MIGUEL ARJONA RAMÍREZ

Predição Linear Harmônica para Processamento Espectral e Temporal de Sinais de Voz

Tese apresentada à Escola Politécnica da
Universidade de São Paulo
para obtenção do
título de Professor Livre-Docente
junto ao Departamento de
Engenharia de Sistemas Eletrônicos.

São Paulo
2006

MIGUEL ARJONA RAMÍREZ

Predição Linear Harmônica para Processamento Espectral e Temporal de Sinais de Voz

Tese apresentada à Escola Politécnica da
Universidade de São Paulo
para obtenção do
título de Professor Livre-Docente
junto ao Departamento de
Engenharia de Sistemas Eletrônicos.

Área:
Processamento de Sinais

São Paulo
2006

Arjona Ramírez, Miguel

Predição Linear Harmônica para
Processamento Espectral e Temporal de Sinais de Voz.
São Paulo 2006.
62p.

Tese (Livre-Docência) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

1. Processamento de Sinais 2. Codificação de Voz 3. Predição Linear 4. Modelagem Espectral 5. Modelos Autorregressivos I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II.t

À Mariko, à Karina, a Carmen, in memoriam, e a Salvador, in memoriam.

Agradecimentos

Aos colegas do LPS Vítor, Flávio, Denise e Mário pelo apoio e incentivo.

Aos orientados de pós-graduação Liselene e Sergio pela dedicação e idéias compartilhadas.

Aos orientados de iniciação científica, em especial ao Alexandre e ao Vinícius, pela curiosidade e persistência.

Ao Prof. Normonds, in memoriam, por instigar a importância nos princípios e na expressão simbólica.

Ao Prof. Max, in memoriam, por instigar a análise textual e a completude de idéias.

Sumário

Lista de Figuras

Lista de Tabelas

Resumo

Propõe-se um método de predição linear (LP) cuja ordem é decomposta em dois estágios. No primeiro estágio o modelo define um filtro que é usado na extração de ciclos de onda do sinal residual LP. Para a extração de ciclos, propõe-se um algoritmo capaz de preservar a reconstrução perfeita da forma e da periodicidade dos ciclos. Em seguida, no segundo estágio, modela-se o espectro harmônico resultante por um algoritmo LP discreto. Este método atinge distorção espectral logarítmica (log SD) inferior a 1 dB no segundo estágio da modelagem espectral LP discreta. É possível obter estatísticas melhores de ajuste da modelagem espectral quando a medida de distorção espectral para a LP discreta é simétrica e definida de acordo com um método proposto de separação do gradiente do erro, cujos componentes resultantes são certamente espectros de potência. Derivam-se duas distorções espectrais desse tipo a partir da log SD e suas aplicações na modelagem LP discreta são comparadas com outros métodos que empregam as medidas de distorção cosh e de Itakura-Saito.

Abstract

A linear prediction (LP) method is proposed whose order is split in two stages. In the first stage the model defines a filter used in extracting cycle waveforms from the LP residual signal. For cycle extraction, an algorithm is proposed that is capable of perfectly reconstructing waveshape and periodicity. Next, in the second stage, the resulting harmonic spectrum is modeled by a discrete LP algorithm. This method achieves log spectral distortion (SD) below 1 dB for second-stage discrete LP spectral modeling. Better model fit statistics are obtained when the spectral distortion measure for discrete LP is a symmetrical one defined according to a proposed separation method for the error gradient whose resulting components are guaranteed to be power spectra. Two such spectral distortion measures are derived from the log SD and their application in discrete LP modeling is compared to methods using the Itakura-Saito and the cosh distortion measures.

1 Introdução

1.1 Predição linear

O sinal de voz varia suas características ao longo do tempo. Por essa razão é capaz de transmitir informação. De um ponto de vista estatístico, o sinal de voz é não-estacionário. Porém, pode ser considerado quase-estacionário durante intervalos da ordem de 10 ms a 20 ms ou mais, dependendo da natureza do som considerado. Quando se faz uma análise individual desses intervalos do sinal, diz-se que ela é de curto prazo. A análise em questão pode ser uma análise de Fourier como pode ser uma análise preditiva, por exemplo.

A predição linear (LP) foi proposta no final da década de 1960 [1, 2, 3] para analisar sinais de voz, produzindo modelos para codificação e transmissão eficiente. A síntese a partir do modelo recebido, entretanto, depende de outros parâmetros e controles para poder efetuar a reconstrução do sinal.

De forma mais específica, a modelagem LP produz modelos espectrais mais suaves que o espectro de curto prazo do segmento de sinal de voz considerado. Como pode ser observado na Fig. 1, o modelo espectral LP parece uma envoltória que acompanha as mais extensas curvaturas espectrais sem se perturbar com os contornos mais localizados.

O modelo espectral LP pode ser obtido como a resposta em freqüência de um filtro $H(z)$ definido pelos parâmetros do modelo. Num raciocínio construtivo, o que falta a esta resposta é sua multiplicação por um espectro que exiba os contornos locais que o modelo LP não foi capaz de absorver. Este espectro faltante pode ser identificado com o espectro de um sinal de excitação a ser aplicado na entrada do filtro de síntese $H(z)$ de acordo com o modelo de síntese representado na Fig. 2, onde se imagina que esse sinal seja produzido pelo gerador. Um gerador capaz de produzir poucos tipos de formas de onda tipifica um sintetizador paramétrico ou vocoder ("voice coder").

Por outro lado, se houver pretensão de se obter a reconstrução perfeita do sinal de

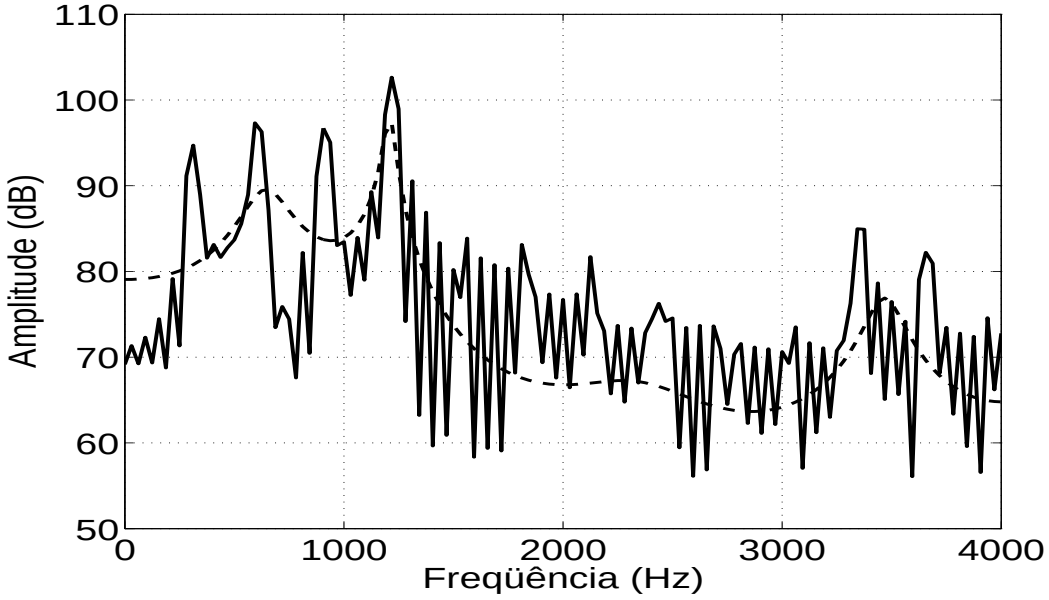


Figura 1: Densidade espectral de potência de um segmento de voz sonoro (curva contínua) e seu correspondente modelo espectral LP de ordem 10 (curva tracejada).

voz, deve-se inverter o modelo de síntese, usando o filtro inverso

$$A(z) = \frac{1}{H(z)} \tag{1.1}$$

no modelo de análise. No filtro inverso injeta-se o sinal de voz original, produzindo o sinal residual da análise LP, cuja forma de onda aparece na Fig. 3 abaixo da forma de onda do segmento de voz sonoro em consideração acima. Nota-se que este sinal apresenta os picos mais afastados de sua excursão média que o sinal de voz. Esta característica facilita a detecção da periodicidade do sinal bem como a extração de seus ciclos, conforme será elaborado no Cap. 2.

Retornando ao espectro do segmento de voz da Fig. 1, que corresponde a uma fatia vertical extraída do espectrograma inferior da Fig. 4 em 750 ms, notam-se picos locais nas freqüências em torno de 300 Hz, 600 Hz, 900 Hz, 1200 Hz, 1800 Hz, 2100 Hz, 2400 Hz, 3300 Hz e 3600 Hz dentro da banda-base de 4 kHz. Estima-se que eles correspondam à freqüência fundamental $f_0 = 300$ Hz e a algumas de suas harmônicas. Agora, olhando mais atentamente o espectro, até seria possível arriscar um palpite de que há harmônicas também em 1500 Hz e em 3000 Hz. Mas, de qualquer forma, todas as harmônicas além da 4ª apresentam amplitude bem inferior, corroborando o modelo senoidal "sinusoidal transform coding" (STC) [4] de que há uma freqüência de corte para os espectros sonoros além da qual eles deixam de ser harmônicos.

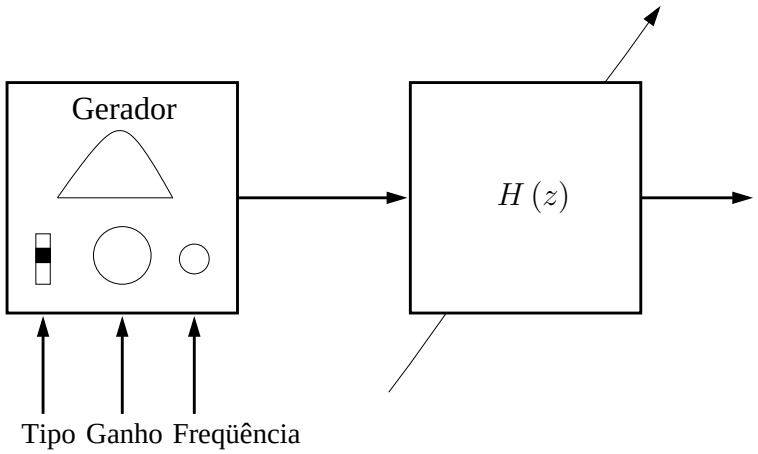


Figura 2: Modelo fonte-filtro de produção da fala em que os coeficientes do filtro podem ser controlados.

Ainda na Fig. 1, observa-se que a envoltória espectral LP apresenta picos locais nas freqüências em torno de 640 Hz, 1200 Hz, 2300 Hz e 3500 Hz. Ela assemelha-se a uma fatia vertical que fosse extraída do espectrograma superior da Fig. 4 em 750 ms, retendo suas regiões mais escuras. Estima-se que elas representem as formantes f_1, f_2, f_3 e f_4, respectivamente, do som representado pelo segmento de sinal analisado. Nota-se ainda que as formantes, em geral, não coincidem com harmônicas. Entretanto, quanto mais forte a harmônica mais próxima, mais "se aproxima" a formante da harmônica. Obviamente, este comportamento reflete em boa medida a natureza do aparelho fonador humano e um pouco os limites da própria modelagem LP.

1.2 Representações tempo-freqüenciais do sinal de voz

Na codificação de voz, à medida que se reduz a taxa de transmissão, principalmente abaixo de 2 kbit/s, torna-se praticamente impossível a transmissão regular de informação sob pena de reduzir excessivamente a definição dos sons individuais [5]. Por um lado, há sons que são longos e que podem ser identificados por seu espectro, que se mantém praticamente constante durante muito tempo. Por outro lado, há sons transitórios que podem ser identificados mais facilmente por seu instante de ocorrência, que deve ser bem definido, exigindo alta resolução temporal. De forma geral, podem-se associar "eventos" da fala com distribuições da potência do sinal em várias regiões do plano tempo-freqüência [6].

A visualização de eventos no domínio tempo-freqüência pode ser feita através de espec-

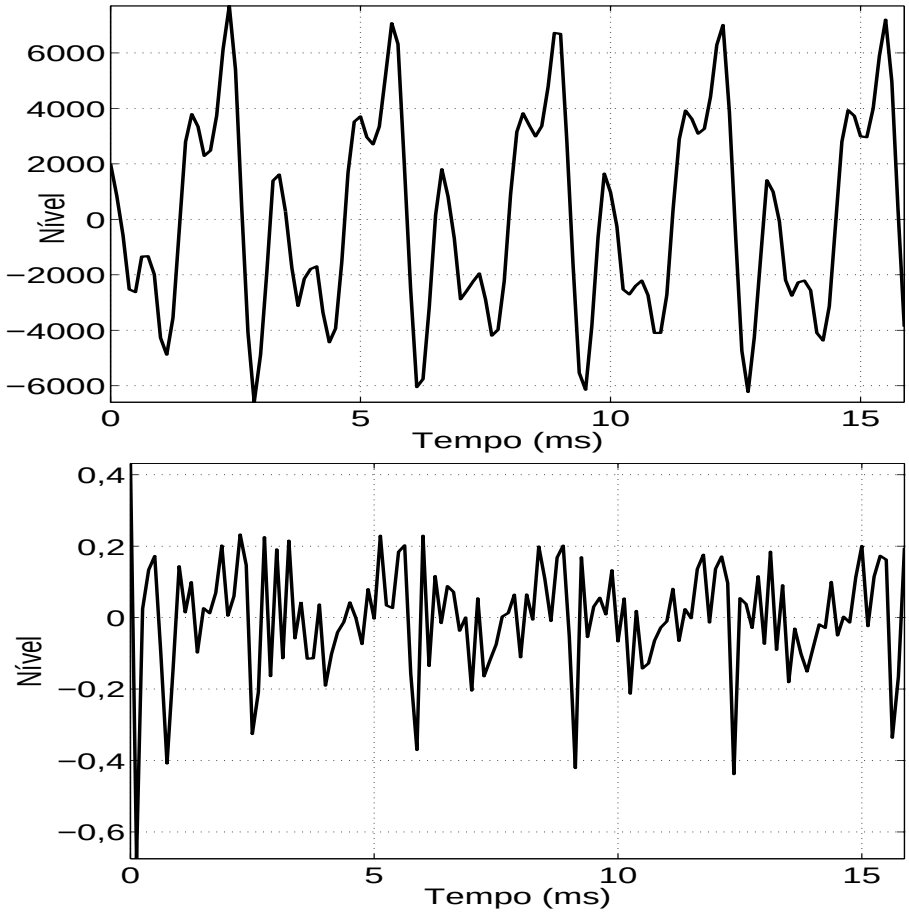


Figura 3: Formas de onda de um segmento de voz sonoro (superior) e correspondente sinal residual LP de ordem 10 (inferior).

trogramas, que constituem um poderoso auxílio instrumental na pesquisa da fala desde a invenção do espectrógrafo de voz na década de 1940 [7]. No espectrograma, a densidade de potência do sinal é indicada pelo tom de cinza, sendo maior quanto mais escuro for o tom.

Os dois modos comuns de qualificação de espectrogramas são de faixa estreita e de faixa larga e ambos estão exemplificados na Fig. 4. O espectrograma de faixa larga apresenta estrias verticais ao longo do eixo do tempo, que sinalizam as ocorrências dos ciclos. Notam-se também trajetórias mais escuras e grossas, que representam a evolução de cada formante. Por outro lado, no espectrograma de faixa estreita são nítidas as trajetórias das harmônicas.

