas e amortecedores.

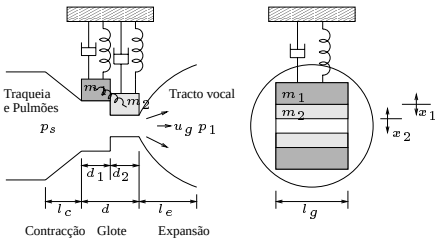


Figura 12 - Modelo de duas massas das cordas vocais [87].

As equações que controlam os movimentos são:

$$\begin{aligned} m_1\frac{d^2x_1}{dt^2} + r_1\frac{dx_1}{dt} + k_1x_1 + k_c(x_1-x_2) + F_1 &= 0 \\ m_2\frac{d^2x_2}{dt^2} + r_2\frac{dx_2}{dt} + k_2x_2 + k_c(x_2-x_1) + F_2 &= 0 \end{aligned}$$

onde x_i representa o deslocamento lateral das massas, F_i representa as forças aerodinâmicas exercidas em cada massa, r_i a resistência devida à viscosidade, $i = 1$ para a massa inferior, e $i = 2$ para a massa superior. No modelo

as molas possuem características não lineares. Durante a fase em que a glote se encontra fechada existe uma força de contacto. O valor da frequência fundamental neste modelo é controlado por um parâmetro, Q, representando a tensão das cordas.

O circuito acústico equivalente encontra-se na Figura 13. R_c representa a contracção abrupta à entrada; Rv_1 e Rv_2 representam as perdas por viscosidade no bordo inferior e superior das cordas, respectivamente; R_{12} representa a variação da energia cinética por unidade de volume na junção das duas massas; R_e a expansão; L_c, Lg_1 e Lg_2 as inertâncias das massas de ar nas três zonas. Os componentes são função da área da primeira secção do tracto vocal, $A1(t)$, e das áreas de abertura glotal de cada uma das massas, $Ag_1(t)$ e $Ag_2(t)$, obtidas com base nos deslocamentos laterais, relativamente a uma posição de repouso, através de

$$\begin{aligned} Ag_1(t) &= Ag_{0_1} + 2l_g x_1(t) \\ Ag_2(t) &= Ag_{0_2} + 2l_g x_2(t)\,, \end{aligned}$$

onde as áreas de repouso, Ag_{0_1} e Ag_{0_2}, são geralmente iguais. A impedância total pode ser representada por uma indutância, $L_g(t)$, em série com uma resistência, $R_g(t)$. A interacção entre o modelo glotal e o tracto é feita através da pressão supraglotal $p_1(t)$.

A.3 Modelos glotais fisiológicos

Os modelos desta categoria são geralmente utilizados em aplicações em que é necessário grande precisão, pois são muito exigentes computacionalmente [110, pág. 31]. Os modelos mais conhecidos são os desenvolvidos por Titze [115]. Outro modelo foi proposto por Hegerl [74] baseado na resolução numérica da equação de Navier-Stokes.

B. Modelo de fonte de ruído

Uma área suficientemente reduzida provoca a mudança de fluxo laminar para um regime turbulento. Para fluxo do ar turbulento, a analogia eléctrica não existe. No entanto, Stevens, Kasowski e Fant, em 1953, mostraram que inserindo um gerador de ruído no ponto da constrição a analogia pode ser mantida [21]. A potência e a resistência interna da fonte depende da área da constrição e do fluxo [10]. A inserção de uma fonte de ruído filtrado, no modelo baseado na analogia com linhas de transmissão, produziu resultados muito aceitáveis, para algumas fricativas pelo menos, apesar de ser apenas uma simples aproximação de um fenómeno complexo e não linear.

Uma abordagem baseada na teoria aeroacústica permitiu a Sinder [79] não considerar o fenómeno de excitação em sons fricativos como algo de separado da propagação e radiação.

Mais detalhes podem ser obtidos em: [108, pág. 42 e seguintes], [10, pág. 53], [6, pág. 413] e [116].

C. Modelamento da excitação glotal no SAP

Os requisitos base para o modelo da excitação foram: permitir o estudo da interacção entre a fonte e as cavidades supra-laríngeas; permitir o controlo directo de parâmetros como a frequência fundamental; contribuir para a obtenção de som sintético de qualidade natural; não ser demasiado pesado computacionalmente.

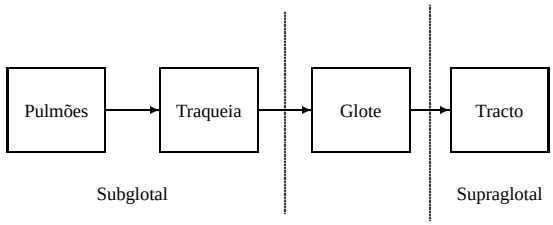


Figura 14 - Vários subsistemas intervenientes na obtenção da onda de excitação glotal.

C.1 Modelamento dos vários subsistemas

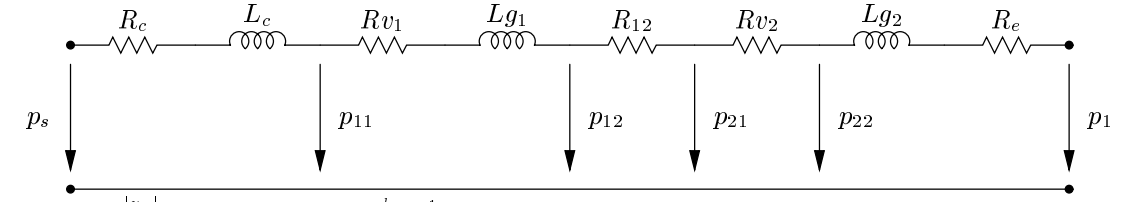
Para a obtenção da excitação glotal, $u_g(t)$, torna-se necessário modelar os vários subsistemas envolvidos: pulmões, cavidades subglotais, a glote e o tracto supraglotal.

O papel dos pulmões é o de fonte de pressão, quase constante, sendo representados no nosso modelo por uma fonte de pressão pulmonar p_p em série com uma resistência R_p. Um valor típico para a pressão pulmonar é o de $10\,cm\,H_2O$, aproximadamente igual a $10000\ dine/cm^2$. Usamos nos nossos estudos $R_p = 8\,\Omega\ cgs$.

Para a representação da parte subglotal, incluindo a traqueia, utilizamos a abordagem de Ananthapadmanabha e Fant [91], com três circuitos RLC ressonantes. A simulação da parte subglotal utilizando modelos similares aos utilizados no caso das cavidades supraglotais não foi tentada, devido à falta de informação detalhada acerca das dimensões.

Várias abordagens foram utilizadas para modelar as cordas vocais: modelos auto-oscilantes, modelos com área glotal parametrizada, etc. Pretendia-se um modelo que permitisse elevada qualidade e com bases fisiológicas, como o modelo de duas massas, (Figura 12), mas que fosse também não muito exigente em termos computacionais. O modelo deveria ainda permitir o controlo directo de parâmetros como a frequência fundamental. Foi utilizado o modelo proposto por Prado [35] em que se parametriza directamente as áreas glotais do modelo de duas massas.

Os sistemas que se encontram acima da glote podem ser modelados por uma impedância de entrada $z_e(t)$ (ou a pressão $p_{sup}(t)$ que se obtém pela convolução dessa impedância de entrada e o fluxo glotal) ou aproximados por uma cascata de circuitos RLC. A utilização da impedância de entrada permite modelar melhor as perdas dependentes da frequência [113, pág. 59]. Foi por isso escolhido este método para o nosso modelo. A impedância de entrada é obtida do modelo acústico das cavidades supralaríngeas. Interessa aqui referir que, na implementação efectuada do cálculo da impedância, é possível calcular a impedância de entrada, para sons nasais, desprezando a impedância de entrada do tracto nasal. Esta facilidade é da máxima utilidade para estudar o efeito adicional do acoplamento do tracto nasal nas características da onda de excitação glotal. No caso de não se pretender incluir o efeito de carga supraglotal $z_e = 0$.



