UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA

APLICAÇÃO DO PROCESSAMENTO HOMOMÓRFICO NA CODIFICAÇÃO DE VOZ A BAIXAS TAXAS

POR : PAULO CESAR DANTAS OLIVEIRA
Engenheiro Eletricista (UFBA - 1988)

ORIENTADOR : PROF. DR. AMAURI LOPES
Professor MS4 do Departamento de Comunicações da
Faculdade de Engenharia Elétrica da UNICAMP

CO - ORIENTADOR : PROF. DR. FÁBIO VIOLARO
Professor MS5 do Departamento de Comunicações da
Faculdade de Engenharia Elétrica da UNICAMP

Dissertação apresentada à Faculdade de Engenharia Elétrica da UNICAMP como requisito parcial para a obtenção do título de Mestre em Engenharia Elétrica.

CAMPINAS, Março de 1992

Este trabalho contou com o apoio financeiro do CNPQ e do CPqD/TELEBRAS, através do convênio UNICAMP/TELEBRAS 387/90.

Agradeço

ao Professor Doutor João Baptista Tadanobu Yabu-uti do Departamento de Comunicações da FEE/UNICAMP, pelo fundamental apoio dedicado na fase inicial do meu curso de Pós-Graduação.

à senhora Iêda Eberlin e senhorita Márcia, pela cooperação na área de infra-estrutura administrativa.

à senhora Lúcia Cardoso, pela elaboração de alguns dos desenhos que ilustram este trabalho.

aos Engenheiros Francisco Egashira, Ernesto Antunes e Leonardo Resende, pelo companheirismo durante o dia a dia da elaboração deste trabalho, além da cooperação na realização dos testes subjetivos.

ao Engenheiro Adrian Batista, pela troca de informações e conhecimentos essenciais na elaboração de parte deste trabalho.

ao Professor José Augusto Fernandes Afonso do Departamento de Comunicações da FEE/UNICAMP, pela cessão do software utilizado na elaboração dos gráficos que ilustram este trabalho.

de maneira especial, aos Professores Doutores Amauri Lopes e Fábio Violaro pela orientação segura e amiga, além do apoio e confiança, sem os quais este trabalho não seria realizado.

A meus pais, Alberto e Therezinha.

Para Tereza.

RESUMO

Este trabalho traz uma análise da aplicação da técnica da Deconvolução Homomórfica na Codificação de Voz a Baixas Taxas. A partir desta técnica é possível obter o cepstrum complexo da resposta impulsiva do filtro digital representativo dos efeitos combinados do Pulso Glótico, do Trato Vocal e da Impedância de Irradiação, segundo o modelo tradicional de produção de sinais de voz. A transmissão de algumas amostras do cepstrum complexo permite a realização de uma estimativa da resposta impulsiva do filtro, a qual, ao ser convoluída com um sinal de excitação adequado, permite reconstruir o sinal de voz no receptor.

Com base na análise anterior, são realizadas simulações de Sistemas Homomórficos de Codificação de Voz, operando a taxas em torno de 4,8 e 9.0 kbits/s. O desempenho destes sistemas é avaliado através de testes subjetivos e comparado ao desempenho de um Sistema de Codificação de Voz baseado na Análise LPC convencional.

Este trabalho traz também um estudo sobre a técnica da Predição Homomórfica que combina a Deconvolução Homomórfica com a Análise Preditiva Linear. Esta técnica possibilita a redução da taxa de transmissão em Sistemas Homomórficos de Fase Mista , além da oportunidade de avaliação dos efeitos da Análise LPC quando aplicada diretamente sobre a resposta impulsiva do Trato Vocal.