Os aspectos de localização dos modelos em tempo-freqüência aparecem com maior ou menor intensidade em todos os casos de processamento do sinal de voz e na sua codificação, em particular, em todos os métodos usados. Em geral, como exposto na

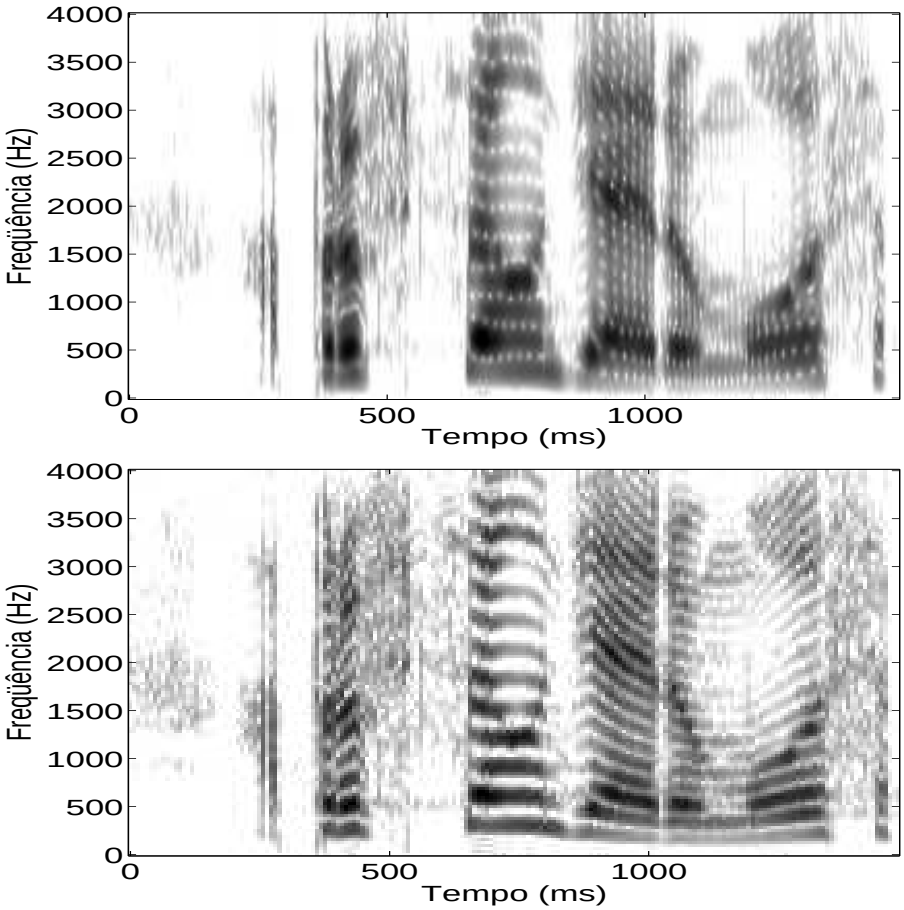


Figura 4: Espectrograma de faixa larga (superior) e espectrograma de faixa estreita (inferior) da frase truncada "A profissão de aeromoça...", pronunciada por uma locutora feminina.

Seção 1.1, segmenta-se o sinal de voz no tempo sem muita precisão e geralmente em segmentos de mesmo comprimento, cuja análise posterior revela suas características espectrais com maior ou menor resolução. No caso dos codificadores de voz senoidais [4], há uma ênfase em melhor resolução espectral, podendo-se livremente associá-los à representação do espectrograma de faixa estreita. Parcialmente ao contrário, os codificadores por interpolação da forma de onda (WI) [8] exigem uma alta resolução temporal que se manifesta na extração dos ciclos de onda como será visto no Cap. 2, o que permite, de certa forma, imaginá-los como o espectrograma de faixa larga. Entretanto, embora em muitos casos de implementação de codificadores WI a resolução espectral seja comparativamente inferior à temporal, a concepção deste modelo de análise e síntese não é incompatível com alta resolução espectral também. Espera-se que a modelagem proposta neste trabalho possa contribuir nessa direção.

1.3 Resultados

As principais contribuições deste trabalho são emprestadas do artigo a ser publicado [9] e dos artigos apresentados em congressos [10, 11, 12]. Além disso, ele contribui com a interconexão dos temas, dando-lhes unidade e contexto. Especificamente, podem ser identificados os seguintes resultados inovadores neste trabalho.

- Na Seção 4.4 apresenta-se um procedimento para obtenção de medidas de distorção espectrais para métodos de estimação espectral iterativos baseados em modelos LP. Ele baseia-se na separação do gradiente do erro de ajuste espectral em componentes espectrais com propriedades de espectros de potência, que podem ser transformadas em seqüências de autocorrelação de forma a constituir algoritmos LP.

- Na Seção 4.4 aplica-se o procedimento proposto acima à distorção espectral logarítmica (log SD), obtendo-se duas medidas de distorção simétricas que a aproximam. As medidas espectrais simétricas são mais apropriadas para a avaliação da modelagem de espectros harmônicos.

- Propõe-se a aplicação do algoritmo LP baseado nas medidas de distorção propostas, bem como outros algoritmos LP para espectros discretos (DAP), ao segundo estágio de análise LP com decomposição da ordem de predição (SOLP), sendo o primeiro estágio de ordem inferior usado para processar o sinal de voz para extração de ciclos, conforme descrito na Seção 4.2. Este método permite atingir log SD em torno de 1 dB na modelagem do espectro harmônico do segundo estágio em comparação com a distorção da ordem de 3 dB a 4 dB que se obtém ao modelar diretamente o espectro do sinal de voz.

- O método de extração de ciclos de onda apresentado nas Seções 2.3 e 2.4 permite captar e preservar as propriedades de periodicidade e de forma do sinal residual da LP, entregando-as como trajetórias com taxa de amostragem uniforme ao estágio de codificação que poderá processá-las de acordo com seus próprios critérios. Posteriormente, as trajetórias reconstruídas pelo estágio de decodificação são reamostradas de acordo com os processos previstos na Seção 2.5 para, finalmente, sintetizar, de acordo com o processo descrito na Seção 2.6, o sinal reconstruído a partir dos ciclos recuperados. Esta metodologia permite a reconstrução perfeita do sinal residual da LP, dependendo dos métodos de interpolação usados.

- A modelagem LP de espectros contínuos é apresentada de forma didática na Seção 3.2,

integrando as derivações dos algoritmos LP para a distorção de Itakura-Saito e para a média da razão de densidades espectrais de potência.

- A análise e a síntese WI são apresentadas em notação de tempo discreto no Cap. 2 devido à constatação da suficiência da taxa de amostragem do sinal como taxa mais alta de representação capaz de produzir a reconstrução perfeita do sinal. Sem essa restrição, na literatura predomina a notação em tempo contínuo.

1.4 Organização

No Cap. 2 é apresentado o método proposto para extração de ciclos de onda do sinal residual da LP juntamente com resultados de simulações para alguns métodos de interpolação das trajetórias e das formas de onda características.

No Cap. 3 são descritos primeiramente os métodos de modelagem LP de espectros contínuos, cujos algoritmos LP e medidas de erro podem ser comparados com os resultados da modelagem LP no domínio do tempo, apresentados nos Apêndices A e B. Além disso, na derivação dos algoritmos LP para espectros contínuos, é necessário utilizar uma propriedade dos modelos espectrais LP apresentada no Apêndice C. Finalmente, na Seção 3.3 é apresentado o algoritmo iterativo DAP para modelagem LP de espectros discretos com apoio em resultados básicos apresentados no Apêndice D.

No Cap. 4 é descrita a proposta de decomposição da ordem LP (SOLP) em dois estágios na Seção 4.2, empregando-se, no primeiro estágio, os algoritmos propostos nas Seções 2.3 e 2.4. Em seguida, na Seção 4.4, propõe-se para o segundo estágio um procedimento para obtenção de medidas de distorção espectrais para métodos de estimação espectral LP iterativos, cuja derivação está apoiada nas apresentaçôes da Seção 3.3 e da Seção 4.3 e nos resultados fundamentais do Apêndice D. Os resultados de simulações do método SOLP proposto em comparação com outros algoritmos de modelagem LP de espectros harmônicos são apresentados e comentados no Cap. 5.

Finalmente, no Cap. 6 conclui-se com a apresentação das implicações dos resultados obtidos.

2 *Extração de ciclos para interpolação da forma de onda*

2.1 Introdução

Os ciclos de onda correspondem a períodos de segmentos sonoros de sinais de voz quando considerados num sentido restrito. Apesar de sua caracterização clássica em análise de voz, sendo mais importantes em codificadores de voz pitch-síncronos, a operação de extração de ciclos de onda (CyWs) tem sido aplicada com sucesso mesmo sobre segmentos surdos do sinal de voz. Quando usada na codificação de voz, isso tem ocorrido conjuntamente com vária técnicas de interpolação. Esta necessidade deve-se à duração variável dos ciclos de onda e à importância que adquire o processamento de sua informação de forma independentemente de sua duração e amplitude.

Efetivamente, a interpolação da forma de onda (WI) representa um modelo flexível de sinal de excitação para a codificação de voz, em geral associada à predição linear para a definição de seu modelo espectral [8]. Entretanto, suas várias formas de onda de parâmetros e de sinais têm difentes larguras de faixa inerentes e taxas de amostragem críticas que geralmente não são uniformes. Dentre estas taxas podem-se enumerar a taxa de ciclos, a taxa da predição linear (LP), a taxa da detecção de pitch, a taxa de amostragem do sinal e a taxa de codificação da forma de onda. Por esse motivo, torna-se mais simples e atraente a descrição do modelo em tempo contínuo como foi feito originalmente. De fato, há representações em tempo contínuo que podem ser descritas digitalmente com as B-splines cúbicas [13], mas a administração efetiva de diferentes taxas para sinais e parâmetros é um obstáculo persistente na implementação dos codificadores. Resolve-se comumente com a imposição da taxa de amostragem do sinal e da taxa de determinação de parâmetros em compatibilidade com a taxa de transmissão do codificador, acarretando imperfeições na modelagem ou ineficiência na codificação.

O ponto de vista tomado aqui considera que os sinais e os parâmetros devem ser

extraídos ou determinados na razão de suas taxas naturais e a conversão para as taxas de codificação é deixada a cargo de interpoladores evolutivos. Conceitualmente, a taxa de amostragem mais alta a ser considerada é a taxa de amostragem do sinal, tornando as representações em tempo discreto suficientes para o processamento e também para a descrição algorítmica [10].

Assim, a parte que cuida da representação da fonte no codificador pode ser controlada pelas características da fonte enquanto a parte responsável pela codificação pode se adaptar aos requisitos de transmissão, que podem demandar que o codificador opere numa taxa fixa ou em taxa variável. Com este procedimento obtém-se alta flexibilidade na escalabilidade da taxa.

O extrator de ciclos descrito também é adequado para codificadores pitch-síncronos [14] contanto que se redefina o pitch como a duração do segmento do sinal que se estende entre dois instantes de baixa amplitude que ocorram entre picos do sinal independentemente da sonoridade.

2.2 Descrição da interpolação de forma de onda

Na interpolação de forma de onda [8], a superfície $u\left(t,\phi(t)\right)$ caracteriza a excitação conjuntamente com a trajetória de fase

$$\phi(t) = \phi\left(t_0\right) + 2\pi \int_{t_0}^{t} \frac{1}{p(t)} dt, \tag{2.1}$$

em que $p(t)$ é a trajetória de pitch. A *forma de onda característica* (CW) $c_{t_0}\left(\phi\right) = u\left(t,\phi\right)|_{t=t_0}$ para $\phi \in [-\pi,\pi)$ descreve o ciclo de onda potencial no instante $t = t_0$, que apenas é indicado pela amostra

$$c_{t_0}\left(\phi\right) = u\left(t_0,\phi\left(t_0\right)\right) = r\left(t_0\right)$$

do sinal residual. Portanto, neste ponto, será exigido para garantia da reconstrução perfeita que a forma de onda característica seja uma versão alongada do segmento do sinal residual que se estende desde t_0 em diante até o próximo instante $t_0 + p\left(t_0\right)$, intermediário no ciclo e situado entre picos assumindo-se que o próprio t_0 seja um instante intermediário no ciclo e situado entre picos.

É importante na amostragem e interpolação das CWs visualizar a superfície de CWs ao longo do eixo do tempo. Para uma fase $\phi = \phi_0$ normalizada, a correspondente *forma*

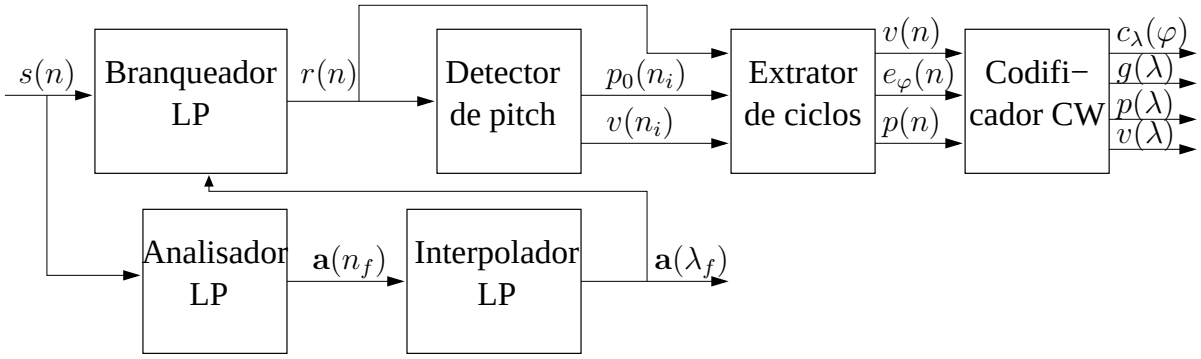


Figura 5: Diagrama de blocos de um codificador de voz genérico usando o extrator de ciclos.

de onda evolutiva (EW) é $e_{\phi_0}(t) = u(\phi, t)|_{\phi=\phi_0}$. Uma evolução mais suave da função característica pode ser obtida pela interpolação das formas de onda extraídas após sua normalização de comprimento.

O procedimento padrão de extração da forma de onda aplica amostragem uniforme [15] mas a extração crítica de ciclos de pitch tem sido usada para reduzir a complexidade da codificação [16] bem como para permitir a reconstrução perfeita da forma de onda [17].

2.3 Extração de ciclos

A forma de onda selecionada para processamento é o sinal residual $r(n)$ da predição linear, que é comumente escolhido devido a suas características periódicas acentuadas em relação ao sinal original. Na seqüência, a periodicidade do sinal residual é analisada em maior detalhe por um detector robusto de pitch baseado na seqüência de autocorrelação. Este detector segue de perto as diretrizes propostas em [18] e [19]. Ele fornece um valor estimado mesmo quando o sinal é surdo no intervalo considerado, sendo necessário um detector de sonoridade independente. Como pode ser visto na Fig. 5, usa-se um detector de sonoridade por autocorrelação, que entrega uma decisão $v(n_i)$ por intervalo também.

As estimativas de período fundamental facilitam a tarefa do demarcador de forma de onda, cuja tarefa é buscar as extremidades dos ciclos fundamentais como indicado na Fig. 7. Tendo em vista a reconstrução perfeita, a extremidade inicial do ciclo n_c é a amostra que ocorre no instante $n = d(n_c - 1) + 1$ que se segue ao término do ciclo anterior extraído enquanto sua extremidade final $n = d(n_c)$ é colocada numa posição de baixa amplitude entre os dois picos de pitch seguintes. O demarcador de ciclos na Fig. 7 busca os dois picos seguintes dentro de uma faixa de tolerância de amplitude em torno do pico atual durante um intervalo de tempo determinado pelo período fundamental atual

dentro de uma tolerância pré-determinada. Em seguida, o detector de ciclos buscará uma amostra de baixa amplitude em torno do ponto médio entre os dois picos seguintes. Na Fig. 6 podem ser observadas algumas demarcações de ciclos de onda.

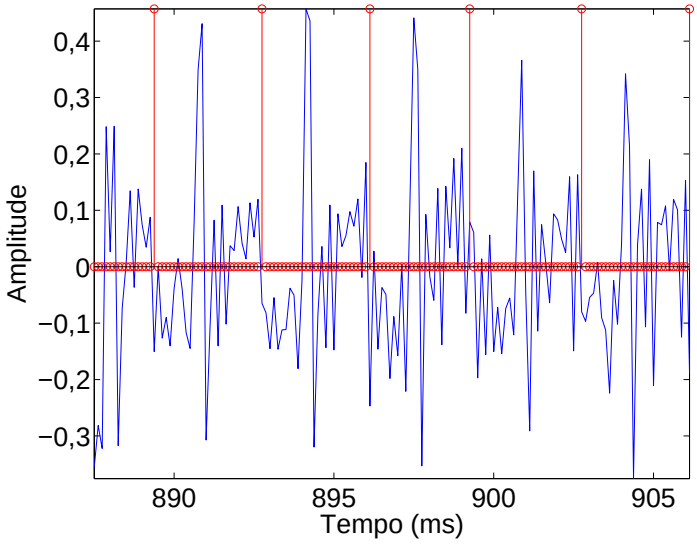
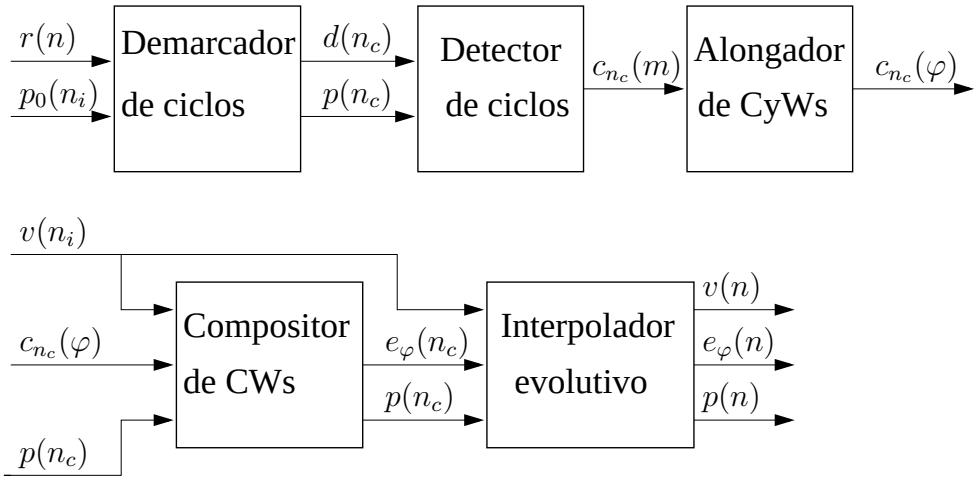


Figura 6: Ciclos de onda demarcados dentro de uma seção de sinal residual.