$R_c = 1.37\frac{\rho}{2}\frac{|u_g|}{Ag_1^2}$ $\quad L_c = \int_0^{l_c}\frac{1}{A_c(x)}dx$ $\quad u_g$= fluxo glotal $\quad p_s$ = pressão subglotal (pulmonar)

$Rv_1 = 12\frac{\mu l_g^2 d_1}{Ag_1^3}$ $\quad Lg_1 = \frac{\rho d_1}{Ag_1}$ $\quad \rho$ = densidade do ar $\quad \mu$ = coeficiente de viscosidade

$R_{12} = \frac{\rho}{2}\left(\frac{1}{Ag_2^2} - \frac{1}{Ag_1^2}\right)|u_g|$ $\quad R_e = -\frac{\rho}{2}\frac{2}{Ag_2 A_1}\left(1 - \frac{Ag_2}{A_1}\right)|u_g|$ $\quad A_c$ = área na região de contracção

$Rv_2 = 12\frac{\mu l_g^2 d_2}{Ag_2^3}$ $\quad Lg_2 = \frac{\rho d_2}{Ag_2}$ $\quad A_1$ = área da primeira secção do tracto vocal

Figura 13 - Circuito equivalente do modelo de duas massas [82].

C.2 Circuito equivalente

Depois de efectuadas as escolhas para a forma de modelar cada um dos subsistemas envolvidos, obtemos o circuito apresentado na Figura 15.

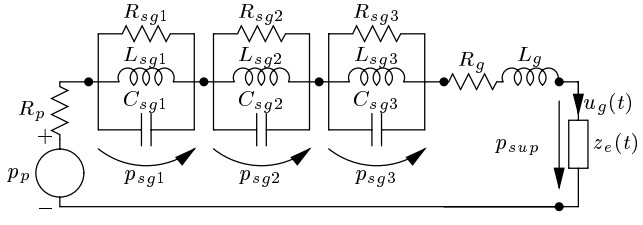


Figura 15 - Análogo eléctrico para obtenção da onda de excitação glotal, $u_g(t)$.

C.3 Modelo paramétrico das áreas

No modelo, a resistência R_g e indutância L_g, representando as cordas vocais, dependem da área de abertura glotal. Como já foi referido, optamos por utilizar um modelo paramétrico de duas massas baseado no trabalho de Prado [35].

As áreas glotais, $Ag_1(t)$ e $Ag_2(t)$, obtêm-se através de

$$Ag_1(t) = \begin{cases} A_1\left(0.5 - 0.5\cos\left(\frac{\pi t}{T_a}\right)\right) + Ag_0, & 0 < t < T_a \\ A_1\cos\left(\frac{\pi(t-T_a)}{2T_f}\right) + Ag_0, & T_a < t < T_a + T_f \\ Ag_0, & T_a + T_f < t < T_0 \end{cases}$$

$$Ag_2(t) = \begin{cases} Ag_0, & 0 < t < \tau \text{ ou } T_a + T_f + \tau < t < T_0 \\ A_2\left(0.5 - 0.5\cos\left(\frac{\pi(t-\tau)}{T_a}\right)\right) + Ag_0, & \tau < t < T_a + \tau \\ A_2\cos\left(\frac{\pi(t-T_a-\tau)}{2T_f}\right) + Ag_0, & T_a + \tau < t < T_a + T_f + \tau \end{cases}$$

em que T_0 é o período de excitação glotal ($T_0 = 1/F_0$); T_a a duração do movimento de abertura das cordas vocais; T_f o tempo que as cordas demoram a fechar; Ag_0 abertura mínima da glote; A_2 e A_1 aberturas máximas; e $\tau = \frac{\Phi T_0}{360}$ com Φ a diferença de fase entre Ag_1 e Ag_2.

O valor de T_a e T_f obtêm-se do quociente de abertura, OQ (do Inglês *Open Quotient*), e quociente de velocidade, SQ (do Inglês *Speed Quotient*), sendo $T_a = OQ \times T_0 \times SQ/(SQ+1)$ e $T_f = T_a/SQ$.

C.4 Cálculo de R_g e L_g

Com base nas áreas, $Ag_1(t)$ e $Ag_2(t)$, obtêm-se os valores para R_g e L_g, usando as expressões [37, equações 6 e 7]:

$$R_g = \frac{\rho}{2}\left[\frac{0.37}{Ag_1^2} + \frac{1 - 2\frac{Ag_2}{area_1}\left(1 - \frac{Ag_2}{area_1}\right)}{Ag_2^2}\right]|u_g| + 12\mu l_g^2\left(\frac{d_1}{Ag_1^3} + \frac{d_2}{Ag_2^3}\right) + R_{12}$$

$$L_g = \rho\left(\frac{d_1}{Ag_1} + \frac{d_2}{Ag_2}\right),$$

em que R_{12} representa as perdas causadas pela diferença de fase entre as duas massas e é dada por [37, equação 9a]

$$R_{12} = \frac{\rho}{2}\eta_{12}\left(\frac{1}{Ag_1} - \frac{1}{Ag_2}\right)^2|u_g|,$$

com

$$\eta_{12} = \begin{cases} 0.4 & \text{se } Ag_1 \geq Ag_2 \\ 1.0 & \text{se } Ag_1 < Ag_2 \end{cases}.$$

Nas equações anteriores Ag_1, Ag_2, $area_1$ e u_g são todas variáveis no tempo, tendo-se omitido essa dependência para simplificar as expressões.

C.5 Cálculo do fluxo glotal

As variações de pressão ao longo do circuito podem ser representadas por:

$$p_p - R_p u_g(t) - \sum_i p_{sgi} - \frac{d(L_g u_g(t))}{dt} - R_g u_g(t) - p_{sup}(t) = 0,$$

onde p_{sgk} , $k = 1, 2, 3$ se obtêm de,

$$u_g(t) = C_{sgk}\frac{d(p_{sgk}(t))}{dt} + \frac{p_{sgk}(t)}{R_{sgk}} + \frac{1}{L_{sgk}}\int_0^t p_{sgk}(\tau)d\tau.$$

A obtenção de $u_g(t)$, ou melhor do seu equivalente discreto $u_g(nT_s)$, implica a resolução das equações diferenciais. A resolução numérica destas equações consegue-se usando uma aproximação por equações diferença. As derivadas e integrais são aproximados por

$$\frac{d}{dt}f(t) \cong \frac{f(t_i) - f(t_{i-1})}{T_s} \quad \text{e} \quad \int_0^t f(t)dt \cong T_s\sum_{j=0}^{i-1} f(t_j)$$

onde $T_s = t_i - t_{i-1}$ representa o intervalo de amostragem.

A pressão supraglotal, $p_{sup}(n)$, pode ser obtida como a convolução da impedância de entrada do tracto $z_e(t)$, com o fluxo glotal $u_g(t)$ [113, pg. 60]: $p_{sup}(t) = u_g(t) * z_e(t)$. A convolução discreta pode ser decomposta na soma de dois termos, a saber:

$$p_{sup}(n) = \sum_{j=0}^{\infty} u_g(n-j)z_e(j) = u_g(n)z_e(0) + \sum_{j=1}^{\infty} u_g(n-j)z_e(j).$$

C.6 Irregularidades

A forma de onda em períodos sucessivos de $u_g(t)$ não é igual. Na literatura sobre o assunto aparecem termos como *jitter*, variação aleatória de período para período na duração do período; *shimmer*, referente à variação da amplitude do pulso glotal entre dois períodos consecutivos; diplofonia; etc.