Página

CAPÍTULO 1

INTRODUÇÃO

Nas últimas décadas, a informação tem se tornado um elemento de fundamental importância no funcionamento e desenvolvimento do sistema sócio-ecônomico em que está inserido o homem moderno. A crescente quantidade de informação a ser transmitida desafiou os pesquisadores no sentido de se obter sistemas de alta capacidade que possibilitem o processamento e a transmissão de informações de maneira rápida e pouco sensível a erros. A digitalização dos sistemas de comunicação é uma resposta a este desafio e representa, atualmente, o objetivo final a ser alcançado no menor espaço de tempo possível.

A transmissão de informação através de sinais de voz ocupa um lugar de grande importância neste contexto, já que o sistema telefônico representa a maior rede de comunicação existente no planeta, interligando praticamente todos os lugares do mundo onde existem sociedades organizadas.

No sistema PCM convencional de transmissão digital, o sinal de voz é amostrado a uma frequência de 8 kHz e codificado em palavras de oito bits, o que corresponde a uma taxa de transmissão de 64 kbits/s. Este sistema pertence ao grupo dos "*Codificadores de Forma de Onda*" e seu desempenho diminue rapidamente quando operando a taxas inferiores a 16 kbits/s [1].

Neste contexto, o fato de que o aumento da capacidade dos canais disponíveis não acompanha a velocidade de crescimento da quantidade de informação a ser transmitida, vem estimulando os grandes avanços que estão sendo obtidos com os sistemas do tipo "*Codificadores de Voz a Baixas Taxas*". Nesta categoria estão reunidos os sistemas que operam a taxas abaixo de 9.6 kbits/s [1]. Estes sistemas são genericamente denominados "*Vocoders*" e utilizam uma representação paramétrica do sinal de voz. Esta representação é baseada num modelo de produção de voz no qual um

sistema digital linear H(z), com variação lenta no tempo, representando as contribuições combinadas do *Pulso Glótico*, do *Trato Vocal* e da *Impedância de Irradiação*, é excitado convenientemente para se produzir o sinal de voz [2]. Este sistema linear é considerado racional em um curto intervalo de tempo e assim, pode ser caracterizado, por exemplo, por um número finito de pólos e zeros, ou equivalentemente, pelos coeficientes do numerador e denominador, os quais constituem os parâmetros do modelo. Dessa maneira, pode-se, em alguns casos, levar a taxa de transmissão a níveis em torno de 2 kbits/s, à custa de uma degradação na qualidade do sinal de voz .

O modelo de produção de voz mais tradicional utiliza como sinais de excitação um trem de impulsos unitários periódico para a geração de sons sonoros e ruído branco para a geração de sons não-sonoros[2]. A figura 1.1 ilustra este modelo:

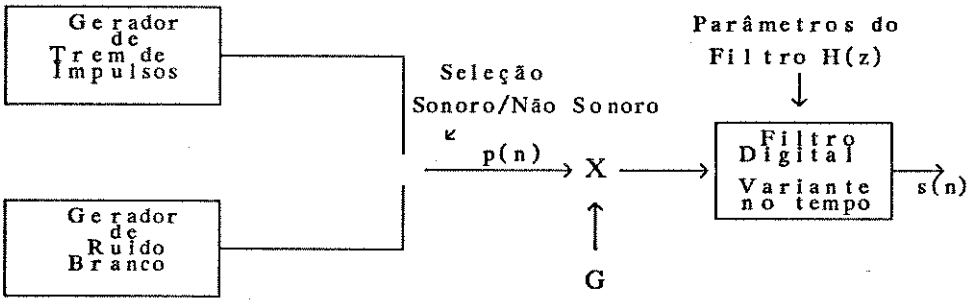


Figura 1.1 – Modelo Tradicional de Producão de Voz

onde p(n) é o sinal de excitação, G é uma informação de amplitude e s(n) é o sinal de voz sintetizado.