Para facilitar a escalabilidade dos resultados da análise, o período fundamental $p_0(n_i)$ determinado pelo detector de pitch para o intervalo n_i em que o ciclo correspondente se situa é substituído pela duração $p(n_c) = d(n_c) - d(n_c - 1)$ do ciclo.

Figura 7: Diagrama de blocos do extrator de ciclos.



2.4 Normalização da duração dos ciclos

Os ciclos de onda extraídos sofrem uma expansão de taxa de amostragem para vir a ocupar um período constante pré-determinado. Em consideração à natureza periódica dos períodos de pitch, a representação original dos ciclos de onda é sua série de Fourier, usada para promover seu alongamento para um período constante. Empregam-se freqüentemente os coeficientes da série de Fourier evolutiva

$$a_t(k) = \frac{1}{p(t)} \int_t^{t+p(t)} r(t) e^{-j\frac{2\pi k}{p(t)}t} dt \tag{2.2}$$

para $k = -K, -K+1, \ldots, K$ com $K = \lfloor f_{\mathrm{Ny}} p(t) \rfloor$, sendo que f_{Ny} é a largura de faixa do sinal ou sua freqüência de Nyquist. Estes coeficientes de Fourier podem ser usados diretamente na análise. Entretanto, uma nova escala de tempo normalizada ϕ torna-se mais eficiente para a codificação porque ela dispensa os processos de nascimento e morte de trajetórias harmônicas como ocorre nos codificadores senoidais [4]. Este eixo é normalmente denominado de eixo da fase e a CW disposta ao longo deste eixo descreve-se como

$$c_t(\phi) = \sum_{k=-\frac{P}{2}}^{\frac{P}{2}} a_t(k) e^{j\frac{2\pi k}{P}\phi} \tag{2.3}$$

em que P/f_s é o período de pitch constante para a freqüência de amostragem f_s. Este alongamento temporal mantém a capacidade de reconstrução perfeita contanto que o período de pitch constante não seja menor que o maior período de pitch do sinal. Além disso, na Eq. (2.3), a série de Fourier é estendida com a incorporação dos termos com coeficientes $a_t(k) = 0$ para $k = \pm(K+1), \pm(K+2), \ldots, \pm\frac{P}{2}$. Inversamente, a série de Fourier original pode ser obtida de volta através do truncamento da série alongada.

Neste ponto, é mais conveniente lidar com uma representação em tempo discreto. No lugar do sinal $r(t)$ em evolução na Eq. (2.2), o ciclo de onda extraído

$$c_{n_c}(m) = r\left(d(n_c - 1) + m + 1\right) \tag{2.4}$$

é usado para $m = 0, 1, \ldots, p(n_c) - 1$. Assim, a série de Fourier discreta do ciclo de onda que começa em n_c é

$$a_{n_c}(k) = \frac{1}{p\left(n_c\right)} \sum_{m=0}^{p(n_c)-1} c_{n_c}(m) e^{-j\frac{2\pi k}{P}m} \tag{2.5}$$

para $k = -K_l, -K_l+1, \ldots, K_u$. Para $p\left(n_c\right)$ par, $K_l = p\left(n_c\right)/2$ e $K_u = K_l - 1$. Caso contrário, para $p\left(n_c\right)$ ímpar, $K_l = K_u = \frac{p(n_c)-1}{2}$. A CW é obtida a partir da série de

Fourier estendida

$$c_{n_c}(\varphi) = \sum_{\varphi=-\left(\frac{P}{2}-1\right)}^{\frac{P}{2}-1} a'_{n_c}(k) e^{j\frac{2\pi k}{P}\varphi}, \tag{2.6}$$

em que se assume que o período de pitch constante P é par, sem perda de generalidade, e os coeficientes estendidos são

$$a'_{n_c}(k) = 0 \text{ for } K_l + 1 \leq |k| \leq \frac{P}{2} - 1 \text{ e } k = -\frac{P}{2}$$

em conjunto com

$$a'_{n_c}(-K_l) = a'_{n_c}(K_l) = \frac{1}{2} a_{n_c}(-K_l)$$

para p_{n_c} par ou

$$a'_{n_c}(-K_l) = a_{n_c}(-K_l) \text{ e } a'_{n_c}(K_u) = a_{n_c}(K_u)$$

para p_{n_c} ímpar. Inversamente, a série de Fourier original pode ser obtida por truncamento e pela inversão da operação de corte pelas extremidades como delineado acima para os coeficientes estendidos.

Tendo em vista a codificação eficiente, pode-se empregar alternativamente interpolação limitada em faixa de freqüência através de funções sinc truncadas para normalizar os comprimentos dos ciclos de onda extraídos [11].

2.5 Recomposição das formas de onda características

A seqüência $\left\{\{c_{n_c}(\varphi)\}_{\varphi=0}^{P-1}\right\}_{n_c}$ de formas de onda características subentende uma superfície característica que emerge quando as formas de onda estão alinhadas e dispostas ordenadamente ao longo do eixo do tempo n, contanto que a taxa de amostragem de ciclos tenha sido suficientemente alta. O processo de extração de ciclos delineado na Seção 2.3 garante um alto grau de alinhamento entre ciclos consecutivamente extraídos devido à colocação do pico na região média de $c_{n_c}(m)$. Entretanto, ainda persiste um desalinhamento residual, causado pelas variações no período de pitch, que o compositor de CWs incluído na Fig. 7 corrige através de deslocamento circular dos ciclos quando a forma de onda for sonora. Em conseqüência, a trajetória de pitch tem que ser ajustada para cancelar o desalinhamento eventual de tal forma que a sincronia possa ser restabelecida. O alinhamento pelos picos revelou-se mais eficaz que o alinhamento baseado na autocorrelação máxima em acordo com [18].

Em relação à disposição das CWs ao longo do eixo do tempo, a melhor estratégia em

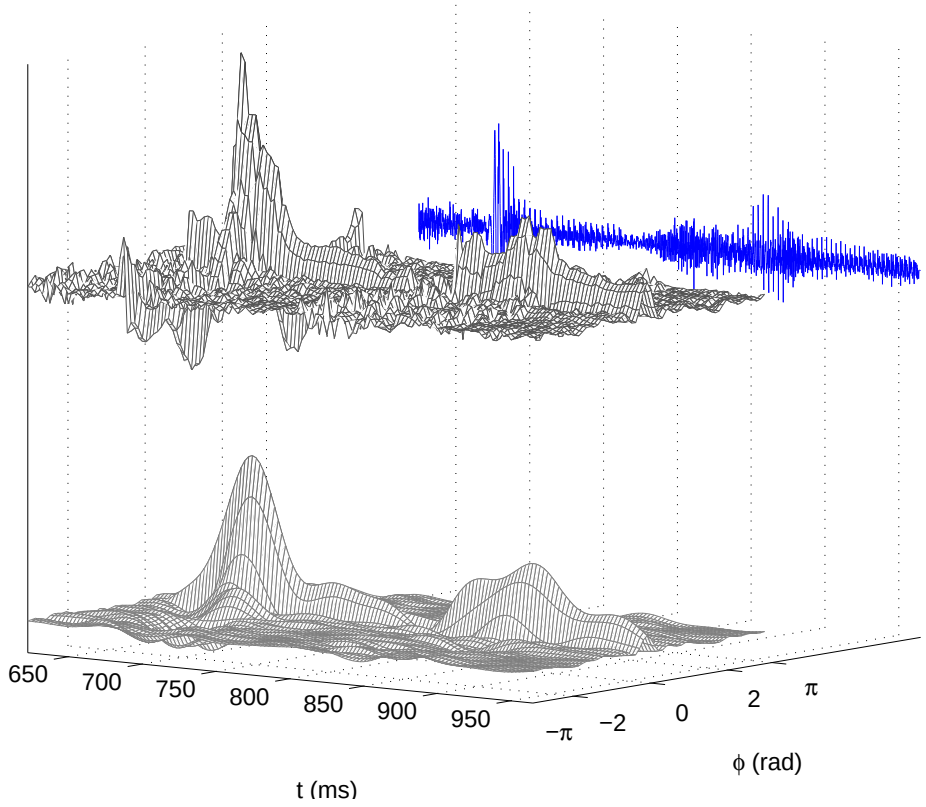


Figura 8: Frontalmente vê-se a superfície CW filtrada de uma seção do sinal residual, que está na região posterior. No plano inferior vê-se a superfície evolutiva alisada.

primeira aproximação é a retenção da mesma CW durante todo o período do ciclo de onda como

$$c_n(\varphi) = c_{n_c}(\varphi) \tag{2.7}$$

para $\varphi = 0, 1, \ldots, P-1$ e $n = d(n_c - 1) + 1, d(n_c - 1) + 2, \ldots, d(n_c)$.

Além disso, podem-se usar várias formas de interpolação para reamostrar uniformemente as CWs em taxa inferior. Aplicando-se interpolação sinc limitada em faixa de freqüência, como foi feito para o alongamento dos ciclos em [11], pode-se obter uma superfície evolutiva mais suave para reamostrar uniformemente em taxa mais baixa. Empregando-se $2D + 1$ amostras originais para efetuar a interpolação, o sinal reamostrado gerado é

$$e_\varphi(\lambda) = Q \sum_{n=\frac{\lambda}{Q}-D}^{\frac{\lambda}{Q}+D} e_\varphi(n)\, h(\lambda - Qn), \tag{2.8}$$

para $\varphi = 0, 1, \ldots, P-1$, sendo $Q = f_s/f_{\mathrm{CW}}$ o fator de subamostragem das CWs desde a taxa de amostragem f_s do sinal até a taxa final f_{CW} de amostragem das CWs. A superfície CW que se obtém com $f_{\mathrm{CW}} = 400$ Hz está ilustrada na Fig. 8.

Para interpolação sinc limitada em faixa de freqüência, as formas de onda evolutivas são superamostradas para a síntese como

$$\tilde{e}_{\varphi}(n) = \sum_{\lambda=Q(n-D)}^{Q(n+D)} e_{\varphi}(\lambda) h(Qn - \lambda). \tag{2.9}$$

2.6 Síntese de ciclos de onda

No sintetizador, as formas de onda evolutivas são superamostradas, como foi exemplificado ao final da seção anterior, para obter as formas de onda características reconstruídas.

Além disso, a trajetória de pitch $\tilde{p}(\lambda)$ decodificada é superamostrada na taxa de amostragem do sinal, produzindo a trajetória interpolada $\tilde{p}(n)$. A trajetória de sonoridade $\tilde{v}(\lambda)$ é superamostrada também e ambas são usadas para contrair as formas de onda características de volta ao ciclos de onda com períodos dados pela trajetória de pitch como esquematizado na Fig. 9.

Assim, a trajetória de pitch conduz o amostrador de ciclos de onda ao longo de uma trajetória de fase derivada como

$$\tilde{m}(n) = \left(\tilde{m}\,(0) + \sum_{i=1}^{n} 1 \right) \mod \tilde{p}(n) \tag{2.10}$$

para regenerar o sinal residual na forma

$$\tilde{r}(n) = \tilde{g}(n) \tilde{c}_n\,(\tilde{m}(n)) \tag{2.11}$$

após a multiplicação de cada amostra do ciclo pelo fator de ganho interpolado $\tilde{g}(n)$.

Para superamostrar os ciclos extraídos, outros tipos de interpolação podem ser usados além da extensão harmônica de séries de Fourier e da interpolação com sincs janeladas [11], como a interpolação com B-splines cúbicas [13].

2.7 Experimentos

O extrator de ciclos foi testado na taxa natural de ciclos, que foi superamostrada para a taxa de amostragem do sinal $f_s = 8$ kHz e tomada com uma taxa intermediária de amostragem de formas de onda características bem como também foi usado para emular a amostragem uniforme de CWs na taxa $f_{\mathrm{CW}} = 400\,\mathrm{Hz}$ estabelecida por [15]. Os ciclos

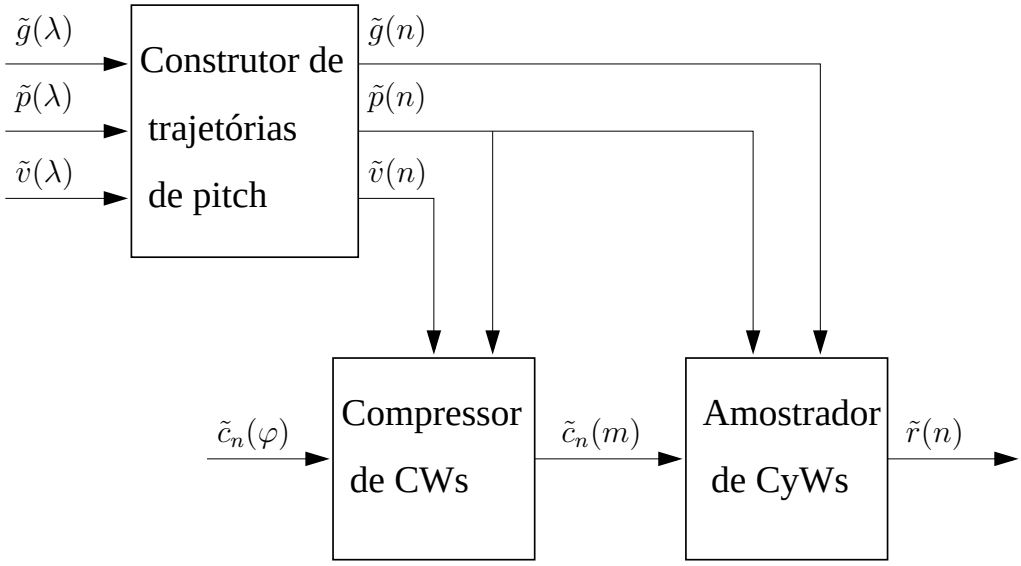


Figura 9: Diagrama de blocos do compositor de ciclos de onda para a síntese do sinal.

de onda são representados no domínio da fase normalizadas, onde eles são estendidos harmônicamente na representação em séries de Fourier e as correspondentes CWs são contraídas de volta aos ciclos por truncamento de suas séries de Fourier. Ainda foi aplicada a interpolação por sincs limitadas em faixa de freqüência para comparação de resultados. Em todos os casos, foi usada a trajetória de pitch extraída com precisão máxima. Os sinais de teste foram oito orações da base de dados TIMIT, tomadas com igual número de locutores femininos e masculinos e totalizando 14,5 s de fala feminina e 12,4 s de fala masculina.

O desempenho de reconstrução do sinal foi avaliado ao nível do sinal residual através da medida da relação sinal-ruído segmentada (SNRSEG) com segmentos de 16 ms entre o sinal residual $r(n)$ na Fig. 5 e sua reconstrução $\tilde{r}(n)$ pela Eq. (2.11), localizada na Fig. 9.

Em primeiro lugar, constatou-se que a extração de ciclos de onda na sua taxa natural com alongamento por série de Fourier até atingir o comprimento $P = 256$ ao longo do eixo da fase alcança praticamente a reconstrução perfeita. Ainda, a superamostragem das formas de onda evolutivas desde a taxa natural de ciclos até a taxa de amostragem do sinal usando interpolação por retenção de ordem zero mantém a reconstrução perfeita. Entretanto, quando se usa interpolação sinc baseada em $2D + 1 = 11$ amostras da forma de onda extraída naturalmente, a SNRSEG cai para valores em torno de 50 dB.