O modelo da fonte inclui a possibilidade de modelar algumas irregularidades, nomeadamente flutuações da frequência fundamental e da abertura máxima da glote. Ao valor, obtido por interpolação, de F_0 é adicionado um valor dependente do parâmetro *jitter*

$$F_{0_{com\ jitter}} = F_0 + random \times 2 \times F_0 \times jitter/100.0\ ,$$

sendo $random$ um valor aleatório entre -0.5 e 0.5. O valor de T_0 correspondente é depois arredondado para o instante de amostragem mais próximo. A implementação é semelhante para o *shimmer*.

C.7 Aspiração

Foi também incluído no modelo a geração de ruído de aspiração, seguindo as propostas de Sondhi e Schroeter [37]. Utilizando a área glotal Ag_2 e o fluxo glotal, é calculado o quadrado do número de Reynolds, Re^2, segundo a fórmula

$$Re^2(n) = \left[\frac{l_g \rho}{\mu A g_2} u_g(n)\right]^2$$

e o fluxo resultante da aspiração, que apenas existe para $Re^2 > Re^2_{critico}$, é igual a

$$u_{asp}(n) = G \times random(n) \frac{Re^2 - Re^2_{critico}}{R_g(n)}\ ,$$

com $G = 0.5 \times 10^{-7}$, $Re^2_{critico} = 2700^2$ e $random(n)$ um número aleatório entre -0.5 e 0.5.

VII. Obtenção da posição dos articuladores

> There is a set of fundamental problems the solutions to which require the determination of vocal tract shape from the acoustic parameters of the speech signal.
>
> Victor N. Sorokin [117]

Os modelos descritos permitem definir configurações do aparelho de produção e obter os sons correspondentes, designado por problema directo. Iremos agora tratar do problema inverso, obter a posição dos articuladores com base no sinal de voz.

Desde a década de 1950, quando se efectuaram as primeiras simulações em computador de produção de voz, que os investigadores desejam conhecer a forma do tracto vocal ao longo do tempo. Este conhecimento é muito importante para os Linguistas, tendo também aplicações em Engenharia e Medicina. Em Engenharia oferece possibilidades de melhorar a codificação, reconhecimento e síntese de fala. No entanto, um processo de análise para obter a área é necessário para tornar essas possibilidades realidade.

Dados sobre o tracto vocal são essenciais. A teoria acústica de produção de voz [24] considera o tracto como um tubo com área variável. Existem basicamente dois tipos de métodos para obter a área: medição directa ou estimação com base no sinal acústico.

A. Medição directa

Existem técnicas que permitem obter imagens de zonas completas do aparelho de produção humano, outras que apenas possibilitam seguir o comportamento no tempo de alguns pontos e, ainda, técnicas que permitem medição de comportamentos complexos como a área de contacto da língua com o palato [118, capítulo 1].

Os métodos directos baseados em raios-X, utilizados já por investigadores como Chiba e Kajiyama [19] e Fant [24], constituem ainda fonte de dados para muitos sintetizadores. Infelizmente esta técnica é laboriosa, necessita de dosagens elevadas, e apenas nos dá imagens bidimensionais. Em especial, pelo risco que a radiação representa, raramente é empregue actualmente.

Mais recentemente, começaram a usar-se técnicas como a Tomografia Axial Computorizada e Ressonância Magnética [53]. A Ressonância Magnética não sofre dos problemas dos raios-X, parece ser a melhor técnica para recolha dos dados necessários. As desvantagens advêm do facto de problemas de resolução e dificuldades com estruturas calcificadas com pouco hidrogénio móvel que não se distinguem. As imagens obtidas por esta técnica têm sido essencialmente de configurações estáticas do tracto. Foram efectuadas medições para vogais [119], de fricativas [120], [119], liquidas [121], laterais [122], do tracto nasal [56], [72] e da posição do velo [123]. Desenvolvimentos recentes, dos equipamentos associados a técnicas de processamento, permitem obter imagens dinâmicas [124]. Imagens obtidas utilizando endoscopia podem também fornecer informação útil [125].

Técnicas como a articulografia electromagnética, designada por EMA ou EMMA, aparecendo o segundo M para explicitar que as imagens são no plano sagital médio (em Inglês *Midsagittal*), permitem obter informação acerca da posição de um conjunto de pontos ao longo do tempo [126], [127]. Têm sido utilizadas para estudar o comportamento da língua, lábios e, em alguns casos, do velo [128], [129]. Os sistemas actualmente em funcionamento apenas permitem obtenção de dados acerca de pontos situados num mesmo plano sagital, limitação que será corrigida pelos novos sistemas 3D em fase final de desenvolvimento [130].

Para a medição de certos articuladores foram desenvolvidos métodos especiais. Exemplo deste caso é o *velotrace* desenvolvido para medir directamente a posição do velo [131].

Uma técnica recente que promete ser da máxima utilidade é a proposta por Burnett [132], baseada na utilização de um

radar de baixa potência. Esta técnica pode ser utilizada para obter informação, tanto acerca da área de abertura glotal, como acerca da posição da língua durante a produção de voz.

B. Métodos indirectos

O mapeamento do domínio acústico para o domínio articulatório consiste em estimar a forma do tracto vocal, usando apenas o sinal acústico. Nesta secção apresenta-se um resumo da investigação efectuada no passado e no presente.

As várias abordagens podem agrupar-se nas seguintes classes: métodos analíticos, métodos de procura em tabelas, redes neuronais e métodos de optimização usando realimentação.

Uma das maiores dificuldades do mapeamento acústico-articulatório é a presença de múltiplas soluções. Este facto foi já bem documentado teórica e empiricamente [10], [24].

B.1 Métodos analíticos

Vários investigadores propuseram métodos analíticos para obter a função de área com base no sinal acústico. As técnicas baseiam-se nos coeficientes de Predição Linear (LPC) ou resposta impulsional do tubo. A abordagem LPC deriva as áreas dos coeficientes de reflexão obtidos por filtragem inversa [28], [29]. O grande problema desta abordagem é o "efeito de ventríloquo", diferentes formas podem produzir as mesmas formantes [39], [133].

O método utilizando a resposta impulsional nos lábios, designado por LIR, assenta no facto de que se a função de transferência do tracto vocal é conhecida então $A(x)$ pode ser obtida [134], [30].

O LIR possui duas limitações principais: é necessário saber as condições de fronteira na glote; para a estimação dos valores principais (*eigenvalues*) é necessário usar um período de tempo longo (10 a 20 ms), perdendo-se a estacionaridade. O LPAT tem diversos problemas: incerteza na fonte; a função de área obtida não é única; exclui nasais e sons surdos; a condição de fronteira, não considerando a carga de radiação, é irrealista; não considera perdas; obriga a estimar correctamente o comprimento do tracto vocal.

B.2 Métodos de procura em tabelas

Faz-se a amostragem dos parâmetros articulatórios e constroem-se tabelas (*codebooks*) com as formas do tracto vocal e as respectivas representações acústicas [39]. Schroeter, Meyer e Parthasarathy [135] fizeram alguns melhoramentos para a geração dos *codebooks* e utilizaram procura por programação dinâmica. Muitos autores se dedicaram, e dedicam (por exemplo [136]), ao melhoramento do processo de geração das tabelas.

Esta técnica tem diversos problemas: carga computacional; sensibilidade à fonte de excitação; ambiguidade no mapeamento; limitações do modelo acústico [137]. Estas técnicas são geralmente adequadas para derivar configurações iniciais dos articuladores para processos de inversão baseados em optimização.

B.3 Redes neuronais

Uma abordagem, mais recente, consiste na aplicação de redes neuronais. A rede é treinada com um conjunto grande de parâmetros acústicos e articulatórios. Um padrão acústico é depois utilizado para obter o padrão articulatório correspondente [138], [70], [139]. No entanto, o processo de treino continua um desafio [138] e não existem ainda vantagens claras destas técnicas em relação a outras [137]. A capacidade actual das redes neuronais é a de fornecer valores iniciais para os parâmetros articulatórios, para um pequeno conjunto de treino. Têm sido utilizados vários tipos de redes.