A técnica mais difundida para a determinação dos parâmetros característicos do sistema H(z) utiliza o método da *Análise Preditiva Linear* e é conhecida como *Análise LPC (Linear Predictive Coding)* [2]. Supondo o filtro H(z) constituído apenas por pólos, a Análise LPC determina de forma bastante eficiente os coeficientes do filtro H(z). A Análise LPC é aplicada diretamente ao sinal de voz e sua função básica é realizar uma modelagem *AR (Auto-Regressive)* [2] do envelope do espectro de voz, não sendo sensível às variações finas causadas pela excitação. O sistema baseado nesta técnica é conhecido como *Vocoder LPC* [2].

Os resultados bastante satisfatórios obtidos com a Análise LPC, aliados a

sua simplicidade e eficiência, foram os responsáveis pela grande difusão desta técnica, a qual é utilizada na maioria dos sistemas de codificação de voz em baixas taxas em desenvolvimento. Entretanto, a qualidade da voz sintetizada não se aproxima daquela obtida com os codificadores de forma de onda , como o PCM convencional, apresentando principalmente uma perda de naturalidade no sinal de voz produzido.

Apesar da técnica da Análise LPC ser limitada no sentido de que impõe a modelagem do filtro H(z) apenas com pólos, a sua simplicidade e eficiência ainda recomendam fortemente o seu emprego. Com isto, os esforços para a obtenção de um melhor desempenho dos codificadores paramétricos de voz tem sido canalizados no sentido do aperfeiçoamento do sinal de excitação, afastando-se da configuração proposta no modelo tradicional de produção de voz . Dessa maneira, as pesquisas atualmente apontam para os sistemas *Multi- Pulse LPC*, *CELP (Code Exciled Linear Predictive Coding)* e *RELP (Residual Excited Linear Predictive Coding)* [1]. Estes sistemas são baseados na Análise LPC, porém utilizam sinais de excitação mais elaborados para melhorar a qualidade da voz sintetizada. Evidentemente, estes sistemas conseguem manter a simplicidade da Análise LPC às custas de um considerável aumento da complexidade na geração do sinal de excitação.

Este trabalho aponta para o sentido oposto e, apostando na simplicidade do modelo tradicional, utiliza um método mais elaborado para se determinar um conjunto de parâmetros característicos do filtro H(z). O objetivo é explorar ao máximo o potencial do modelo tradicional na reprodução da voz, buscando determinar o limite máximo de desempenho que pode ser atingido.

Neste sentido, é utilizada a técnica da *Desconvolução Homomórfica* [3], que é uma alternativa à Análise LPC na obtenção de um conjunto de parâmetros característicos do filtro H(z). Com base nesta técnica, é possível gerar, a partir do sinal de voz, uma sequência $\hat{h}(n)$ denominada *cepstrum complexo da resposta impulsiva* h(n), que caracteriza o filtro H(z). Como as amostras da sequência $\hat{h}(n)$ diminuem de amplitude rapidamente, é possível obter um pequeno conjunto de amostras suficiente para a realização de uma boa estimativa da resposta impulsiva do filtro H(z), a qual é utilizada na geração do sinal de voz a partir de sua convolução com o sinal de excitação.

Apesar de sua maior complexidade, a Desconvolução Homomórfica não faz nenhuma restrição quanto à forma do filtro H(z), como realizado na Análise LPC, e assim um desempenho superior ao do Vocoder LPC pode ser obtido, mesmo com a utilização do modelo tradicional de produção de voz.

O estudo da utilização da técnica da Desconvolução Homomórfica na

codificação de voz a baixas taxas, proporciona também conhecimentos valiosos que podem ser aplicados por exemplo, em Sistemas de Sonar e Radar , Prospecção de Petróleo, Equalização Cega, entre outros[3].

ORGANIZAÇÃO DA DISSERTAÇÃO

No Capítulo 2 é apresentada a técnica do Processamento Homomórfico de Sinais, com particular ênfase aos Sistemas Homomórficos Convolucionais.

No Capítulo 3 são analisados os aspectos específicos do Processamento Homomórfico quando aplicado à Análise-Síntese de sinais de voz.