Quando as formas de onda evolutivas são filtradas passa-baixas e reamostradas na taxa $f_{\mathrm{CW}} = 400$ Hz através de interpolação sinc baseada em $2D + 1 = 11$ amostras das formas de onda evolutivas tomadas na taxa de amostragem do sinal como ilustrado na

Fig. 8, a SNRSEG média é de aproximadamente 30 dB, conferindo com o desempenho da extração convencional de ciclos [13].

Além disso, verificou-se em experimentos preliminares com extração múltipla de ciclos um aumento de mais de 7 dB sobre a extração WI convencional. A representação de múltiplos ciclos já foi usada anteriormente no caso da representação por uma transformada pitch-síncrona [14].

2.8 Conclusão

A descrição algorítmica da extração e interpolação de ciclos de onda foi posta no domínio do tempo discreto. Esta representação apresenta passos que abrangem o estágio de análise de um codificador de voz. Esta análise inclui um extrator de ciclos de onda que opera na taxa de ciclos natural não-uniforme do sinal residual da predição linear de acordo com a proposta de acompanhamento da evolução do sinal pelo estágio de análise para possibilitar o alcance da reconstrução perfeita. Por outro lado, o estágio de codificação pode operar nas suas taxas intrínsecas, sendo conectado com o estágio de análise através de um interpolador da forma de onda evolutiva. Testaram-se vários tipos de interpoladores, que propiciaram resultados situados entre a reconstrução perfeita e a extração e representação WI tradicional.

3 Modelos preditivos espectrais

3.1 Introdução

Conforme a apresentação da análise LP na Seção 1.1, ela pode ser usada para projetar um filtro capaz de reproduzir a envoltória espectral do segmento de sinal de voz e até para estimar características espectrais da produção da voz como as formantes. Isto mostra a capacidade que a modelagem LP apresenta de separar as características do filtro das características da fonte no modelo de síntese da Fig. 2.

Neste capítulo, em primeiro lugar, será explorada essa capacidade de ajuste dos modelos LP a espectros contínuos. Porém, em seguida, será abordada a possibilidade de modelagem LP de espectros discretos. Neste aspecto, já foi tocada uma limitação da modelagem LP na extração das formantes durante a análise do segmento sonoro exemplificado na Seção 1.1. Reproduzindo aqui a observação, quanto mais próxima de uma harmônica forte estiver uma formante, mais sua estimativa se deslocará em direção a essa harmônica.

A partir dessa observação, torna-se mais clara a compreensão de que, se forem ajustados modelos LP de ordem bem superior a 10, eventualmente será possível reproduzir todos os contornos do espectro harmônico [20]. No limite, a ordem do modelo LP terá que ser pelo menos igual ao dobro do número de harmônicas, que, no exemplo da Seção 1.1, acarretaria ordem 24 pelo menos.

O interessante seria ter a capacidade de ajuste de modelos LP de ordens menores. Uma abordagem para possibilitar este ajuste consiste na interpolação do espectro entre harmônicas para eliminar os vales profundos que soem ocorrer nessas regiões espectrais, com a redução da gama dinâmica do espectro [4]. Outra abordagem alternativa [21] consiste na seleção apenas das harmônicas, formando um espectro discreto, eliminando por construção as regiões entre harmônicas. Esta última abordagem será seguida na Seção 3.3 e seu modelo de síntese é analisado na Seção 4.3 do Cap. 4.

3.2 Modelagem de espectros contínuos

Embora a maior parte das aplicações atuais de análise LP partam do domínio do tempo pela disponibilidade do sinal de voz amostrado, originalmente Fumitada Itakura e Shuzo Saito [1, 2] propuseram a medida de distorção espectral

$$E_{\mathrm{ISC}} = \frac{1}{\pi}\int_0^{\pi}\left[\frac{P(\omega)}{\hat{P}(\omega)} - \log\frac{P(\omega)}{\hat{P}(\omega)} - 1\right]d\omega, \tag{3.1}$$

que passou a ser conhecida como distorção de Itakura-Saito, a partir da função de verossimilhança logarítmica para o que vieram a denominar de sintetizador de voz PARCOR ("partial autocorrelation") [22]. Numa denominação mais geral, os coeficientes PARCOR também são conhecidos como coeficientes de correlação parcial.

Com o modelo LP representado como

$$\hat{P}(\omega) = \frac{G^2}{\left|A\left(e^{j\omega}\right)\right|^2} \tag{3.2}$$

$$= \frac{G^2}{\left|1+\sum_{i=1}^{p} a_i e^{-j\omega i}\right|^2}, \tag{3.3}$$

a distorção de Itakura-Saito, em função dos parâmetros G e $\boldsymbol{a} = \begin{bmatrix} a_1 & a_2 & \dots & a_p \end{bmatrix}^T$, torna-se

$$E_{\mathrm{ISC}}(G,\boldsymbol{a}) = \frac{1}{\pi}\int_0^{\pi}\left[\frac{\left|A\left(e^{j\omega}\right)\right|^2 P(\omega)}{G^2} - \log\left(\left|A\left(e^{j\omega}\right)\right|^2 P(\omega)\right) + \log G^2 - 1\right]d\omega, \tag{3.4}$$

Primeiramente, determina-se o fator de ganho G em função do vetor de coeficientes de predição $\boldsymbol{a}$ através da minimização de $E_{\mathrm{ISC}}(G,\boldsymbol{a})$, que necessita a anulação da derivada parcial

$$\frac{\partial E_{\mathrm{ISC}}(G,\boldsymbol{a})}{\partial G^2} = \frac{1}{\pi}\int_0^{\pi}\left(-\frac{\left|A\left(e^{j\omega}\right)\right|^2 P(\omega)}{G^4} + \frac{1}{G^2}\right)d\omega, \tag{3.5}$$

que produz

$$G^2 = \frac{1}{\pi}\int_0^{\pi}\left|A\left(e^{j\omega}\right)\right|^2 P(\omega)\,d\omega \tag{3.6}$$

Substituindo o ganho dado por (3.6) na Eq. (3.4), obtém-se o erro apenas em função

dos coeficientes $\boldsymbol{a}$ como

$$\begin{aligned} E_{\mathrm{ISC}}(\boldsymbol{a}) &= 1 - \frac{1}{\pi}\int_0^\pi \log\left(\left|A\left(e^{j\omega}\right)\right|^2 P\left(\omega\right)\right) d\omega \\ &+ \log\left(\frac{1}{\pi}\int_0^\pi \left|A\left(e^{j\omega}\right)\right|^2 P\left(\omega\right) d\omega\right) - 1 \qquad (3.7) \\ &= -\frac{1}{\pi}\int_0^\pi \log\left(\left|A\left(e^{j\omega}\right)\right|^2 d\omega - \frac{1}{\pi}\int_0^\pi \log P\left(\omega\right) d\omega\right. \\ &+ \log\left(\frac{1}{\pi}\int_0^\pi \left|A\left(e^{j\omega}\right)\right|^2 P\left(\omega\right) d\omega\right) \qquad (3.8) \\ &= -\frac{1}{\pi}\int_0^\pi \log P\left(\omega\right) d\omega + \log\left(\frac{1}{\pi}\int_0^\pi \left|A\left(e^{j\omega}\right)\right|^2 P\left(\omega\right) d\omega\right) \qquad (3.9) \end{aligned}$$

em que foi usada a propriedade da média nula da densidade espectral logarítmica de potência da resposta em freqüência do filtro inverso, expressa pela Eq. (C.17) do Apêndice C.

Para a minimização da distorção de Itakura-Saito $E_{\mathrm{ISC}}(\boldsymbol{a})$ na Eq. (3.9), é necessário anular seu gradiente em relação a $\boldsymbol{a}$. Como a primeira parcela independe de $\boldsymbol{a}$ e a função $\log(\cdot)$ é monotônica crescente, a minimização de $E_{\mathrm{ISC}}(\boldsymbol{a})$ é equivalente à minimização da razão espectral desescalada [20], que pode ser representada como

$$\begin{aligned} E_{\mathrm{LPGC}} &= G^2 \frac{1}{\pi}\int_0^\pi \frac{P(\omega)}{\hat{P}(\omega)} d\omega \qquad (3.10) \\ &= \frac{1}{\pi}\int_0^\pi \left|A\left(e^{j\omega}\right)\right|^2 P\left(\omega\right) d\omega \qquad (3.11) \\ &= \frac{1}{2\pi}\int_{-\pi}^\pi \left|A\left(e^{j\omega}\right)\right|^2 P\left(\omega\right) d\omega \qquad (3.12) \end{aligned}$$

com extensão par do espectro dado às freqüências negativas, sendo $A\left(e^{j\omega}\right) = \sum_{i=0}^{p} a_i e^{-j\omega i}$ com $a_0 \triangleq 1$ de acordo com o modelo (3.3).

Inserindo-se o espectro de potência do filtro inverso do modelo acima na medida de distorção (3.12), obtém-se

$$\begin{aligned} E_{\mathrm{LPGC}} &= \frac{1}{2\pi}\int_{-\pi}^\pi \sum_{i=0}^{p}\sum_{m=0}^{p} a_i a_m P(\omega) e^{j\omega(i-m)} d\omega \qquad (3.13) \\ &= \sum_{i=0}^{p}\sum_{m=0}^{p} a_i a_m \frac{1}{2\pi}\int_{-\pi}^\pi P(\omega) e^{j\omega(i-m)} d\omega. \qquad (3.14) \end{aligned}$$

Sabe-se que, pelo Teorema de Wiener-Khinchine [23], a seqüência de autocorrelação é a transformada inversa de Fourier em tempo discreto da densidade espectral de potência, isto é,

$$R(m) = \frac{1}{2\pi}\int_{-\pi}^\pi P(\omega) e^{j\omega m} d\omega. \qquad (3.15)$$

Substituindo-se a antitransformada de Fourier (3.15) na expressão (3.14) do erro espectral LP contínuo, vem

$$E_{\mathrm{LPGC}} = \sum_{i=0}^{p} \sum_{m=0}^{p} a_i R(i-m) a_m, \tag{3.16}$$

que coincide com a Eq. (A.8) do erro quadrático temporal no caso particular em que a matriz de correlação estendida $\boldsymbol{\Psi}$ é a matriz de autocorrelação estendida $\boldsymbol{R}$.

Cada componente do gradiente da distorção LP contínua em relação a um coeficiente de predição é obtido como

$$\begin{aligned} \frac{\partial E_{\mathrm{LPGC}}}{\partial a_l} &= 2a_l R(0) + 2 \sum_{\substack{i=0 \\ i \neq l}}^{p} a_i R(i-l) && (3.17) \\ &= 2 \sum_{i=0}^{p} a_i R(i-l). && (3.18) \end{aligned}$$

Para minimização da distorção LP contínua, anula-se cada componente (3.18) do gradiente, obtendo-se o sistema de equações

$$\sum_{i=0}^{p} a_i R(i-l) = 0, \tag{3.19}$$

que, ao substituir-se $a_0 = 1$, torna-se

$$\sum_{i=1}^{p} a_i R(i-l) = -R(l), \tag{3.20}$$

para $l = 1, 2, \ldots, p$, que correspondem ao sistema de equações normais (B.1) particularizadas para o método da autocorrelação.

Com a determinação do vetor $\boldsymbol{a}$ de coeficientes de predição pela resolução do sistema de equações normais (3.20), pode-se voltar à Eq. (3.16) para determinar o fator de ganho que completa a determinação do modelo espectral e à Eq. (3.12) que agora determina o erro da razão espectral desescalada como

$$G^2 = E_{\mathrm{LPGC\,min}} = R(0) + \sum_{i=1}^{p} a_i R(i), \tag{3.21}$$

que coincide com o erro quadrático mínimo (B.16) para o método da autocorrelação. Ainda pode-se retornar à Eq. (3.9) para determinar a distorção de Itakura-Saito mínima

$$E_{\mathrm{ISC\,min}} = -c_0 + \log G^2 \tag{3.22}$$

em que

$$c_0 = \frac{1}{\pi}\int_0^{\pi} \log P(\omega)\, d\omega \tag{3.23}$$

representa o coeficiente cepstral de qüefrência nula do espectro original.

Alternativamente, pode-se usar o modelo de densidade espectral de potência

$$\hat{P}(\omega) = \left|H\left(e^{j\omega}\right)\right|^2 = \frac{1}{\left|A\left(e^{j\omega}\right)\right|^2} \tag{3.24}$$

$$= \frac{1}{\left|\sum_{i=0}^{p} a_i e^{-j\omega i}\right|^2} \tag{3.25}$$

em que $H\left(e^{j\omega}\right)$ e $A\left(e^{j\omega}\right)$ são as respostas em freqüência do filtro de síntese e do filtro inverso ou filtro de análise, respectivamente. Em comparação com o modelo (3.3), o fator de ganho G foi absorvido pelos coeficientes de predição, eliminando-se a restrição $a_0 = 1$ e deixando esse coeficiente livre para assumir qualquer valor. Para este modelo espectral, é conveniente usar a medida de distorção espectral dada pela média da razão espectral de densidade de potência na seguinte forma

$$E_{\mathrm{LPC}} = \frac{1}{\pi}\int_0^{\pi} \frac{P(\omega)}{\hat{P}(\omega)} d\omega \tag{3.26}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)} d\omega, \tag{3.27}$$

que corresponde ao erro quadrático no domínio do tempo, apresentado no Apêndice A, porque ambos conduzem ao sistema de equações normais discutido no Apêndice B, como será visto abaixo. De forma mais concreta, esta modelagem subentende um modelo de síntese fonte-filtro, representado esquematicamente na Fig. 2. Note-se bem que os modelos espectrais contínuos de ordem média menor que 20, tipicamente 10, não conseguem representar as componentes harmônicas dos sons sonoros, sendo mais eficiente neste caso a modelagem espectral discreta apresentada na Seção 3.3 e usada extensivamente no Cap. 4.

Inserindo-se o modelo (3.25) na medida de distorção (3.27), obtém-se

$$E_{\mathrm{LPC}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{i=0}^{p}\sum_{m=0}^{p} a_i a_m P(\omega) e^{j\omega(i-m)} d\omega \tag{3.28}$$

$$= \sum_{i=0}^{p}\sum_{m=0}^{p} a_i a_m \frac{1}{2\pi}\int_0^{\pi} P(\omega) e^{j\omega(i-m)} d\omega. \tag{3.29}$$

Substituindo-se a antitransformada de Fourier (3.15) na expressão (3.29) do erro es-

pectral LP contínuo, vem

$$E_{\mathrm{LPC}} = \sum_{i=0}^{p} \sum_{m=0}^{p} a_i R(i-m) a_m, \tag{3.30}$$

que coincide com a Eq. (A.8) do erro quadrático temporal no caso particular em que a matriz de correlação estendida $\boldsymbol{\Psi}$ é a matriz de autocorrelação estendida $\boldsymbol{R}$.

Cada componente do gradiente da distorção LP contínua em relação a um coeficiente de predição é obtido como

$$\frac{\partial E_{\mathrm{LPC}}}{\partial a_l} = 2a_l R(0) + 2 \sum_{\substack{i=0 \\ i \neq l}}^{p} a_i R(i-l) \tag{3.31}$$

$$= 2 \sum_{i=0}^{p} a_i R(i-l). \tag{3.32}$$

Para minimização da distorção LP contínua, anula-se cada componente (3.32) do gradiente, obtendo-se o sistema de equações

$$\sum_{i=0}^{p} a_i R(i-l) = 0, \tag{3.33}$$

que correspondem ao sistema de equações normais (B.1) para $l = 1, 2, \ldots, p$, particularizadas para o método da autocorrelação. Pode-se ainda acrescentar a equação $E_{\mathrm{LPC\,min}} = a_0 \sum_{i=0}^{p} a_i R(i)$, obtida a partir da Eq. (3.21), ao sistema de equações (3.33), que se torna, então,

$$\sum_{i=0}^{p} a_i R(i-l) = \frac{1}{a_0} \delta(l)$$

para $l = 0, 1, \ldots, p$.

Note-se, então, que, como $a_0 = 1/G$, a média da razão espectral de potência assume seu valor mínimo

$$E_{\mathrm{LPC\,min}} = 1, \tag{3.34}$$

que corresponde à densidade espectral média da fonte na Fig. 2 ou, equivalentemente, seu valor eficaz. Entretanto, nesse caso de coincidência, a distorção de Itakura-Saito assume o valor mínimo

$$E_{\mathrm{ISC\,min}} = 0, \tag{3.35}$$

que pode ser verificado diretamente da definição (3.1) e conferida na Eq. (3.22).