B.4 Métodos de optimização usando realimentação

Tentativas para uma resolução analítica do problema produziram resultados insatisfatórios [29]. As soluções podem não ser únicas, devido ao sinal ser limitado em frequência e a assunção de ondas planas não ser válida a frequências elevadas [134]. Têm portanto de se utilizar métodos numéricos. Os processos convencionais de optimização são processos iterativos, tipicamente utilizando alguma forma de procura por gradiente. Técnicas não usando gradiente, como o *simulated annealing* [85] e algoritmos genéticos [140] são também usadas.

A optimização pode ser feita em apenas uma *frame* temporal, ao longo de várias [141], ao longo de uma trajectória parametrizada [142], ou em termos de configurações alvo (em Inglês *targets*) [140].

Nos métodos utilizando apenas uma *frame*, o sinal é dividido em secções de 5 a 40 ms, onde se pode considerar o sinal como estacionário. Esta forma, conveniente, é a usada na maioria dos trabalhos publicados. A configuração do tracto vocal é estimada independentemente para cada *frame*. Não se utiliza o facto de a configuração do tracto vocal variar lentamente ao longo do tempo.

Se os parâmetros forem estimados conjuntamente para várias *frames*, a correlação entre elas pode ser explorada através de restrições ou parametrização das trajectórias dos articuladores. As trajectórias podem representar o tracto vocal ao longo de muitas *frames* eficientemente e podem aliviar o problema do mapeamento não ser unívoco [143], [142]. O problema é que a eficácia dos processos de optimização diminui com o aumento do número de parâmetros a optimizar, a chamada "praga da dimensionalidade".

Como alternativa, o movimento dos articuladores pode ser representado como um sistema dinâmico de que se estimam as entradas. Desta forma, o movimento pode ser representado por alvos, facilmente relacionados com fonemas, ou *gestures* [142], [140].

As representações da entrada e saída influenciam os resultados da optimização através do significado das propriedades, a dimensão das propriedades, a métrica escolhida para o erro e as restrições (em Inglês *constraints*). Devido à existência de várias soluções e para evitar mínimos locais utilizam-se uma variedade de restrições, técnicas de inicialização e regularização.

Têm sido empregues processos de optimização como: algoritmo de Hookes e Jeeves [40], [142], [141]; algoritmo de gradiente óptimo [144]; combinações do método de Fletcher-Reeves e aproximações sucessivas [35]; métodos

estocásticos como algoritmos genéticos [140] e *simulated annealing* [85].

Muitas representações foram propostas para representar a configuração do tracto de uma forma mais eficiente do que utilizando directamente a função de área. Foi, por exemplo, utilizada a decomposição em série de Fourier da forma da língua e da função de área [145]. Os modelos articulatórios sagitais de Mermelstein [33] e Coker [50] são os mais populares, mas muitos outros são utilizados. Mesmo os modelos paramétricos de área continuam a ser utilizados [43].

Muitas representações do domínio acústico e métricas foram utilizadas: a distância Euclidiana entre as primeiras 3 a 5 frequências [35], distâncias espectrais lineares e logarítmicas [40], distâncias cepstrais [146]. Outras representações incluem: distância Euclidiana entre o logaritmo das frequências das formantes, algumas vezes complementados com a amplitude das formantes [117] e distâncias LPC. Recentemente, foram utilizadas medidas de erro utilizando informação perceptual, como a distância das 3 primeiras formantes em Barks [147]. Alguns investigadores propuseram critérios de erro múltiplos [142].

Nem sempre adicionar mais informação acústica conduz a melhores resultados, como Sorokin descobriu, ao utilizar os logaritmos das primeiras 3 e 4 frequências com piores resultados no segundo caso [117].

B.5 Mapeamento acústico-articulatório de consoantes e sons nasais

Geralmente, apenas se tem abordado a inversão de sons sonoros orais não obstruentes. Pouco trabalho foi feito para consoantes e sons nasais.

Sorokin [148] e Shirai abordaram a inversão de fricativas surdas com resultados razoáveis. O caso das fricativas sonoras foi abordado, recentemente, por Riegelsberger [108]. A obtenção do lugar de constrição para oclusivas foi investigada por Galván-Rodríguez [149].

Em relação aos sons nasais é também reduzido o número de trabalhos, estando o problema longe de resolvido. Foi estudada a obtenção da posição do velo utilizando uma rede neuronal treinada com dados de articulografia electromagnética (EMMA) por Richmond [150]. Rossato efectuou experiências relacionadas com a inversão de vogais nasais, mas não utilizou voz natural, apenas propriedades acústicas derivadas do próprio modelo articulatório usado [151].

VIII. Aplicações

Têm sido desenvolvidos ao longo dos anos vários sintetizadores articulatórios. A variedade de técnicas, aplicações a que se destinam e limitações é muito grande.

Apesar do estado de desenvolvimento deste tipo de sintetizadores não os tornar ainda utilizáveis em sistemas comerciais de conversão de texto em fala existem já primeiros protótipos deste tipo [152], [142], [153], [31], [154].

Muitos dos sintetizadores articulatórios têm sido utilizados em estudos de Fonética e Fonologia. Sem o acesso a esta ferramenta teria sido muito difícil, ou mesmo impossível, o aparecimento e desenvolvimento de algumas destas teorias que tanto têm contribuído para o aumento de conhecimento acerca dos processos de produção e percepção de voz. Têm sido propostas, nos últimos anos, teorias fonológicas baseadas total ou parcialmente na utilização de descrições articulatórias. Exemplos destas teorias são a Fonologia Articulatória [155], [156], desenvolvida fazendo uso do sintetizador CASY dos laboratórios Haskins [11] e a Fonologia Funcional proposta por Boersma [59] que levou o autor a desenvolver um sintetizador articulatório muito completo.

Outra área de aplicação de interesse é a sua utilização em codificação de voz. O reduzido número de articuladores, aliado à lenta variação no tempo das suas posições, torna-os candidatos ideais para codificação. O grande problema reside na obtenção automática dos articuladores apenas com base no sinal acústico, problema ainda não completamente resolvido. Exemplo deste tipo de aplicação é o projecto de desenvolvimento de um *voice mimic* por J. Flanagan e colaboradores [136], [157]-[159].

Tem, também, sido estudada a forma de tornar um sintetizador articulatório capaz de aprender a falar, imitando o processo de aprendizagem de uma criança [160]

Embora não relacionado com a voz de uma forma directa, as técnicas referentes ao modelamento acústico do tracto são úteis no modelamento de instrumentos musicais de sopro [89, por exemplo]. Neste tipo de aplicação, as técnicas baseadas em filtros de onda digitais são geralmente as utilizadas, devido às paredes rígidas dos instrumentos musicais.

Referências

[1] Gunnar Fant, "What can basic research contribute to speech synthesis", *Journal of Phonetics*, vol. 19, pp. 75–90, 1991.

[2] D.H. Klatt, "Software for a cascade/parallel formant synthesizer", *Journal of the Acoustic Society of America*, vol. 67, no. 3, pp. 971–995, Março de 1980.

[3] Christian Benoît, "Speech synthesis: Present and future", em *The Landscape of Future Education in Speech Communication Sciences*, Gerrit Bloothooft et al., Ed., pp. 119–123. OTS Publications, 1997.

[4] L. C. Oliveira, *Síntese de Fala a Partir do Texto*, Dissertação de doutoramento, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, Outubro de 1996.

[5] D. H. Klatt, "Review of text-to-speech conversion for english", *J. Acoust. Soc. Am.*, vol. 82, no. 3, pp. 737–793, 1987.

[6] D. Childers, *Speech Processing and Synthesis Toolboxes*, 2000.

[7] R. Sproat, Ed., *Multilingual Text-to-Speech Synthesis: the Bell Labs Approach*, Kluwer, 1998.