3.3 Modelagem espectral discreta

Como comentado na Seção 3.1, quando o espectro a ser modelado é em espectro de raias, isto é, apresenta regiões espectrais estreitas com alta concentração de potência em contraposição a outras regiões espectrais com baixíssimo conteúdo de potência, seriam necessários modelos espectrais de ordem elevada para ter capacidade de reproduzir esse amplo contraste de densidade de potência, sendo que o interesse situa-se apenas nas regiões de maior densidade espectral, podendo-se assumir densidade nula entre elas.

Assim, é interessante que a medida de distorção espectral para espectros discretos se atenha às regiões espectrais em torno das harmônicas. Será usada nesta seção a discretização da distorção de Itakura-Saito proposta em [21], que é dada por

$$E_{\mathrm{IS}} = \frac{1}{N}\sum_{k=1}^{N} \frac{P\left(\omega_k\right)}{\hat{P}\left(\omega_k\right)} - \log\frac{P\left(\omega_k\right)}{\hat{P}\left(\omega_k\right)} - 1, \tag{3.36}$$

onde $P\left(\omega_k\right)$ é o espectro discreto original e $\hat{P}\left(\omega_k\right)$ é o espectro do modelo, ambos tomados para as componentes $k = 1, 2, \ldots, N$. Usa-se nesta seção o modelo espectral dado pela Eq. (3.25).

Para a minimização da distorção espectral E_{IS} é necessário que se anule seu gradiente em relação ao vetor estendido de coeficientes de predição $\boldsymbol{\alpha} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_p \end{bmatrix}^T$, isto é, cada derivada parcial

$$\frac{\partial E_{\mathrm{IS}}}{\partial a_l} = \frac{1}{N}\sum_{k=1}^{N}\left[\left(1 - \frac{\hat{P}\left(\omega_k\right)}{P\left(\omega_k\right)}\right)\frac{\partial}{\partial a_l}\frac{P\left(\omega_k\right)}{\hat{P}\left(\omega_k\right)}\right] \tag{3.37}$$

para $l = 0, 1, \ldots, p$.

Introduzindo-se o modelo espectral (3.25) na derivada parcial do segundo membro da Eq. (3.37) e assumindo-se a convergência dos somatórios que compõem as derivadas parciais indicadas por esta, pode-se alterar a ordem de seus fatores para a seguinte com-

posição

$$\frac{\partial E_{\mathrm{IS}}}{\partial a_l} = \frac{1}{N}\sum_{k=1}^{N}\left[\left(1-\frac{\hat{P}\left(\omega_k\right)}{P\left(\omega_k\right)}\right)\frac{\partial}{\partial a_l}\sum_{i=0}^{p}\sum_{m=0}^{p}a_i a_m P\left(\omega_k\right)e^{j\omega_k(i-m)}\right] \tag{3.38}$$

$$= \frac{1}{N}\frac{\partial}{\partial a_l}\sum_{i=0}^{p}\sum_{m=0}^{p}a_i a_m \sum_{k=1}^{N}\left[\left(1-\frac{\hat{P}\left(\omega_k\right)}{P\left(\omega_k\right)}\right)P\left(\omega_k\right)e^{j\omega_k(i-m)}\right] \tag{3.39}$$

$$= \frac{\partial}{\partial a_l}\sum_{i=0}^{p}\sum_{m=0}^{p}a_i a_m \frac{1}{N}\sum_{k=1}^{N}\left[\left(P\left(\omega_k\right)-\hat{P}\left(\omega_k\right)\right)e^{j\omega_k(i-m)}\right] \tag{3.40}$$

$$= \frac{\partial}{\partial a_l}\sum_{i=0}^{p}\sum_{m=0}^{p}a_i a_m \frac{1}{N}\sum_{k=1}^{N}\left[\left(P\left(\omega_k\right)-\hat{P}\left(\omega_k\right)\right)\cos\left(\omega_k(i-m)\right)\right] \tag{3.41}$$

para $l = 0, 1, \ldots, p$, tendo-se usado a igualdade (D.9) na última passagem.

Como $\hat{P}\left(\omega_k\right)$ na Eq. (3.41) também depende do vetor $\boldsymbol{\alpha}$, ela representa um sistema não-linear de equações. Entretanto, um artifício adotado em [21] consiste em mantê-lo independente, o que conduz a um sistema linear de equações a ser resolvido por um algoritmo iterativo. Para obtê-lo, usa-se a representação discretizada

$$R(m) = \frac{1}{N}\sum_{k=1}^{N}P(\omega_k)\cos\left(\omega_k m\right) \tag{3.42}$$

da expressão (3.15) do Teorema de Wiener-Khinchine para transformar o sistema de equações (3.41) em

$$\frac{\partial E_{\mathrm{IS}}}{\partial a_l} = \frac{\partial}{\partial a_l}\sum_{i=0}^{p}\sum_{m=0}^{p}a_i a_m\left(R(i-m)-\hat{R}(i-m)\right) \tag{3.43}$$

$$= 2\left(\sum_{i=0}^{p}a_i R(i-l)-\sum_{i=0}^{p}a_i\hat{R}(i-l)\right) \tag{3.44}$$

para $l = 0, 1, \ldots, p$, cujas componentes devem ser uma a uma anuladas conforme comentado acima quando da introdução da Eq. (3.37). Assim, obtém-se o sistema de equações

$$\boldsymbol{R\alpha} = \hat{\boldsymbol{R}}\boldsymbol{\alpha}, \tag{3.45}$$

que pode ser usado em conjunto com um modelo de síntese para a obtenção de um sistema de equações lineares para resolução iterativa como delineado na Seção 4.3.

4 *Modelos preditivos para processamento temporal e espectral*

4.1 Introdução

As técnicas de análise por predição linear (LP) permitem obter modelos que podem ser usados para processamento temporal, definindo-se filtros com seus parâmetros, ou para processamento espectral em que amostras das respostas em freqüência dos modelos são tomadas para compor um sinal. Usualmente, os codificadores de voz aplicam uma ou outra dessas modalidades. Entretanto, neste trabalho o interesse recai na decomposição em ordem dos modelos LP, aplicados para pré-processamento temporal de filtragem anterior à extração de ciclos de onda e para pós-processamento de modelagem espectral harmônica para quantização vetorial dos espectros de evolução lenta. Para essa decomposição em ordem, desenvolveram-se métodos chamados de predição linear por decomposição em ordem ou "split-order LP" (SOLP) [9].

A filtragem pelo filtro inverso do modelo LP produz um sinal residual com picos mais uniformes e localizados, facilitando a estimação do período fundamental e a extração dos ciclos de onda fundamentais, como foi observado na Seção 1.1. Por outro lado, o ajuste de modelos LP aos espectros harmônicos dos ciclos de evolução lenta permite representá-los por conjuntos de pares de raias espectrais (LSPs), que são usados comumente para quantização vetorial (VQ) [24] em codificadores de voz devido a sua baixa sensibilidade às operações de quantização e de interpolação [25]. Propõe-se usar uma ordem baixa para a filtragem e uma ordem bem mais alta para a modelagem espectral de tal forma que o modelo conjunto seja de ordem 10, usual em codificadores WI apenas para filtragem [8], ou de ordem 14, usual em modelos espectrais para codificadores senoidais [26]. Portanto, no caso de codificadores WI, um dos objetivos deste trabalho é codificar em conjunto a informação da configuração do filtro e da sua excitação, pelo menos sua componente

periódica, com o mesmo número de bits usados apenas para a codificação da configuração do filtro num codificador WI comum.

De um ponto de vista estatístico, a estimação do modelo espectral para o espectro harmônico é o segundo estágio, dado o primeiro estágio que é a estimação da periodicidade. De fato, a própria existência explícita do espectro harmônico decorre da estimativa da periodicidade e da conseqüente localização do ciclo.

Em relação à métrica usada para avaliar o erro de predição em cada análise LP, as necessidades discriminatórias são diferentes para cada estágio. No primeiro estágio, necessitam-se métricas enviesadas em direção aos picos espectrais pois as formantes devem ser bem discriminadas. Já no segundo estágio, em que se modela um espectro harmônico, os desvios positivos e negativos da modelagem devem ser igualmente repelidos. Em outras palavras, no segundo estágio devem-se usar métricas simétricas. Entretanto, é comum o emprego de medidas de distorção assimétricas, como a distorção de Itakura-Saito, mesmo neste caso [21] embora haja registro do emprego bem sucedido da distorção *cosh*, simétrica [27].

Para a proposta de uma medida de distorção simétrica, toma-se como alvo a medida espectral logarítmica ("log SD") porque ela é considerada a distorção ótima em quantização de alta resolução ("high rate quantization") [28] tendo em vista que as distâncias euclidianas ponderadas entre vetores de LSPs se aproximam da log SD à medida que a resolução da quantização é aumentada. Além disso, a log SD é equivalente à distância cepstral enquanto a distorção cosh a aproxima por cima [29].

Apesar da divergência da log SD em algoritmos iterativos de análise LP para espectros discretos, descreve-se aqui um método para a geração de medidas de distorção simétricas. O método baseia-se na separação das componentes positiva e negativa do gradiente da log SD. Ainda, mostra-se que o emprego de aproximações linear e cúbica ao gradiente da log SD com este método resulta em menores distorções espectrais que as obtidas quando se usam medidas de distorção assimétricas.

4.2 O contexto para extração de ciclos e modelagem espectral

Os ciclos de onda são extraídos do sinal residual $r(n)$, obtido pela filtragem inversa do sinal de voz $s(n)$ com modelos LP calculados pelo método da autocorrelação como mostrado esquematicamente na Fig. 10. Em seguida, os ciclos residuais $\{r_c(n,m)\}_{m=0}^{l_0(n)-1}$

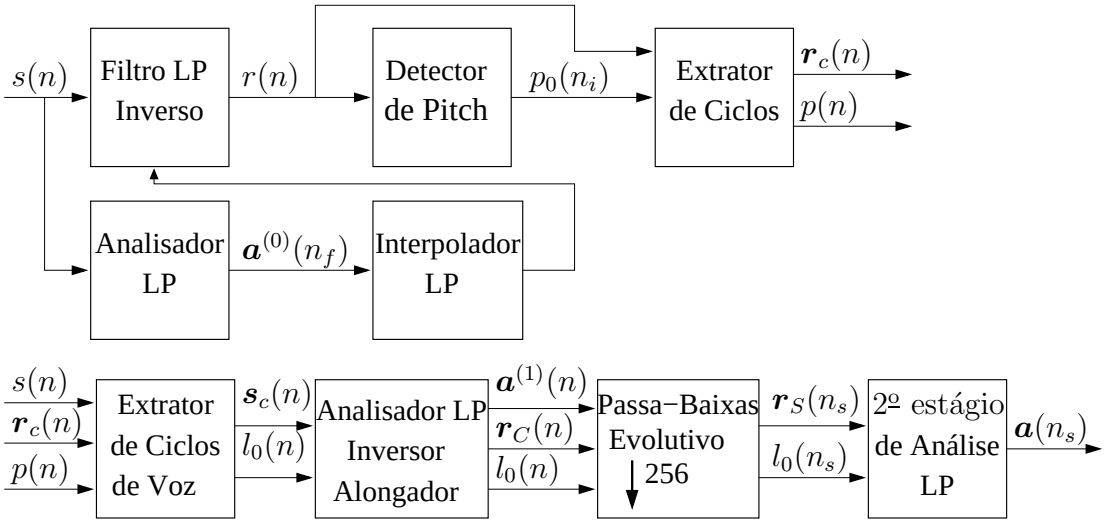


Figura 10: Extração de ciclos com análise LP de primeiro estágio e modelagem espectral no segundo estágio de análise LP da SOLP.

são extraídos [11, 10] com o algoritmo descrito na Seção 2.3 e também os correspondentes ciclos do sinal de voz $\{s_c(n,m)\}_{m=0}^{l_0(n)-1}$, onde $l_0(n)$ representa o comprimento do ciclo no instante n, estimado a partir do período de pitch $p_0(n_i)$ estimado pelo detector robusto de "pitch". Com as extremidades dos ciclos localizadas, os modelos LP do primeiro estágio $\boldsymbol{a}^{(1)}(n) = \left\{a_i^{(1)}(n)\right\}_{i=0}^{p_1}$ são determinados através de uma análise LP baseada nos coeficientes de autocorrelação circulares [30]. Estes modelos são, então, utilizados para determinar os ciclos residuais em regime periódico permanente. Os comprimentos $l_0(n)$ dos ciclos residuais são registrados antes de ser alongados para um comprimento comum, como descrito na Seção 2.4, produzindo as formas de onda características (CWs) $\{r_C(n,m)\}_{m=0}^{L-1}$. Em seguida, a superfície CW é filtrada passa-baixas na direção da evolução [31], obtendo-se formas de onda de evolução lenta (SEWs) $\{r_S(n_s,m)\}_{m=0}^{L-1}$ na taxa de 31,25 CW/s pela aplicação de um fator 256 de subamostragem. Em algumas condições de teste no Capítulo 5, a subamostragem é aplicada sem a prévia filtragem para avaliação da influência da filtragem sobre a modelagem espectral. Finalmente, efetua-se a análise LP de segundo estágio, obtendo o conjunto de vetores de coeficientes $\boldsymbol{a}(n_s) = \{a_i(n_s)\}_{i=0}^{p}$ e seus modelos espectrais resultantes da amostragem de cada densidade espectral de potência LP

$$\hat{P}(\omega) = \frac{1}{|A\left(e^{j\omega}\right)|^2} = \frac{1}{|\sum_{i=0}^{p} a_i e^{-j\omega i}|^2}. \tag{4.1}$$

nas freqüências $\omega = \omega_k = k\omega_1$, onde $\omega_1 = 2\pi/l_0$.

4.3 Modelagem espectral baseada na excitação periódica de filtros definidos por modelos autorregressivos

Um modelo de síntese para o problema da modelagem espectral discreta consiste num filtro excitado por um pente ou seqüência de pulsos periódicos de período l_0. Este modelo leva à seguinte recorrência da seqüência de autocorrelação da resposta impulsiva $\{h(n)\}$ do filtro autorregressivo $H(z)$

$$\sum_{i=0}^{p} a_i R_{hh}(m-i) = h(-m), \tag{4.2}$$

onde $\{R_{hh}(m)\}$ é um ciclo dessa seqüência de autocorrelação. Assume-se também que os ciclos $\{R_{hh}(m)\}$ e $\{h(n)\}$ na Eq. (4.2) estejam centrados em $m = 0$ e que os índices $m-i$ e $-m$ na Eq. 4.2 sejam tomados módulo l_0.

No clássico modelo autorregressivo discreto ("discrete all-pole" DAP) [21], a medida de distorção é a versão discreta da distorção de Itakura-Saito, expressa como

$$E_{\mathrm{IS}} = \frac{1}{N}\sum_{k=1}^{N} \frac{P(\omega_k)}{\hat{P}(\omega_k)} - \log\frac{P(\omega_k)}{\hat{P}(\omega_k)} - 1, \tag{4.3}$$

onde $P(\omega_k)$ é o espectro discreto original e $\hat{P}(\omega_k)$ é o espectro do modelo, ambos tomados para $k = 1, 2, \ldots, N$.

A minimização de E_{IS} na Eq. (4.3), após substituir $\hat{P}(\omega_k)$ da Eq. (4.1), com respeito ao conjunto $\{a_i\}_{i=0}^{p}$ de coeficientes de predição produz o conjunto de equações não-lineares deduzido na Seção 3.3 e reproduzido aqui como

$$\boldsymbol{R}\boldsymbol{a} = \hat{\boldsymbol{R}}\boldsymbol{a}, \tag{4.4}$$

onde $\boldsymbol{R}$ é a matriz de Toeplitz simétrica determinada pelo vetor $\begin{bmatrix} R(0) & R(1) & \cdots & R(p) \end{bmatrix}^T$ de coeficientes de autocorrelação circular associado ao espectro de potência dado, $\boldsymbol{a} = \begin{bmatrix} a_0 & a_1 & \ldots & a_p \end{bmatrix}^T$ é o vetor estendido de coeficientes de predição e $\hat{\boldsymbol{R}}$ é a matriz de autocorrelação $(p+1)\times(p+1)$ associada ao modelo espectral $\hat{P}(\omega_k)$.

O modelo de síntese (4.2) pode ser usado para expressar a Eq. (4.4) como

$$\boldsymbol{R}\boldsymbol{a} = \hat{\boldsymbol{h}} \tag{4.5}$$

Esta fórmula é empregada no algoritmo DAP [21] como um passo intermediário na

resolução do sistema de equações não-lineares (4.4) por um algoritmo linear iterativo constituído de duas fases. Na etapa 1, calcula-se o vetor $\hat{\boldsymbol{h}}$ de amostras da resposta impulsiva periódica a partir de um modelo de síntese $\hat{H}(z)$ inicial. Então, na etapa 2, o sistema linear de equações (4.5) é resolvido em busca do vetor de coeficientes $\boldsymbol{a}$, que redefine o modelo $\hat{H}(z)$ para a fase 1 da iteração seguinte. De fato, introduz-se o fator de esquecimento α entre as iterações consecutivas m e $m+1$, que é empregado como

$$\boldsymbol{a}_{m+1} = \alpha \boldsymbol{a}_m + (1-\alpha)\boldsymbol{R}^{-1}\hat{\boldsymbol{h}}. \tag{4.6}$$

4.4 Medidas de distorção simétricas para modelagem espectral harmônica

No caso da predição linear com decomposição da ordem (SOLP), é mais adequado empregar-se uma medida de distorção simétrica no segundo estágio que envolve a modelagem de amplitudes harmônicas. Isto já foi salientado para o caso geral de modelagem espectral harmônica [27] usando a medida de distorção cosh . Como indicado na Seção 4.1, a medida cosh

$$E_{\mathrm{C}} = \frac{1}{N}\sum_{k=1}^{N} \frac{P(\omega_k)}{\hat{P}(\omega_k)} + \frac{\hat{P}(\omega_k)}{P(\omega_k)} - 2 \tag{4.7}$$

aproxima-se da log SD por cima.

Para obter uma redução maior da distorção espectral, adota-se a log SD

$$E_{\mathrm{SD}} = \frac{1}{N}\sum_{k=1}^{N} \log^2 \frac{P(\omega_k)}{\hat{P}(\omega_k)} \tag{4.8}$$

como medida de distorção. Para sua minimização, de acordo com o Apêndice D, cada componente de seu gradiente

$$\frac{\partial}{\partial a_l} E_{\mathrm{SD}} = \frac{4}{N}\sum_{i=0}^{p} a_i \sum_{k=1}^{N} \log \frac{P(\omega_k)}{\hat{P}(\omega_k)} \hat{P}(\omega_k) \cos\left(w_k(i-l)\right) \tag{4.9}$$

em relação aos coeficientes de predição a_l, para $l = 0, 1, \ldots p$, dever ser nulo.

Conjetura-se que se poderia obter um sistema de equações semelhante à Eq. (4.5) se cada componente do gradiente pudesse ser expressa na forma

$$\frac{\partial}{\partial a_l} E_{\mathrm{SD}} = \frac{4}{N}\sum_{i=0}^{p} a_i \sum_{k=1}^{N} \left[P_o(\omega_k) - P_r(\omega_k)\right] \cos\left(\omega_k(i-l)\right) \tag{4.10}$$

em que $P_o(\omega_k)$ são componentes espectrais preponderantemente baseadas nas componen-

tes originais e $P_r(\omega_k)$ são componentes espectrais baseadas principalmente em componentes espectrais estimadas anteriormente. Se esta separação for atingida, o sistema de equações que deverá ser resolvido em cada iteração do algoritmo será

$$\boldsymbol{R}_o \boldsymbol{a} = \boldsymbol{h}_r, \tag{4.11}$$

em que $\boldsymbol{h}_r = \boldsymbol{R}_r \boldsymbol{a}$ e $\boldsymbol{R}_o$ e $\boldsymbol{R}_r$ são matrizes de Toeplitz $(p+1) \times (p+1)$ simétricas geradas pelas séries de Fourier inversas de $\{P_o(\omega_k)\}_{k=1}^{N}$ e $\{P_r(\omega_k)\}_{k=1}^{N}$, respectivamente.

Em primeiro lugar, expandindo-se a razão espectral logarítmica na Eq. (4.9), obtém-se as componentes espectrais

$$P_o(\omega_k) = \log\left(P(\omega_k)\right) \hat{P}(\omega_k) \tag{4.12}$$

$$P_r(\omega_k) = \log\left(\hat{P}(\omega_k)\right) \hat{P}(\omega_k). \tag{4.13}$$

Ainda é necessário garantir que estas componentes sejam verdadeiramente espectros de potência. Para isso, há que se recorrer a multiplicadores adaptativos tanto para as componentes espectrais originais $P(\omega_k)$ quanto para as estimadas $\hat{P}(\omega_k)$, que devem ser determinados de forma que seus logaritmos modificados se tornem não-negativos. Mesmo com este esforço adicional não é possível garantir a convergência do algoritmo.

O escalonamento espectral poderia ser evitado se a razão espectral logarítmica que existe no gradiente puder ser escrita como a diferença entre duas componentes positivas. Esta situação pode ser alcançada pela aproximação da função logaritmo por sua série de Taylor truncada. A expansão

$$\log\frac{1+x}{1-x} = 2\sum_{m=1}^{\infty}\frac{x^{2m-1}}{2m-1} \tag{4.14}$$

em série pode ter sua convergência garantida quando o argumento da função logarítmica for a razão espectral logarítmica. Para que isto aconteça, a variável de expansão deve ser

$$x = \frac{P(\omega_k) - \hat{P}(\omega_k)}{P(\omega_k) + \hat{P}(\omega_k)}. \tag{4.15}$$

Substituindo x da Eq. (4.15) na Eq. (4.14) e reagrupando os termos positivos e negativos da aproximação linear, resulta nas seguintes componentes

$$P_o(\omega_k) = 2\frac{P(\omega_k)\,\hat{P}(\omega_k)}{P(\omega_k) + \hat{P}(\omega_k)} \tag{4.16}$$

$$P_r(\omega_k) = 2\frac{\hat{P}^2(\omega_k)}{P(\omega_k) + \hat{P}(\omega_k)} \tag{4.17}$$

que, ao ser aplicada à Eq. (4.11), dá origem ao método DAPSD1 de estimação harmônica. A aproximação seguinte é a cúbica, que gera as seguintes componentes

$$P_o(\omega_k) = 2\left[\frac{P(\omega_k)\hat{P}(\omega_k)}{P(\omega_k)+\hat{P}(\omega_k)} + \frac{1}{3}\frac{P^3(\omega_k)\hat{P}(\omega_k)+3P(\omega_k)\hat{P}^3(\omega_k)}{\left(P(\omega_k)+\hat{P}(\omega_k)\right)^3}\right] \tag{4.18}$$

$$P_r(\omega_k) = 2\left[\frac{\hat{P}^2(\omega_k)}{P(\omega_k)+\hat{P}(\omega_k)} + \frac{1}{3}\frac{3P^2(\omega_k)\hat{P}^2(\omega_k)+\hat{P}^4(\omega_k)}{\left(P(\omega_k)+\hat{P}(\omega_k)\right)^3}\right] \tag{4.19}$$

que, quando aplicada à Eq. (4.11), constitui o método DAPSD3 de estimação harmônica.

Enquanto as Eqs. (4.16) e (4.17) são computacionalmente tratáveis, as Eqs. (4.18) e (4.19) são muito complexas para implementação mas serão analisadas também para ajudar a estabelecer a tendência de desempenho que se pode esperar a partir do aumento progressivo de precisão no truncamento da série de Taylor da razão espectral logarítmica.

As Eqs. (4.5) ou (4.11) são nucleares para os métodos DAP e DAPSD1, respectivamente. A resolução destes sistemas de equações pode ser simplificada observando-se que $\boldsymbol{R}$ e $\boldsymbol{R}_o$ são matrizes de Toeplitz simétricas cujas inversas podem ser determinadas eficientemente pela relação de Gohberg-Semencul [32]. Então, a matriz inversa decomposta é multiplicada pelo vetor $\hat{\boldsymbol{h}}$ ou $\boldsymbol{h}_r$, respectivamente, determinando-se assim o vetor de coeficientes de predição. Como $\boldsymbol{R}$ é constante em cada ciclo, estima-se que a complexidade operacional de DAPSD1 seja aproximadamente o dobro da complexidade do método DAP.

5 *Experimentos com modelagem espectral*

5.1 Condições de experimentação

O processo conjunto de extração de ciclos e de modelagem espectral foi testado para ordens conjuntas 10 e 14 de modelagem LP, que são comuns na análise LP para processamento temporal e para modelagem espectral discreta de sinais amostrados em 8 kHz, respectivamente. Além disso, a ordem de predição conjunta foi decomposta de tal forma a ter modelos de segunda e quarta ordem no primeiro estágio LP. No segundo estágio de análise LP para modelagem espectral harmônica, foram comparados os seguintes métodos:

i) Modelagem espectral discreta "all-pole" (DAP) usando a medida de distorção de Itakura-Saito;

ii) Modelagem espectral discreta "all-pole" (DAP) usando a medida de distorção cosh;

iii) Modelagem espectral discreta "all-pole" usando uma aproximação linear da razão espectral logarítmica - DAPSD1;

iv) Modelagem espectral discreta "all-pole" usando uma aproximação cúbica da razão espectral logarítmica - DAPSD3.

Estes quatro algoritmos iterativos de modelagem espectral discreta "all-pole" foram simulados com fator $\alpha = 0,5$ de esquecimento até que o máximo de $i_{\max} = 16$ iterações seja atingido ou até que a flutuação entre iterações sucessivas seja menor que $1 \cdot 10^{-4}$ dB em E_{IS} para o primeiro algoritmo e em E_{SD} para os três últimos.

Como sinais de teste, foram usadas todas as 1680 orações na partição de teste da base de dados de voz TIMIT, reamostrados em 8 kHz e distribuídos como dois terços de falantes masculinos e um terço de falantes femininos, totalizando um tempo de gravação de 1781,1 s de fala feminina e 3405,6 s de fala masculina. Convém enfatizar que todos

os segmentos dos sinais de voz foram modelados, independentemente de sua sonoridade, porque o modelo WI de evolução do sinal admite a extração de ciclos de onda de duração arbitrária quando o segmento em questão for surdo, de acordo com a exposição do Cap. 2. A distorçao espectral (SD) logarítmica foi usada para comparar os valores espectrais discretos modelados em freqüências múltiplas da fundamental com os valores espectrais originais correspondentes. Ela foi calculada como

$$D_{\mathrm{S}} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(10\log_{10}P\left(\omega_k\right) - 10\log_{10}\hat{P}\left(\omega_k\right)\right)^2} \tag{5.1}$$

sendo que ambos os espectros logarítmicos, $\{10\log_{10}P\left(\omega_k\right)\}_{k=1}^{N}$ e $\left\{10\log_{10}\hat{P}\left(\omega_k\right)\right\}_{k=1}^{N}$, são anteriormente normalizados de forma que seus valores médios sejam nulos como sugerido por [33] e com base nessa propriedade espectral dos espectros obtidos por LP, cuja demonstração é apresentada no Apêndice C.

5.2 Resultados dos experimentos

A característica mais evidente nos resultados apresentados da Tabela 1 à Tabela 3 são os níveis de log SD no segundo estágio da modelagem LP com decomposição de ordem, situados significativamente abaixo de 1 dB quando $p_1 = 2$. A título de comparação, este nível é um limiar típico de distorção para projeto de quantizadores [24, 34]. Portanto, deve-se ter cautela nesta comparação porque se trata de uma modelagem neste caso. De qualquer forma, constitui-se numa melhora significativa alcançada pela análise LP em dois estágios sobre a modelagem espectral LP em estágio único uma vez que, tanto em tarefas de interpolação de forma de onda, como mostrado em estudos preliminares [12], quanto na codificação senoidal [33, 35], os métodos de estimação espectral apresentam desempenho dentro do mesmo intervalo de níveis de distorção de 3 dB a 4 dB.

Para avaliar o que acontece quando o contexto de modelagem usado nesta pesquisa é aplicado num único estágio de ordem p_2, o sinal de voz é segmentado nas mesmas extremidades que o sinal residual de ordem p_0 seria e não se insere nenhuma filtragem evolutiva. Os resultados são apresentados nas Tabelas 4 e 5, em que as distorções médias se situam muito próximas para as duas medidas de distorção ao passo que a fração de excedentes de 6 dB é consideravelmente inferior para a DAPSD1. Além disso, comparando-se os resultados para a mesma ordem de modelagem, depreende-se que a estimação do "pitch" e a segmentação dos ciclos não apresentam muita sensibilidade à ordem de predição do

Tabela 1: Distorção espectral logarítmica para as modelagens espectrais DAP, DAPC, DAPSD1 e DAPSD3 de ciclos processados sem filtragem evolutiva, extraídos de locuções masculinas.

Modelagem ordens		Distorção espectral logarítmica (dB) do segundo estágio e fração de excedentes de 2 dB (%)							
p_1	p_2	DAP		DAPC		DAPSD1		DAPSD3	
		(dB)	(%)	(dB)	(%)	(dB)	(%)	(dB)	(%)
2	8	0,73	9,87	0,73	9,4	0,72	9,12	0,72	9,10
2	12	0,42	3,40	0,42	3,10	0,41	2,93	0,41	2,91
4	6	1,56	31,23	1,56	30,9	1,53	29,73	1,53	29,71
4	10	0,90	9,67	0,91	9,10	0,88	8,50	0,88	8,47

Tabela 2: Distorção espectral logarítmica para as modelagens espectrais DAP, DAPC, DAPSD1 e DAPSD3 de ciclos processados sem filtragem evolutiva, extraídos de locuções femininas.

Modelagem ordens		Distorção espectral logarítmica (dB) do segundo estágio e fração de excedentes de 2 dB (%)							
p_1	p_2	DAP		DAPC		DAPSD1		DAPSD3	
		(dB)	(%)	(dB)	(%)	(dB)	(%)	(dB)	(%)
2	8	0,78	11,83	0,78	11,01	0,76	10,84	0,76	10,82
2	12	0,46	3,90	0,46	3,45	0,45	3,37	0,45	3,34
4	6	1,88	43,84	1,86	43,22	1,82	42,08	1,82	42,06
4	10	1,18	18,44	1,18	17,10	1,14	16,46	1,14	16,38

sinal residual.

Além disso, uma comparação geral dos resultados de distorção espectral logarítmica da Tabela 1 à Tabela 3 leva à conclusão de que os resultados das análise DAPSD (1 e 3) são uniformemente menores que aqueles das análises DAP e DAPC, com particular destaque para o caso em que $p_1 = 4$.

Portanto, um exame preliminar geral dos resultados revela que a análise DAPSD1 é o método de preferência para o segundo estágio da predição linear com decomposição de ordem contanto que a complexidade adicional possa ser absorvida. Entretanto, para a aproximação linear o acréscimo de complexidade não é tão alto enquanto que para a aproximação cúbica o acréscimo de complexidade não compensa sua aplicação.

Como resultado mais geral, ficou estabelecido que a medida de distorção simétrica proposta neste trabalho é mais apropriada para a modelagem de amplitudes harmônicas, proporcionando melhor ajuste de modelagem como assinalado pela menor proporção de

Tabela 3: Distorção espectral logarítmica das modelagens espectrais DAP e DAPSD1 de ciclos processados com filtragem evolutiva, extraídos de locuções masculinas e femininas.

Modelagem ordens		Distorção espectral logarítmica (dB) do segundo estágio e fração de excedentes de 2 dB (%)							
		masculinas				femininas			
p_1	p_2	DAP		DAPSD1		DAP		DAPSD1	
		(dB)	(%)	(dB)	(%)	(dB)	(%)	(dB)	(%)
2	8	0,90	9,34	0,88	8,45	0,91	9,72	0,90	8,67
2	12	0,61	4,44	0,60	3,96	0,59	3,90	0,60	3,40
4	6	1,60	28,27	1,56	26,87	1,79	36,99	1,77	35,05
4	10	1,07	13,60	1,04	12,51	1,20	17,23	1,19	15,53

Tabela 4: Distorção espectral logarítmica modelagens espectrais em estágio único DAP e DAPSD1 de ciclos extraídos de locuções masculinas.

Ordens de modelagem			Distorção espectral log em estágio único(dB)		Fração de excedentes de 6 dB (%)	
p_0	p_1	p_2	DAP	DAPSD1	DAP	DAPSD1
2	0	10	4,09	4,08	10,22	8,87
2	0	14	3,44	3,53	4,94	4,26
4	0	10	4,15	4,13	10,39	8,96
4	0	14	3,52	3,58	4,91	4,20

Tabela 5: Distorção espectral logarítmica modelagens espectrais em estágio único DAP e DAPSD1 de ciclos extraídos de locuções femininas.