[8] T. Dutoit, *An Introduction to Text-to-Speech Synthesis*, Kluwer Academic Publisher, 1997.

[9] J. P. H. van Santen, J. Olive, J. Hirschberg, e R. Sproat, Eds., *Speech Synthesis*, Springer-Verlag, 1996.

[10] James L. Flanagan, *Speech Analysis, Synthesis and Perception*, Springer-Verlag, New York, 1972.

[11] Philip Rubin, Thomas Baer, e Paul Mermelstein, "An articulatory synthesizer for perceptual research", *J. Acoust. Soc. Am.*, vol. 70, no. 2, pp. 321–328, Agosto de 1981.

[12] Shinji Maeda, "The role of the sinus cavities in the production of nasal vowels", em *Proc. ICASSP*, 1982, pp. Vol. 2, 911–914.

[13] M. M. Sondhi e J. Schroeter, "A hybrid time-frequency domain articulatory speech synthesizer", *IEEE Trans. Acoust. Sp. Sig. Proc.*, 1987.

[14] T. Koizumi, S. Tanigushi, e S. Hiromitsu, "Glottal source-vocal tract interaction", *J. Acoust. Soc. Am.*, vol. 78, no. 5, pp. 1541–1547, 1985.

[15] R. Linggard, *Electronic Synthesis of Speech*, Cambridge University Press, 1985.

[16] Homer Dudley e T. H. Tarnoczy, "The speaking machine of Wolfgang von Kempelen", *J. Acoust. Soc. Am.*, vol. 22, no. 2, pp. 151–166, Março de 1950.

[17] J. Q. Steward, "An electrical analog of the vocal organs", *Nature*, vol. 110, pp. 311–312, 1922.

[18] Franklin S. Cooper, Alvin M. Liberman, e John M. Borst, "The interconversion of audible and visible patterns as a basis for research in the perception of speech", em *Proc. Natl. Acad. Sci.*, 37, Ed., 1951, pp. 318–325.

[19] Tsutomu Chiba e Masato Kajiyama, *The Vowel, its nature and structure*, Phonetic Society of Japan, Tokyo, 1958.

[20] H. K. Dunn, "The calculation of vowel ressonances, and an electrical vocal tract", *J. Acoust. Soc. Am.*, vol. 22, pp. 740–753, 1950.

[21] K. N. Stevens, S. Kasowski, e C. Gunnar Fant, "An electrical analog of the vocal tract", *J. Acoust. Soc. Am.*, vol. 25, pp. 734–742, 1953.

[22] G. Rosen, "Dynamic analog speech synthesizer", *J. Acoust. Soc. Am.*, vol. 30, no. 3, pp. 201–209, 1958.

[23] J. van den Berg, "An electrical analog of the trachea, lungs and tissues", *Acta Physiol. Pharmacol. Neerl.*, vol. 9, pp. 361–385, 1960.

[24] G. Fant, *Acoustic theory of speech production*, Mouton and Co., Gravenhage, The Netherlands., 1960.

[25] John L. Kelly Jr. e Carol C. Lochbaum, "Speech synthesis", em *Proc. Fourth Intern. Congr. Acoust., Paper G42*, 1962, vol. 22, pp. 1–4.

[26] James L. Flanagan e Lorinda L. Landgraf, "Self-oscilating source for vocal-tract synthesizers", *IEEE Trans. Audio and Electr.*, vol. AU-16, no. 1, pp. 57–64, Março de 1968.

[27] J. L. Flanagan e L. Cherry, "Excitation of vocal-tract synthesizers", *J. Acoust. Soc. Am.*, vol. 45, no. 3, pp. 764–769, 1969.

[28] Hishashi Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms", *IEEE Trans. Audio and Electr.*, vol. AU-21, no. 5, October de 1973.

[29] Hishashi Wakita, "Estimation of vocal-tract shapes from acoustical analysis of the speech wave: The state of the art", *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-27, no. 3, pp. 281–285, Junho de 1979.

[30] M. M. Sondhi e J. R. Resnick, "The inverse problem for the vocal tract: Numerical methods, acoustical experiments, and speech synthesis", *J. Acoust. Soc. Am.*, vol. 73, no. 3, pp. 985–1002, Março de 1983.

[31] Cecil H. Coker, "Synthesis by rule from articulatory parameters", em *Proc. 1967 Conference Speech Commun. Process.* IEEE, 1967, pp. 52–53.

[32] W. L. Henke, *Dynamic articulatory model of speech production using computer simulation*, Phd thesis, MIT, Cambridge, MA, 1966.

[33] P. Mermelstein, "Articulatory model for the study of speech production", *J. Acoust. Soc. Am.*, vol. 53, no. 4, pp. 1070–1082, 1973.

[34] J. Liljencrants, *Speech Synthesis with a Reflection-Type Line Analog*, Ds dissertation, Dept. of Speech Comm. and Music Acoust., Royal Inst. of Tech., Stockolm, Sweden, 1985.

[35] Pedro P. L. Prado, *A Target-Based Articulatory Synthesizer*, PhD thesis, University of Florida, 1991.

[36] M. G. Rahim e C. C. Goodyear, "Estimation of vocal tract filter parameters using a neural net", *Speech Communication*, vol. 9, no. 1, pp. 49–55, Fevereiro de 1990.

[37] Juergen Schroeter e Man Mohan Sondhi, "Speech coding based on physiological models of speech production", em *Advances in Speech Siganl Processing*, Sadaoki Furui e Man Mohan Sondhi, Eds., pp. 231–268. Marcel Dekker,Inc., New York, 1992.

[38] Kenneth N. Stevens e Arthur S. House, "Development of a quantitative description of vowel articulation", *J. Acoust. Soc. Am.*, vol. 27, pp. 484–493, 1955.

[39] B. S. Atal, J. J. Chang, M. V. Mathews, e J. W. Tukey, "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique", *J. Acoust. Soc. Am.*, vol. 63, no. 5, pp. 1535–1555, 1978.

[40] J. L. Flanagan, K. Ishizaka, e K. L. Shipley, "Signal models for low bit-rate coding of speech", *J. Acoust. Soc. Am.*, vol. 68, no. 3, pp. 780–791, 1980.

[41] Qiguang Lin, *Speech production theory and Articulatory Speech Synthesis*, PhD thesis, Dept. of Speech Comm. & Music Acoustics, Royal Institute of Technology (KTH), Stockolm, Sweden, 1990.

[42] Zhenli Yu, "A method to determine the area function of speech based on perturbation theory", *Speech Transmission Laboratory, Quarterly Progress ans Status Report*, vol. STL-QPSR 4, pp. 77–95, 1993.

[43] Mats Båvegård, "Towards an articulatory speech synthesizer: Model development and simulations", *Speech, Music and Hearing, Quarterly Progress and Status Report*, vol. TMH-QPSR 1, pp. 1–15, 1996.

[44] William J. Hardcastle e Alain Marchal, Eds., *Speech Production and Speech Modelling*, NATO Advanced Science Institute Series: D, Volume 55. Kluwer Academic Publishers, Dordrecht, Junho de 1990.

[45] M. Båvegård, "Introducing a parametric consonantal model to the articulatory speech synthesizer", em *Proc. Eurospeech*, Madrid, Setembro de 1995, pp. III, 1857–1860.

[46] Soumya Bouabana, *Modélisation des Mouvements Articulatoires de la Langue par la Méthode de la LPC Multi-impulsionnelle*, Mémoire de soutenance de thése, ENST, Paris, 1995.

[47] J. S. Perkell, *A physiologically-oriented model of tongue activity in speech production*, Phd thesis, MIT, Cambridge, MA, 1974.

[48] Elliot Saltzman e K. Munhall, "A dynamic approach to gestural patterning in speech production", *Ecological Psychology*, vol. 1/3, pp. 333–382, 1989.