Ordens de modelagem			Distorção espectral log em estágio único (dB)		Fração de excedentes de 6 dB (%)	
p_0	p_1	p_2	DAP	DAPSD1	DAP	DAPSD1
2	0	10	3,89	3,88	9,38	8,05
2	0	14	3,16	3,29	4,66	4,01
4	0	10	3,93	3,91	9,51	8,06
4	0	14	3,22	3,32	4,69	3,98

excedentes.

Comparando a Tabela 3 com a Tabela 1 ou a Tabela 2, pode-se notar que o desempenho melhora para todos os métodos quando a filtragem evolutiva é inserida no caso das ordens de predição $p_0 = 4$ e $p_1 = 6$, principalmente para voz feminina, sendo este um resultado muito interessante para a codificação.

Uma amostra de espectro harmônico de um ciclo de voz feminina é mostrado na Fig. 11 em que os melhores ajustes da modelagem espectral são obtidos com as estimativas DAPSD1 e DAPC, em particular para a primeira e a última componentes harmônicas, mas também para várias subseqüências de componentes harmônicas intermediárias.

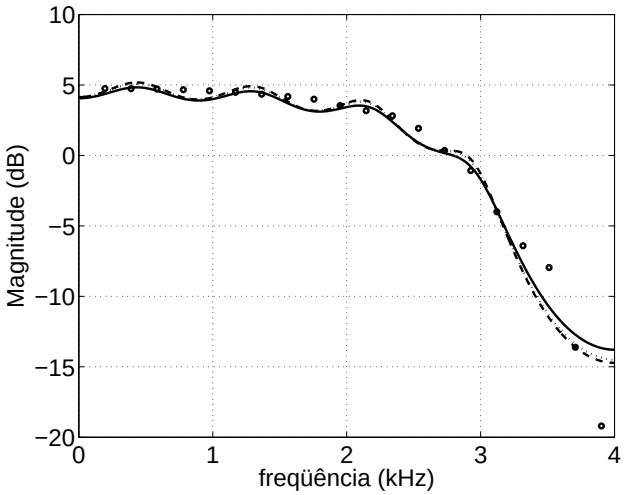


Figura 11: O espectro discreto de um ciclo de voz feminina (círculos), seu modelo DAP (curva contínua), seu modelo DAP cosh (curva tracejada) e seu modelo DAP SD linearmente aproximado (curva pontilhada), que se situa sempre entre os dois últimos modelos. Todos os modelos são de 8ª ordem após um primeiro estágio de 2ª ordem cujos ciclos residuais sofrem filtragem passa-baixas evolutiva.

Finalmente, no primeiro estágio um filtro de análise LP de segunda ordem parece ser mais eficiente para a extração dos ciclos bem como para deixar suficiente redundância no espectro harmônico residual que garante uma preditibilidade elevada no segundo estágio. Por outro lado, o primeiro estágio de quarta ordem, apesar de ser muito conveniente para a predição completa de espectros surdos e de ter um desempenho superior com a filtragem evolutiva, não atinge uma redução de distorção tão acentuada quanto o primeiro estágio de segunda ordem, talvez por ser seguido por um modelo espectral de ordem mais baixa.

5.3 Conclusão

O problema da modelagem espectral harmônica, comumente abordado como uma tarefa única, foi tratado com uma análise preditvia em dois estágios. A ordem total de predição é decomposta entre os dois estágios. O primeiro estágio de LP é responsável pela competente extração de ciclos, que prepara o terreno para o segundo estágio de LP em que a tarefa de modelagem espectral é realmente executada. Foram revistos vários métodos de análise, algumas medidas simétricas de distorção espectral foram propostas para esta tarefa e um método foi proposto que gera várias medidas simétricas que aproximam a medida de distorção espectral logarítmica. Nesta nova tarefa de LP com decomposição de ordem (SOLP) todos os métodos de modelagem discreta "all-pole" testados resultaram em distorções espectrais logarítmicas inferiores a 1 dB no segundo estágio quando o primeiro estágio é de segunda ordem. Além disso, a aproximação linear da razão espectral logarítmica, denominada de DAPSD1, proporciona melhor ajuste de modelagem para todas as situações testadas. Estes resultados em relação à qualidade e suas naturais conseqüências para a redução da taxa de transmissão são promissores para a codificação de voz em taxas baixas.

6 Conclusão

A modelagem LP é usada principalmente para representar espectros contínuos de sinais de voz com larga aplicação nos codecs, nos vocoders e nos reconhecedores de voz. Entretanto, os segmentos sonoros do sinal de voz apresentam espectros harmônicos, cujo ajuste por modelos contínuos é imperfeito. Especialmente, no caso geral dos codificadores senoidais e dos codificadores WI em particular, a natureza discreta do espectro de voz é ressaltada devido à singularização das harmônicas no caso senoidal e à singularização dos ciclos de onda no caso da WI. A proposta de modelagem LP em tempo-freqüência que foi exposta neste trabalho situa-se principalmente neste último contexto em que a seqüência das modelagens é temporal primeiro e depois espectral.

Nesse contexto, verificou-se a possibilidade de manutenção da ordem total de predição próxima aos valores usados na modelagem LP temporal, que é 10 no caso dos codificadores de voz CELP ("code-excited LP") para sinais com largura de faixa telefônica estreita. Também foi estudado o caso de ordem total 14, usada, via de regra, na modelagem LP espectral dos codificadores senoidais. Assim, na modelagem SOLP proposta, o primeiro estágio LP é usado para definir um filtro inverso que produz um sinal residual usado pelo extrator de ciclos de onda para separar a periodicidade e a forma de cada ciclo. A ordem deste primeiro estágio LP pode ser baixa, tendo sido estudados os casos de 2ª e de 4ª ordem. Dessa forma, as possibilidades de ordens LP para o segundo estágio foram 8 e 12 no primeiro caso e 6 e 10 no segundo.

Constatou-se que a modelagem LP discreta ajusta-se com log SD inferior a 1 dB aos espectros harmônicos associados aos ciclos de onda extraídos. Em termos de modelagem espectral, este valor é muito inferior ao erro de modelagem espectral direta, que se situa entre 3 dB e 4 dB. Mais importante ainda, do ponto de vista da codificação de voz, abre-se a possibilidade de o modelo SOLP, que já embute a representação do espectro harmônico, ser incorporado exatamente no lugar em que nos codificadores normais é colocado apenas o modelo LP, tornando-se desnecessária a transmissão do espectro harmônico à parte. A ordem do modelo SOLP é equivalente à ordem do modelo LP, como foi discutido acima.

Resta ainda pesquisar se o número de bits necessário para representar o modelo SOLP pode ser próximo ao número de bits usado para a representação do modelo LP. Em caso de confirmação, haveria um ganho de compressão que pode ser considerável.

Para o segundo estágio da modelagem SOLP, foi proposto e desenvolvido um método de determinação de distorções espectrais simétricas capazes de produzir sistemas de equações LP com resolução iterativa. A essência do método está na separação do gradiente do erro de ajuste espectral em dois espectros que são garantidamente espectros de potência. No caso em que o erro espectral é a log SD, o método se traduz na representação da razão espectral logarítmica de potência através de uma série de potências. Em testes de modelagem SOLP, verificou-se que os erros log SD de modelagem apresentam melhor distribuição estatística neste caso do que quando se usam a distorção de Itakura-Saito ou a distorção *cosh*. Há detalhes de processamento matricial na implementação da modelagem LP discreta com medidas de distorção simétricas que podem vir a se configurar em novos algoritmos.

APÊNDICE A – Erro quadrático de predição linear

Na predição linear (LP) uma amostra $s(n)$ do sinal é predita como a combinação linear

$$\hat{s}(n) = -\sum_{i=1}^{p} a_i s(n-i), \tag{A.1}$$

sendo $a_1, a_2, \cdots, a_p$ os coeficientes de predição e p a ordem de predição.

Define-se o sinal de erro de predição pela subtração do sinal predito do sinal original como

$$d(n) = s(n) - \hat{s}(n). \tag{A.2}$$

Busca-se na análise LP determinar um conjunto de coeficientes de predição que minimizem o erro quadrático de predição ε. Este é o erro quadrático de predição tomado sobre um segmento de voz e aqui será indicado genericamente por

$$\varepsilon = \sum_{n} d^2(n) \tag{A.3}$$

porque há duas formas de se tomar os limites do somatório [3] que não é necessário abordar para o propósito geral deste apêndice.

Substituindo-se a Eq. (A.1) na Eq. (A.2 e o resultado duas vezes na Eq. (A.3), obtém-se

$$\varepsilon = \sum_{n} \left(s(n) + \sum_{i=1}^{p} a_i s(n-i) \right) \left(s(n) + \sum_{j=1}^{p} a_j s(n-j) \right) \tag{A.4}$$

Distribuindo-se os produtos e trocando-se a ordem dos somatórios, obtém-se a forma

$$\varepsilon = \sum_{n} s^2(n) + \sum_{i=1}^{p} a_i \sum_{n} s(n-i)s(n) + \sum_{j=1}^{p} a_j \sum_{n} s(n-j)s(n) + \sum_{i=1}^{p} \sum_{j=1}^{p} a_i a_j \sum_{n} s(n-i)s(n-j) \tag{A.5}$$

Definindo-se os coeficientes de correlação

$$\varphi_{ij} = \sum_{n} s(n-i)s(n-j), \tag{A.6}$$

eles podem ser empregados para colocar a Eq. (A.5) na forma

$$\varepsilon = \varphi_{00} + 2\sum_{i=1}^{p} a_i\varphi_{i,0} + \sum_{i=1}^{p}\sum_{j=1}^{p} a_i\varphi_{ij}a_j \tag{A.7}$$

A Eq. (A.7) é uma forma quadrática que pode ser representada por uma expressão matricial como

$$\varepsilon = \boldsymbol{\alpha}^T\boldsymbol{\Psi}\boldsymbol{\alpha}, \tag{A.8}$$

sendo $\boldsymbol{\alpha} = \begin{bmatrix} 1 & a_1 & a_2 & \cdots & a_p \end{bmatrix}^T$ o vetor de coeficientes de predição estendido e $\boldsymbol{\Psi}$ é a matriz $(p+1)\times(p+1)$ de correlação estendida. Os elementos da matriz $\boldsymbol{\Psi}$ são os coeficientes de correlação

$$\psi_{ij} = \varphi_{i-1,j-1}. \tag{A.9}$$

Assim, $\boldsymbol{\Psi}$ é uma matriz simétrica que pode ser representada em blocos, a partir da matriz $p\times p$ de correlação simétrica $\boldsymbol{\Phi}$ e do vetor de correlações $\boldsymbol{\psi}$, definidos no Apêndice B, como

$$\boldsymbol{\Psi} = \begin{bmatrix} \varphi_{00} & \boldsymbol{\psi}^T \\ \boldsymbol{\psi} & \boldsymbol{\Phi} \end{bmatrix}. \tag{A.10}$$

Finalmente, quando o vetor $\boldsymbol{\alpha}$ contém os coeficientes de predição que resolvem o sistema de equações normais (B.1) do Apêndice B, o erro quadrático na Eq. (A.8) reduz-se a

$$\varepsilon_{min} = \varphi_{00} + \boldsymbol{\psi}^T\boldsymbol{a}, \tag{A.11}$$

que é o erro quadrático mínimo.

APÊNDICE B – Predição linear em desenvolvimento matricial

A predição linear (LP) é, essencialmente, a resolução do sistema de equações normais

$$\boldsymbol{\Phi a} = -\boldsymbol{\psi} \tag{B.1}$$

em que se busca o vetor $\boldsymbol{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix}^T$ de coeficientes de predição que a satisfaça, sendo p a ordem de predição. O elemento na linha i e coluna j da matriz $\boldsymbol{\Phi}$ de correlação é o coeficiente de correlação φ_{ij} do segmento de sinal de voz $\{s(n)\}$, que pode ser calculado por meio do produto interno de polinômios [3]

$$\varphi_{ij} = \langle z^{-i}, z^{-j} \rangle \tag{B.2}$$

$$= \sum_n s(n-i)s(n-j) \tag{B.3}$$

e o vetor $\boldsymbol{\psi}$ tem componentes definidas por coeficientes de correlação da seguinte forma

$$\psi(i) = \varphi_{i,0} \tag{B.4}$$

para $i = 1, 2, \ldots, p$.

Os algoritmos eficientes de resolução do sistema de equações normais são iterativos na ordem, tomando-se ordens de predição na seqüência $m = 1, 2, \cdots, p$. O filtro inverso $A_m(z)$ de ordem m pode ser escrito em função dos polinômios ortogonais $\{B_i(z)\}_{i=1}^m$ como

$$A_m(z) = 1 + \sum_{i=1}^{m} k_i B_{i-1}(z). \tag{B.5}$$

Esta expressão polinomial pode ser colocada na seguinte forma matricial

$$\boldsymbol{a} = \boldsymbol{Bk} \tag{B.6}$$

em que $\boldsymbol{k} = \begin{bmatrix} k_1 & k_2 & \cdots & k_p \end{bmatrix}^T$ é o vetor de coeficientes de correlação parcial (PARCOR). Por sua vez, a matriz $\boldsymbol{B}$ é definida em função dos filtros inversos $\{B_i(z)\}_{i=1}^{p-1}$ da predição regressiva como

$$\boldsymbol{B} = \begin{bmatrix} 1 & b_{11} & b_{21} & \cdots & b_{p-1,1} \\ 0 & 1 & b_{12} & \cdots & b_{p-1,2} \\ 0 & 0 & 1 & \cdots & b_{p-1,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{B.7}$$

cujos coeficientes aparecem em suas colunas.

Como os polinômios $B_m(z)$ e $B_n(z)$, com $n \neq m$, são ortogonais, isto é,

$$\langle B_m(z), B_n(z) \rangle = \sum_{i=1}^{m+1} \sum_{j=1}^{n+1} b_{mi} \varphi_{ij} b_{nj} = \delta_{mn} \beta_n, \tag{B.8}$$

em que a norma β_n do polinômio $B_n(z)$ pode ser calculada pelo produto interno de polinômios

$$\beta_n = \langle B_n(z), B_n(z) \rangle \tag{B.9}$$

e $\delta_{mn} = \begin{cases} 0 & \text{se } m \neq n \\ 1 & \text{se } m = n \end{cases}$ é a função delta de Kronecker.

Em forma matricial, a Eq. (B.8) torna-se

$$\boldsymbol{B}^T \boldsymbol{\Phi} \boldsymbol{B} = \boldsymbol{\beta}, \tag{B.10}$$

sendo $\boldsymbol{\beta}$ a matriz diagonal cuja diagonal principal é formada por $\beta_0, \beta_1, \ldots, \beta_{p-1}$.

O sistema de equações normais (B.1), usando a Eq. (B.6), pode ser posto como

$$\boldsymbol{\Phi} \boldsymbol{B} \boldsymbol{k} = -\boldsymbol{\psi}. \tag{B.11}$$

Multiplicando a Eq. (B.11) membro a membro à esquerda por $\boldsymbol{B}^T$ e aplicando, então, a Eq. (B.10), vem

$$\boldsymbol{\beta} \boldsymbol{k} = -\boldsymbol{B}^T \boldsymbol{\psi}. \tag{B.12}$$

Como $\boldsymbol{\beta}$ é diagonal, admitindo-se que não seja singular, sua inversão é trivial, obtendo-se o vetor de coeficientes PARCOR como

$$\boldsymbol{k} = -\boldsymbol{\beta}^{-1} \boldsymbol{B}^T \boldsymbol{\psi} \tag{B.13}$$

e, aplicando-se a Eq. (B.6), obtém-se também os coeficientes de predição como

$$\boldsymbol{a} = -\boldsymbol{B}\boldsymbol{\beta}^{-1}\boldsymbol{B}^T\boldsymbol{\psi}. \tag{B.14}$$

Comparando-se o sistema de equações normais (B.1) original com sua resolução através da Eq. (B.14), nota-se que o processo de resolução acarretou a decomposição da matriz $\boldsymbol{\Phi}^{-1}$ como

$$\boldsymbol{\Phi}^{-1} = \boldsymbol{B}\boldsymbol{\beta}^{-1}\boldsymbol{B}^T. \tag{B.15}$$

A análise LP comumente consiste na resolução de um sistema de equações normais como a Eq. (B.1). Neste caso, após a determinação dos coeficientes de predição $a_1, a_2, \ldots, a_p$ que resolvem o sistema de equações normais de ordem p, pode-se calcular o erro quadrático mínimo resultante pela Eq. (A.11) do Apêndice A.

Uma abordagem alternativa para a análise LP consiste na absorção da equação do erro mínimo dentro do sistema de equações de análise. Para determinação do novo sistema de equações, atribui-se erro mínimo unitário e admite-se como grau de liberdade adicional a determinação do coeficiente de predição a_0, que, eventualmente, pode ser diferente da unidade. Assim, o vetor de coeficientes de predição estendido torna-se $\boldsymbol{\alpha} = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_p \end{bmatrix}^T$ e a Eq. (A.11) do erro mínimo se converte em

$$\varepsilon_{min} = a_0\varphi_{00} + \boldsymbol{\psi}^T\boldsymbol{a} = 1. \tag{B.16}$$

Isto resulta no sistema de equações LP estendido

$$\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{u} \tag{B.17}$$

em que $\boldsymbol{u} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T$ é um vetor $(p+1) \times 1$.

A resolução eficiente deste sistema de equações estendido envolve a decomposição da inversa $\boldsymbol{\Psi}^{-1}$ da matriz de correlação estendida da mesma forma que a resolução do sistema de equações normais implica na decomposição da inversa $\boldsymbol{\Phi}$ da matriz de correlação conforme a Eq. (B.15). Entretanto, neste último caso, a decomposição da matriz inversa é implícita enquanto no caso do sistema estendido é necessário obter a matriz inversa, principalmente quando ele se apresenta na forma

$$\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{b} \tag{B.18}$$

em que $\boldsymbol{b}$ pode ser um vetor qualquer.

APÊNDICE C – Média do espectro logarítmico de potência

A média do espectro logarítmico de potência $\log P(\omega)$ é

$$\overline{P_L} = \frac{1}{\pi}\int_0^\pi \log P(\omega) d\omega. \tag{C.1}$$

Em especial, o modelo LP

$$\begin{aligned} \hat{P}(\omega) &= \left|H\left(e^{j\omega}\right)\right|^2 = \frac{1}{\left|A\left(e^{j\omega}\right)\right|^2} && \text{(C.2)} \\ &= \frac{1}{\left|\sum_{i=0}^{p} a_i e^{-j\omega i}\right|^2} && \text{(C.3)} \end{aligned}$$

tem média logarítmica

$$\overline{\hat{P}_L} = -\frac{1}{\pi}\int_0^\pi \log\left|A\left(e^{j\omega}\right)\right|^2 d\omega. \tag{C.4}$$

Tomando-se o espectro bilateral, tem-se a média espectral logarítmica

$$\begin{aligned} \overline{\hat{P}_L} &= -\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|A\left(e^{j\omega}\right)\right|^2 d\omega && \text{(C.5)} \\ &= -\frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\log A\left(e^{j\omega}\right) + \log A\left(e^{-j\omega}\right)\right) d\omega && \text{(C.6)} \\ &= -\frac{1}{2\pi}\int_{-\pi}^{\pi} 2\Re\left[\log A\left(e^{-j\omega}\right)\right] d\omega && \text{(C.7)} \\ &= -\frac{2}{2\pi}\Re\left[\int_{-\pi}^{\pi} \log A\left(e^{-j\omega}\right) d\omega\right] && \text{(C.8)} \end{aligned}$$

a partir da qual a extensão analítica da função no plano complexo z permite escrever

$$\overline{\hat{P}_L} = -\frac{2}{2\pi j}\Re\left[\oint_C \log A\left(1/z\right)\frac{dz}{z}\right] \tag{C.9}$$

em que C é a circunferência de raio unitário com centro na origem do plano z percorrida no sentido anti-horário.

Como conseqüência de sua obtenção através de análise LP, $1/A(z)$ é estável e, assim,

os zeros de $A(z)$ estão no interior do círculo unitário de forma que os zeros de $A(1/z)$ estarão no exterior do círculo unitário. Por outro lado, no interior do círculo unitário, a única singularidade de $\frac{\log A(1/z)}{z}$ está em $z = 0$ e, aplicando o teorema dos resíduos [36], pode-se calcular seu resíduo como

$$\begin{aligned} \operatorname{Res}[z=0] &= \lim_{z\to 0} \log A\left(1/z\right) \frac{z-0}{z} && \text{(C.10)} \\ &= \log A\left(\infty\right) && \text{(C.11)} \\ &= \log a_0. && \text{(C.12)} \end{aligned}$$

A igualdade (C.12) permite escrever a média log-espectral (C.9) como

$$\overline{\hat{P}_L} = -2 \log a_0. \tag{C.13}$$

No caso em que o modelo LP está representado como

$$\hat{P}(\omega) = \frac{G^2}{\left|1 + \sum_{i=1}^{p} a_i e^{-j\omega i}\right|^2}, \tag{C.14}$$

sua média log-espectral (C.13) representa-se como

$$\overline{\hat{P}_L} = 2 \log G \tag{C.15}$$

e a média da densidade espectral logarítmica de potência do filtro inverso é

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left|A\left(e^{j\omega}\right)\right|^2 d\omega &= 2 \log 1 && \text{(C.16)} \\ &= 0. && \text{(C.17)} \end{aligned}$$

Este resultado de média nula da densidade espectral logarítmica de potência do filtro inverso é muito importante na análise espectral da predição linear, tratada extensamente no Cap. 6 de [3].

APÊNDICE D – O gradiente da distorção espectral logarítmica

Deseja-se ajustar a um espectro harmônico discreto um modelo espectral contínuo (3.25) amostrado nas freqüências $\omega = \omega_k = k\omega_1$, onde $\omega_1 = 2\pi/l_0$.

Quando se emprega a medida de distorção espectral (SD) logarítmica

$$E_{\mathrm{SD}} = \frac{1}{N}\sum_{k=1}^{N} \log^2 \frac{P\left(\omega_k\right)}{\hat{P}\left(\omega_k\right)} \tag{D.1}$$

diretamente num processo de modelagem espectral, é necessário obter-se seu gradiente para prosseguir com um processo de minimização.

A partir da Eq. (D.1), cada derivada parcial pode ser expressa como

$$\frac{\partial}{\partial a_l} E_{\mathrm{SD}} = \frac{2}{N}\sum_{k=1}^{N} \log\left(\frac{P\left(\omega_k\right)}{\hat{P}\left(\omega_k\right)}\right) \frac{\hat{P}\left(\omega_k\right)}{P\left(\omega_k\right)} P\left(\omega_k\right) \frac{\partial}{\partial a_l} \frac{1}{\hat{P}\left(\omega_k\right)} \tag{D.2}$$

pela aplicação sucessiva da regra da cadeia. Esta equação pode ser imediatamente simplificada para

$$\frac{\partial}{\partial a_l} E_{\mathrm{SD}} = \frac{2}{N}\sum_{k=1}^{N} \log\left(\frac{P\left(\omega_k\right)}{\hat{P}\left(\omega_k\right)}\right) \hat{P}\left(\omega_k\right) \frac{\partial}{\partial a_l} \frac{1}{\hat{P}\left(\omega_k\right)}, \tag{D.3}$$

sendo que a última derivada parcial no segundo membro pode ser calculada pela substi-

tuição do modelo espectral da Eq. (3.25), resultando

$$\frac{\partial}{\partial a_l}\frac{1}{\hat{P}\left(\omega_k\right)} = \frac{\partial}{\partial a_l}\left|\sum_{i=0}^{p} a_i e^{-j\omega_k i}\right|^2 \tag{D.4}$$

$$= \frac{\partial}{\partial a_l}\sum_{i=0}^{p} a_i e^{j\omega_k i}\sum_{m=0}^{p} a_m e^{-j\omega_k m} \tag{D.5}$$

$$= \frac{\partial}{\partial a_l}\sum_{i=0}^{p}\sum_{m=0}^{p} a_i a_m e^{j\omega_k (i-m)} \tag{D.6}$$

$$= \sum_{\substack{m=0\\ m\neq l}}^{p} a_m e^{j\omega_k (l-m)} + \sum_{\substack{i=0\\ i\neq l}}^{p} a_i e^{j\omega_k (i-l)} + 2\sum_{i=0}^{p} a_i \tag{D.7}$$

$$= \sum_{i=0}^{p} a_i e^{-j\omega_k (i-l)} + \sum_{i=0}^{p} a_i e^{j\omega_k (i-l)} \tag{D.8}$$

$$= 2\sum_{i=0}^{p} a_i \cos\left(\omega_k (i-l)\right). \tag{D.9}$$

Finalmente, substituindo esta última equação na Eq. (D.3), última versão da derivada parcial de E_{SD} acima, e rearranjando somatórios, obtém-se

$$\frac{\partial}{\partial a_l}E_{\mathrm{SD}} = \frac{4}{N}\sum_{i=0}^{p} a_i \sum_{k=1}^{N} \log\left(\frac{P\left(\omega_k\right)}{\hat{P}\left(\omega_k\right)}\right)\hat{P}\left(\omega_k\right)\cos\left(w_k (i-l)\right). \tag{D.10}$$

Referências Bibliográficas

[1] ITAKURA, F.; SAITO, S. Analysis-synthesis telephony based on the maximum likelihood method. In: KOHASI, Y. (Ed.). *Proc. 6th International Congress Acoustics*. Tokyo: Maruzen, Elsevier, 1968. p. C–17–C–20. Paper C-5-5.

[2] ITAKURA, F.; SAITO, S. Analysis-synthesis telephony based on the maximum likelihood method. In: FLANAGAN, J. L.; RABINER, L. R. (Ed.). *Speech Synthesis*. Stroudsburg, PA: Dowden, Hutchinson and Ross, 1973. cap. IV: Predictive coding of speech, p. 289–292. Distributor: John Wiley & Sons.

[3] MARKEL, J. D.; GRAY,Jr., A. H. *Linear Prediction of Speech*. Berlin: Springer, 1976.

[4] McAULAY, R. J.; QUATIERI, T. F. Sinusoidal coding. In: BASTIAAN KLEIJN, W.; PALIWAL, K. K. (Ed.). *Speech Coding and Synthesis*. Amsterdam: Elsevier Science, 1995. p. 121–173.

[5] ARJONA RAMÍREZ, M.; MINAMI, M. Low bit rate speech coding. In: PROAKIS, J. G. (Ed.). *The Wiley Encyclopedia of Telecommunications*. New York: Wiley, 2003. v. 3, p. 1299–1308.

[6] ARJONA RAMÍREZ, M.; MINAMI, M. Technology and standards for low-bit-rate vocoding methods. In: BIDGOLI, H. (Ed.). *The Handbook of Computer Networks*. New York: Wiley, 2006. II. A ser publicado.

[7] KOENIG, W.; DUNN, H. K.; LACEY, L. Y. The sound spectrograph. *J. Acoust. Soc. Am.*, v. 18, p. 19–49, 1946.

[8] BASTIAAN KLEIJN, W.; HAAGEN, J. Waveform interpolation for coding and synthesis. In: BASTIAAN KLEIJN, W.; PALIWAL, K. K. (Ed.). *Speech Coding and Synthesis*. Amsterdam: Elsevier Science, 1995. p. 175–207.

[9] ARJONA RAMÍREZ, M.; MINAMI, M. Split-order linear prediction for segmentation and harmonic spectral modeling. *IEEE Signal Processing Lett.*, v. 13, n. 4, Apr. 2006. A ser publicado.

[10] ARJONA RAMÍREZ, M. Cycle extraction for perfect reconstruction and rate scalability. In: *Proc. of EUROSPEECH, European Conference on Speech Communication and Technology*. Geneva: [s.n.], 2003. v. 4, p. 2945–2948.

[11] ARJONA RAMÍREZ, M. A waveform extractor for scalable speech coding. In: *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing*. Hong Kong: [s.n.], 2003. v. 2, p. 169–172.

[12] ARJONA RAMÍREZ, M.; BEZERRA de MELO, A. Excitation models for speech interpolation coding. In: *Proc. of the International Workshop on Telecommunications*. Santa Rita do Sapucaí: [s.n.], 2004. p. 238–241.

[13] RUOPPILA, V. T.; TAMMI, M.; SAARINEN, J. Waveform extraction for perfect reconstruction in WI coding. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Istanbul: [s.n.], 2000. v. 3, p. 1359–1362.

[14] YANG, H. et al. Pitch synchronous modulated lapped transform of the linear prediction residual of speech. In: *Proc. of IEEE Int. Conf. on Signal Processing*. Beijing: [s.n.], 1998. v. 1, p. 591–594.

[15] BASTIAAN KLEIJN, W.; HAAGEN, J. A speech coder based on decomposition of characteristic waveforms. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Detroit: [s.n.], 1995. v. 1, p. 508–511.

[16] KLEIJN, W. B. et al. A low-complexity waveform interpolation coder. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Atlanta: [s.n.], 1996. v. 1, p. 212–215.

[17] CHONG, N. R.; BURNETT, I. S.; CHICHARO, J. F. Adapting waveform interpolation (with pitch-spaced subbands) for quantisation. In: *Proc. IEEE Workshop on Speech Coding*. Porvoo: [s.n.], 1999. p. 96–98.

[18] CHONG-WHITE, N. R.; BURNETT, I. S. Accurate, critically sampled characteristic waveform surface construction for waveform interpolation decomposition. *IEE Electronics Letters*, v. 36, n. 14, p. 1245–1247, Jul. 2000.

[19] KLEIJN, W. B. et al. A 5.85 kbit/s CELP algorithm for cellular applications. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Minneapolis: [s.n.], 1993. v. 2, p. 596–599.

[20] MAKHOUL, J. Spectral linear prediction: Properties and applications. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-23, n. 3, p. 283–296, June 1975.

[21] EL-JAROUDI, A.; MAKHOUL, J. Discrete all-pole modeling. *IEEE Trans. Signal Processing*, v. 39, n. 2, p. 411–423, Feb. 1991.

[22] ITAKURA, F. et al. An audio response unit based on partial autocorrelation. *IEEE Trans. Commun.*, COM-20, n. 4, p. 792–797, Aug. 1972.

[23] JAYANT, N. S.; NOLL, P. *Digital coding of waveforms*. Englewood Cliffs: Prentice-Hall, 1984.

[24] PALIWAL, K. K.; ATAL, B. S. Efficient vector quantization of LPC parameters at 24 bits/frame. *IEEE Trans. Speech Audio Processing*, v. 1, n. 1, p. 3–14, Jan. 1993.

[25] FURUI, S. *Digital speech processing, synthesis, and recognition*. New York: Marcel Dekker, 1985.

[26] CHAMPION, T. G.; McAULAY, R. G.; QUATIERI, T. F. High-order allpole modelling of the spectral envelope. In: *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*. Adelaide: [s.n.], 1994. v. 1, p. 529–532.

[27] WEI, B.; GIBSON, J. D. A new discrete spectral modeling method and an application to CELP coding. *IEEE Signal Processing Lett.*, v. 10, n. 4, p. 101–103, Apr. 2003.

[28] GARDNER, W. R.; RAO, B. D. Theoretical analysis of the high-rate vector quantization of LPC parameters. *IEEE Trans. Speech Audio Processing*, v. 3, n. 5, p. 367–381, Sept. 1995.

[29] MARKEL, J. D.; Gray,Jr., A. H. Distance measures for speech processing. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-24, n. 5, p. 380–391, Oct. 1976.

[30] ERTAN, A. E.; Barnwell III, T. P. Circular LPC modeling and constant pitch transformation for scalable bit rate, scalable quality speech coding. In: *2002 IEEE Speech Coding Workshop Proc.* Tsukuba: [s.n.], 2002. p. 50–52.

[31] KLEIJN, W. B.; HAAGEN, J. Waveform interpolation for coding and synthesis. In: KLEIJN, W. B.; PALIWAL, K. K. (Ed.). *Speech Coding and Synthesis*. Amsterdam: Elsevier Science, 1995. p. 175–207.

[32] CERNUSCHI-FRIAS, B. A derivation of the Gohberg-Semencul relation. *IEEE Trans. Signal Processing*, v. 39, n. 1, p. 190–192, Jan. 1991.

[33] RAMABADRAN, T.; SMITH, A.; JASIUK, M. An iterative interpolative transform method for modeling harmonic magnitudes. In: *2002 IEEE Speech Coding Workshop Proc.* Tsukuba: [s.n.], 2002. p. 38–40.

[34] BASTIAAN KLEIJN, W. Signal processing representations of speech. *IEICE Trans. Inf. & Syst.*, E86-D, n. 3, p. 359–376, March 2003.

[35] MOLYNEUX, D. J.; HO, M. S.; CHEETHAM, B. M. G. Robust application of all-pole modeling to sinusoidal transform coding. In: *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing*. Istanbul: [s.n.], 2000. v. 3, p. 1455–1458.

[36] LEPAGE, W. R. *Complex Variables and the Laplace Transform for Engineers*. New York: Dover, 1980.