[49] Shinji Maeda, "A digital simulation method of vocal-tract system", *Speech Communications*, vol. 1, pp. 199–229, 1982.

[50] Cecil H. Coker, "A model of articulatory dynamics and control", *Proc. IEEE*, vol. 64, no. 4, pp. 452–460, Abril de 1976.

[51] J. M. Heinz e K. N. Stevens, "On the derivation of area functions and acoustic spectra from cineradiographic films of speech", *J. Acoust. Soc. Am.*, vol. 36(A), pp. 1037–1038, 1964.

[52] J. Sundberg, C. Johansson, H. Wilbrand, e Christer Ytterbergh, "From sagittal distance to area - A study of transverse, vocal tract cross-sectional area", *Phonetica*, vol. 44, pp. 76–90, 1987.

[53] T. Baer, J. C. Gore, L. C. Gracco, e P. W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging:

Vowels", *J. Acoust. Soc. Am.*, vol. 90, no. 2 (Pt 1), pp. 799–828, Agosto de 1991.

[54] Denis Beautemps, Pierre Badin, e Rafael Laboissiére, "Deriving vocal-tract area functions from midagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data", *Speech Communication*, vol. 16, no. 1, pp. 27–47, 1995.

[55] Peter Ladefoged, James Anthony, e Cordell Riley, "Direct measurement of the vocal tract", *J. Acoust. Soc. Am.*, vol. 49, no. 1 (Pt 1), pp. 1971, 1971.

[56] J. Dang e K. Honda, "MRI measurements and acoustic of the nasal and paranasal cavities", *J. Acoust. Soc. Am.*, 1994.

[57] Olov Engwall, "Modelling of the vocal tract in three dimensions", em *Proc. Eurospeech*, Géza Gordos e Géza Németh, Eds., Budapest, Hungary, Setembro de 1999, vol. 1, pp. 113–116.

[58] Reiner Wilhelms-Tricarico, "Physiological modeling of speech production: Methods for modeling soft-tissue articulators", *J. Acoust. Soc. Am.*, vol. 97, no. 5, pp. 3085–3098, 1995.

[59] Paul Boersma, *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*, Hollan Academic Graphics (HAG), 1998.

[60] Arthur S. House e Kenneth S. Stevens, "Analog studies of the nasalization of vowels", *Journal of Speech and Hearing Disorders*, vol. 21, no. 2, pp. 218–232, Junho de 1956.

[61] Osamu Fujimura e Jan Ludqvist, "Sweep-tone measurements of vocal-tract characteristics", *J. Acoust. Soc. Am.*, vol. 49, no. 2 (Pt. 2), pp. 541–558, 1971.

[62] M. H. L. Hecker, "Dynamic analog of the nasal cavities", *Quarterly Progress Report, Research Laboratory of Technology, M. I. T.*, vol. 62, pp. 196–197, Julho de 1961.

[63] Michael H. L. Hecker, "Studies of nasal consonants with an articulatory speech synthesizer", *J. Acoust. Soc. Am.*, vol. 34, no. 2, pp. 179–188, Fevereiro de 1962.

[64] D. G. Childers e C. Ding, "Articulatory synthesis: nasal sounds and male female voices", *Journal of Phonetics*, vol. 19, pp. 453–464, 1991.

[65] Gloria J. Borden, Katherine S. Harris, e Lawrence J. Raphael, *Speech Science Primer - Physiology, Acoustics, and Perception of Speech*, William-Wilkins, 2 edição, 1994.

[66] G. Feng, "Modélisation acoustique et traitement du signal de parole: le cas des voyelles nasales et la simulation des poles et des zéros", *Bull. Lab. Commun. Parlée, Grenoble*, vol. 16, pp. 1–102, 1987.

[67] Gang Feng e Eric Castelli, "Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization", *J. Acoust. Soc. Am.*, vol. 99, no. 6, pp. 3694–3706, 1996.

[68] Marilyn Y. Chen, "Acoustic correlates of english and french nasalized vowels", *J. Acoust. Soc. Am.*, vol. 102, no. 4, pp. 2360–2370, 1997.

[69] Kenneth N. Stevens, *Acoustic Phonetics*, Current Studies in Linguistics. MIT Press, 1998.

[70] Mats Båvegård, Gunnar Fant, Jan Gauffin, e Johan Liljencrants, "Vocal tract sweeptone data and model simulations of vowels, laterals and nasals", *Speech Transmission Laboratory, Quarterly Progress ans Status Report*, vol. STL-QPSR 4, pp. 43–75, 1993.

[71] Qiguang Lin, "A three-channel model for nasals and nasalization", *J. Acoust. Soc. Am.*, vol. 95, no. 5, Pt 2, pp. 2922, Maio de 1994.

[72] Brad Hudson Story, *Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract*, PhD thesis, University of Iowa, 1995.

[73] T. J. Thomas, "A finite element model of fluid flow in the vocal tract", *Computer Speech and Language*, vol. 1, pp. 131–151, 1986.

[74] Gabriele C. Hegerl e Harald Höge, "Numerical simulation of the glottal flow by model based on the compressible navier-stokes equations", em *Proc. ICASSP*, 1991, pp. 477–480.

[75] Man Mohan Sondhi, "Resonances of a bent vocal tract", *J. Acoust. Soc. Am.*, vol. 79, no. 4, pp. 1113–1116, Abril de 1986.

[76] Thomas D. Rossing e Neville H. Fletcher, *Principles of Vibration and Sound*, Springer-Verlag, 1994.

[77] Philip McCord Morse, *Vibration and Sound*, Acoustical Society of America, 2nd edição, 1991.

[78] Philip McCord Morse e K. Uno Ingard, *Theorectical Acoustics*, McGraw Hill, New York, 1968.

[79] Daniel Jared Sinder, *Speech Synthesis using an aeroacoustic fricative model*, PhD thesis, Rutgers, The State University of New Jersey, October de 1999.

[80] Arthur Gordon Webster, "Acoustical impedance, and the theory of horns and the Phonograph", *Proceeding of the National Academy of Sciences of the United States of America*, vol. 5, pp. 275–282, 1919.

[81] H. Wakita e Gunnar Fant, "Toward a better vocal tract model", *Speech Transmission Laboratory, Quarterly Progress ans Status Report*, vol. STL-QPSR 1, pp. 9–29, 1978.

[82] J. L. Flanagan, K. Ishizaka, e K. L. Shipley, "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", *The Bell System Technical Journal*, vol. 54, no. 3, pp. 485–506, Março de 1975.

[83] James L. Flanagan e Kenzo Ishizaka, "Automatic generation of voiceless excitation in a vocal cord - vocal tract speech synthesizer", *IEEE Trans. Ac. Sp. Sig. Proc*, vol. ASSP-24, no. 2, pp. 163–170, Abril de 1976.

[84] Kenzo Ishizaka, J. C. French, e James L. Flanagan, "Direct determination of vocal tract wall impedance", *IEEE Trans. Ac. Sp. Sig. Proc*, vol. ASSP-23, no. 4, pp. 370–373, Agosto de 1975.

[85] Yu-Fu Hsieh, *A Flexible and High Quality Articulatory Speech Synthesizer*, PhD thesis, University of Florida, 1994.

[86] Pierre Badin e Gunnar Fant, "Notes on vocal tract computation", *Speech Transmission Laboratory, Quarterly Progress ans Status Report*, vol. STL-QPSR 2-3, pp. 53–108, 1984.

[87] K. Ishizaka e J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords", *The Bell System Technical Journal*, vol. 51, pp. 1233–1268, 1972.

[88] M. M. Sondhi, "Model for wave propagation in a lossy vocal tract", *J. Acoust. Soc. Am.*, vol. 55, no. 5, pp. 1070–1075, Maio de 1974.

[89] Gary Paul Scavone, *An Acoustic Analysis of Single-Reed Woodwind Instruments with an Emphasis on Design and Performance Issues and Digital Waveguide Modeling Techniques*, Phd thesis, Stanford University, 1997.

[90] K. Ishizaka, M. Matsudaira, e T. Kaneko, "Input acoustic-impedance measurement of the subglottal system", *J. Acoust. Soc. Am.*, vol. 60, no. 1, pp. 190–197, Julho de 1976.

[91] T. V. Ananthapadmanabha e G. Fant, "Calculation of true glottal flow and its components", *Speech Communication*, vol. 1, pp. 167–187, 1982.

[92] Philip McCord Morse, *Vibration and Sound*, McGraw Hill Book Co., New York, 2nd edição, 1948.

[93] Stanley J. Farlow, *Partial Differential Equations for Scientists and Engineers*, Dover Publications, Inc., New York, 1993.

[94] L. R. Rabiner e R. W. Schafer, *Digital processing of speech signals*, Prentice-Hall, Englewood Cliffs,NJ, 1978.

[95] Hans Werner Strube, "Time-varying wave digital filters and vocal-tract models", em *Proc. ICASSP*, 1982, pp. 923–926.

[96] Stuart Lawson e Ahmad Mirzai, *Wave Digital Filters*, Ellis Horwood Limited, Chichester, UK, 1990.

[97] Peter Meyer, Reiner Wilhelms, e Hans Werner Strube, "A quasiarticulatory speech synthesizer for german language running in real time", *J. Acoust. Soc. Am.*, vol. 86, no. 2, pp. 523–539, Agosto de 1989.

[98] G. T. H. Wright e F. J. Owens, "An optimized multirate sampling technique for the dynamic variation of vocal tract length in the Kelly-Lochbaum speech synthesis model", *IEEE Trans. Speech Audio Proc.*, vol. 1, no. 1, pp. 109–113, Janeiro de 1993.

[99] Vesa Välimäki, *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters*, Phd thesis, Helsinky University of Technology, Finland, 1995.

[100] Vesa Välimäki, Matti Karjalainen, e Timo Kuisma, "Articulatory speech synthesis based on fractional delay waveguide filters", em *Proc. ICASSP*, 1994, pp. 585–588.

[101] Enrico L. Bocchieri e Donald G. Childers, "Interactive graphics editor permits study of animated speech articulation", *Speech Technology*, pp. 10–14, Janeiro/Fevereiro de 1984.

[102] A. Teixeira, *Síntese Articulatória das Vogais Nasais do Português Europeu*, Tese de doutoramento, Universidade de Aveiro, 2000.

[103] Roland van Praag, *Formulation variationelle du champ sonore dans une arborescence de conduits non uniformes. Application à l'appareil vocal*, Dissertation docteur en sciences, Université Libre de Bruxelles, 1997.

[104] Paul Jospa e Roland Van Praag, "Sound field computaion in a network of non uniform ducts. Application to the vocal tract", em *Proc. ICPhS*, 1999, pp. 2141–2144.

[105] Samir El-Masri, Xavier Pelorson, Pierre Saguet, e Pierre Badin, "Vocal tract acoustics using the transmission line matrix (TLM) method", em *Proc. ICSLP*, 1996.

[106] G. Richard, M. Liu, D. Sinder, H. Duncan, Q. Lin, J. Flanagan, S. Levinson, D. Davis, e S. Slimon, "Numerical simulations of fluid flow in the vocal tract", em *Proc. Eurospeech*, Madrid, Spain, 1995.

[107] D. Sinder, Richard G, H. Duncan adn J. Flanagan, M. Krane, S. Levinson, S. Slimon, e D. Davis, "Flow visualization in stylized vocal tracts", em *Proc. ASVA*, Tokyo, Japan, Abril de 1997.

[108] Edward L. Riegelsberger, *The Acoustic-to-Articulatory Mapping of Voiced and Fricated Speech*, PhD thesis, The Ohio State University, 1997.

[109] Matteo Frigo, *FFTW 1.1 User's Manual*, Massachusetts Institute of Technology, 1997.
URL: *http://theory.lcs.mit.edu/~fftw*

[110] Kathleen E. Cummings e Mark A. Clements, "Glottal models for digital speech processing: A historical survey and new results", *Digital Signal Processing*, vol. 5, pp. 21–42, 1995.

[111] A. E. Rosemberg, "Effect of glottal pulse shape on the quality of natural vowels", *J. Acoust. Soc. Am.*, vol. 49, no. 2, Part 2, pp. 583–590, 1971.

[112] Gunnar Fant, Johan Liljencrants, e Qi-Guang Lin, "A four-parameter model of glottal flow", *Speech Transmission Laboratory, Quarterly Progress ans Status Report*, vol. STL-QPSR 4, pp. 1–13, 1985.

[113] D. Allen e W. Strong, "A model for the synthesis of natural sounding vowels", *J. Acoust. Soc. Am.*, vol. 78, pp. 58–69, Julho de 1985.

[114] M. Rothemberg, "An interactive model for voice source", *Speech Transmission Laboratory, Quarterly Progress ans Status Report*, vol. STL-QPSR 4, pp. 1–17, 1981.

[115] Ingo R. Titze, "A four-parameter model of the glottis and vocal fold contact area", *Speech Communication*, vol. 8, no. 3, pp. 191–201, Setembro de 1989.

[116] Kenneth N. Stevens, "Airflow and turbulence noise for fricative and stop consonants: Static considerations", *J. Acoust. Soc. Am.*, vol. 50, no. 4, Part 2, pp. 1180–1192, Maio de 1971.

[117] Victor N. Sorokin, "Determination of vocal tract shape for vowels", *Speech Communication*, vol. 11, no. 1, pp. 71–85, 1992.

[118] William J. Hardcastle e John Laver, Eds., *Handbook of Phonetic Sciences*, Blackwell, Oxford, 1996.

[119] Brad H. Story, Ingo R. Titze, e Eric A. Hoffman, "Vocal tract shapes and area functions from magnetic resonance imaging (MRI)", *J. Acoust. Soc. Am.*, vol. 98, no. 5, Pt. 2, pp. 2930, Novembro de 1995.

[120] S. Narayanan, A. Alwan, e K. Haky, "An articulatory study of fricative consonants using MRI", *J. Acoust. Soc. Am.*, vol. 98, no. 3, pp. 1325–1347, Setembro de 1995.

[121] S. Narayanan, A. Alwan, e K. Haky, "Towards articulatory-acoustic models for liquid consonants based on MRI and EPG data - part i: the laterals", *J. Acoust. Soc. Am.*, vol. 101, no. 2, pp. 1064–1077, Fevereiro de 1997.

[122] S. Narayanan, A. Alwan, e K. Haky, "Towards articulatory-acoustic models for liquid consonants based on MRI and EPG data - part ii: the rotics", *J. Acoust. Soc. Am.*, vol. 101, no. 2, pp. 1078–1089, Fevereiro de 1997.

[123] Didier Demolin, Véronique Lecuit, Thierry Metens, Bruno Nazarian, e Alain Soquet, "Magnetic resonance measurements of the velum port opening", em *Proc. ICSLP '98*, Sydney, Australia, 1998.

[124] Christine H. Shadle, Mohammad Mohammad, John N. Carter, e Philip J. B. Jackson, "Multi-planar dynamic magnetic resonance imaging: New tools for speech research", em *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1999, pp. 623–626.

[125] John H. Esling, Jocelyn Clayards, Jerold A. Edmondson, Qiu Fuyuan, e Jimmy G. Harris, "Quantification of pharyngeal articulations using measurements from laryngoscopic images", em *Proc. ICSLP '98*, Sydney, Australia, 1998.

[126] Joseph S. Perkell, Marc H. Cohen, Mario A. Svirsky, Melanie L. Natthies, Iñaki Garabieta, e Michell T. T. Jackson, "Electromagnetic midsagital articulometer systems for transducing speech articulatory movements", *J. Acoust. Soc. Am.*, vol. 92, no. 6, pp. 3078–3096, Dezembro de 1992.

[127] Philip Hoole e Noel Nguyen, "Electromagnetic articulography in coarticulation research", *Forschungsberichte des Instituts für Pho-*

netik und Sprachliche Kommunikation der Universität München (FIPKM), vol. 35, pp. 177–184, 1997.

[128] António Teixeira e Francisco Vaz, "European Portuguese nasal vowels: An EMMA study", em *Proc. EuroSpeech*, Alaborg, Dinamarca, Setembro de 2001, vol. 2, pp. 1843–1846.

[129] Alan A. Wrench, "An investigation of sagittal velar movement and its correlation with lip, tongue and jaw movement", em *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1999, pp. 435–438.

[130] Andreas Zierdt, Philip Hoole, e Hans G. Tillmann, "Development of a system for three-dimensional fleshpoint measurement of speech movements", em *Proc. ICPhS*, 1999, pp. 73–75.

[131] Satoshi Horiguchi e Fredericka Bell-Berti, "The Velotrace: A device for monitoring velar position", *The Cleft Palate Journal*, vol. 24, no. 2, pp. 104–111, Abril de 1987.

[132] G. C. Burnett, J. F. Holzrichter, T. J. Gable, e L. C. Ng, "Direct and indirect measures of speech articulator motions using low power EM sensors", em *Proc. ICPhS*, 1999, pp. 2247–2249.

[133] Francis Charpentier, "Determination of the vocal tract shape from the formants by analysis of the articulatory-to-acoustic nonlinearities", *Speech Communication*, vol. 3, no. 4, pp. 291–308, Dezembro de 1984.

[134] Man Mohan Sondhi, "Estimation of vocal-tract areas: The need for acoustical measurements", *IEEE Transactions on Acoustic, Speech, and Signal Processing*, vol. ASSP-27, no. 3, pp. 268–273, Junho de 1979.

[135] J. Schroeter, P. Meyer, e S. Parthasarathy, "Evaluation of improved articulatory codebooks and codebook access distance measures", em *Proc. ICASSP*, Albuquerque, USA, Abril de 1990, vol. 1, pp. 393–396.

[136] C. Silva, S. Chennoukh, e I. Trancoso, "On improving the decision algorithm for articulatory codebook search", em *Proc. Eurospeech*, Géza Gordos e Géza Németh, Eds., Budapest, Hungary, Setembro de 1999, vol. 1, pp. 153–156.

[137] Juergen Schroeter e Man Mohan Sondhi, "Techniques for estimating vocal-tract shapes from speech signal", *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 1, Part II, pp. 133–150, Janeiro de 1994.

[138] Qiuzhen Xue, Yu Hen Hu, e Paul Milenkovic, "Analyses of the hidden units of the multi-layer perceptron and its application in acoustic-to-articulatory mapping", em *Proc. ICASSP*, 1990, pp. 869–872.

[139] Mazin G. Rahim, Colin C. Goodyear, W. Bastiaan Kleijn, Juergen Schroeter, e M. Mohan Sondhi, "On the use of neural networks in articulatory speech synthesis", *J. Acoust. Soc. Am.*, vol. 93, no. 2, pp. 1109–1121, Fevereiro de 1993.

[140] Richard S. McGowan, "Recovering articulatory movement from formant frequency trajectories using task dynamics and a genetic algorithm: Preliminary model tests", *Speech Communication*, vol. 4, no. 1, Fevereiro de 1994.

[141] Sunil K. Gupta e Juergen Schroeter, "Pitch-synchronous frame-by-frame and segment-based articulatory analysis by synthesis", *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2517–2530, Novembro de 1993.

[142] S. Parthasarathy e Cecil H. Coker, "On automatic estimation of articulatory parameters in a text-to-speech system", *Computer Speech and Language*, vol. 6, pp. 37–75, 1992.

[143] Katsuhiko Shirai e Tetsunori Kobayashi, "Estimating articulatory motion from speech wave", *Speech Communication*, vol. 5, pp. 159–170, 1986.

[144] S. E. Levinson e C. E. Schmidt, "Adaptive computation of articulatory parameters", *J. Acoust. Soc. Am.*, vol. 74, no. 4, pp. 1145–1154, October de 1983.

[145] Hani Yehia e Fumitada Itakura, "Determination of human vocal-tract dynamic geometry from formant trajectories using spatial and temporal fourier analysis", em *Proc. ICASSP*, 1994, vol. I, pp. 477–480.

[146] Peter Meyer, Juergen Schroeter, e Man Mohan Sondhi, "Design and evaluation of optimal cepstral lifters for accessing articulatory codebooks", *IEEE Trans. Ac. Sp. Sig. Proc*, vol. 39, no. 7, Julho de 1991.

[147] Mats Båvegård e Gunnar Fant, "From formant frequencies to VT area fuction parameters", *Speech Transmission Laboratory, Quarterly Progress ans Status Report*, vol. STL-QPSR 4, pp. 55–66, 1995.

[148] Victor N. Sorokin, "Inverse problem for fricatives", *Speech Communication*, vol. 14, no. 3, pp. 249–262, Junho de 1994.

[149] Arturo Galván-Rodríguez, *Étude dans le cadre de l'inversion acoustico-articulatoire: Amélioration d'un modèle articulatoire, normalisation du locuteur et récuperation du lieu de constriction des plosives*, Thése, Institut National Polytechnique de Grenoble, 1997.

[150] Korin Richmond, "estimating velum height from acoustics during continuous speech", em *Proc. Eurospeech*, Géza Gordos e Géza Németh, Eds., Budapest, Hungary, Setembro de 1999, vol. 1, pp. 149–152.

[151] Solange Rossato, Gang Feng, e Rafaël Laboissière, "Recovering gestures from speech signals: A preliminary study for nasal vowels", em *Proc. ICSLP '98*, Sydney, Australia, 1998.

[152] Cecil H. Coker, "Systems and methods for performing phonemic synthesis", 27 Maio de 1997.

[153] C. H. Coker, N. Umeda, e C. P. Browman, "Automatic synthesis from ordinary English text", *IEEE Trans. Audio and Electr.*, vol. AU-21, no. 3, pp. 293–298, Junho de 1973.

[154] Celia Scully, "Linguistic units and units of speech production", *Speech Communication*, vol. 6, no. 2, pp. 77–142, Junho de 1987.

[155] Catherine P. Browman e Luis Goldstein, "Articulatory phonology: An overview", *Phonetica*, vol. 49, pp. 155–180, 1992.

[156] Catherine P. Browman e Louis Goldstein, "Articulatory gestures as phonological units", *Phonology*, vol. 6, pp. 201–251, 1989.

[157] Samir Chennoukh, Daniel Sinder, Gael Richard, e James Flanagan, "Voice mimic system using articulatory codebook for estimation of vocal tract shape", em *Proc. Eurospeech '97*, Rhodes, Greece, Set. de 1997, pp. 429–432.

[158] F. Zussa, Q. Lin, G. Richard, D. Sinder, e J. Flanagan, "Open-loop acoustic-to-articulatory mapping", *J. Acoust. Soc. Am.*, vol. 98, no. 5, Pt. 2, pp. 2931, Novembro de 1995.

[159] Frédéric Zussa, "A new design for articulatory parametrization of speech: Application to low-bit rate coding and recognition", Industrial thesis report, CAIP, Rutgers University, Agosto de 1995.

[160] Gérard Bailly, Rafaël Laboissière, e Arturo Galván, "Learning to speak: Speech production and sensori-motor representations", em *Self-Organization, Computational Maps and Motor Control*, Pietro Morasso e Vittorio Sanguineti, Eds., pp. 593–615. Elsevier, Amsterdam, 1997.