O Capítulo 4 traz uma descrição completa de cada uma das operações que compõem um Sistema Homomórfico de codificação de voz. Este capítulo traz também um relato do desempenho deste sistema em testes subjetivos informais e uma comparação de seu desempenho em relação ao Vocoder LPC tradicional.

O Capítulo 5 analisa a técnica da Predição Homomórfica que reúne a Análise LPC com a Desconvolução Homomórfica. Esta técnica é apresentada devido ao seu potencial de aplicação na modelagem *ARMA (Auto-Regressive Moving Average)* de sinais de voz e na redução da taxa de transmissão do Vocoder Homomórfico.

Finalmente o Capítulo 6 traz uma análise geral dos resultados obtidos e descreve as principais conclusões deste trabalho.

REFERÊNCIAS

[1] B. S. Atal e L. R. Rabiner, *"Speech Research Directions"*, AT&T Techinical Journal, September/October 1986, Vol.65

[2] L. R. Rabiner and R. W. Schafer, *"Digital Processing of Speech Signals"*, Englewood Cliffs, NJ, Prentice-Hall, 1978.

[3] A. V. Oppenheim and R. W. Schafer, *"Digital Signal Processing"*. Englewood Cliffs, NJ,Prentice - Hall, 1975.

CAPÍTULO 2

PROCESSAMENTO HOMOMÓRFICO DE SINAIS

CONTEÚDO

2.1 INTRODUÇÃO

Neste capítulo são detalhados os princípios gerais do Processamento Homomórfico de Sinais. Em particular, são analisados os Sistemas Homomórficos Convolucionais, os quais se aplicam ao processamento de sinais de voz. São também abordados alguns aspectos computacionais de relevância na implementação destes sistemas.

2.2 Princípio da Superposição Generalizado

Seja **T** uma transformação que caracteriza um sistema genérico linear invariante com o deslocamento . Sejam também $x_1(n)$ e $x_2(n)$ duas entradas quaisquer e **C** um escalar. O Princípio da Superposição estabelece que :

$$T [x_1(n) + x_2(n)] = T [x_1(n)] + T [x_2(n)] \quad (2.1)$$

$$T [C\, x_1(n)] = C \,.\, T [x_1(n)] \quad (2.2)$$

Para realizar a generalização do Princípio da Superposição, adota-se o símbolo "□" para representar combinações de entradas e o símbolo ":" para representar combinações de entradas com escalares. Da mesma forma, denota-se as regras para combinações de saídas pelo símbolo "o" e as regras para combinações de saídas com escalares pelo símbolo "⌊". Sendo **H** a transformação que caracteriza o sistema, o Princípio da Superposição Generalizado estabelece que :

$$H [x_1(n) \,\square\, x_2(n)] = H [x_1(n)] \,o\, H [x_2(n)] \quad (2.3)$$

$$H [C : x_1(n)] = C \,\lfloor\, H [x_1(n)] \quad (2.4)$$

Os sistemas lineares constituem uma particularização, na qual tem-se :

□ = o = + (adição) e
: = ⌊ = . (multiplicação)

Os sistemas que obedecem ao Princípio da Superposição Generalizado podem ser representados como na figura 2.1 e são denominados *Sistemas Homomórficos* [1].

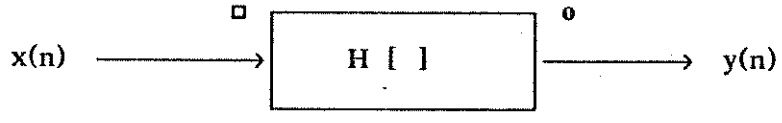


Fig 2.1 - O Sistema Homomórfico

Nesta representação x(n) e y(n) constituem os sinais de entrada e saída, respectivamente.

Demonstra-se [1] que um Sistema Homomórfico pode ser representado como uma cascata de 3 (três) sistemas, como mostrado na figura 2.2:

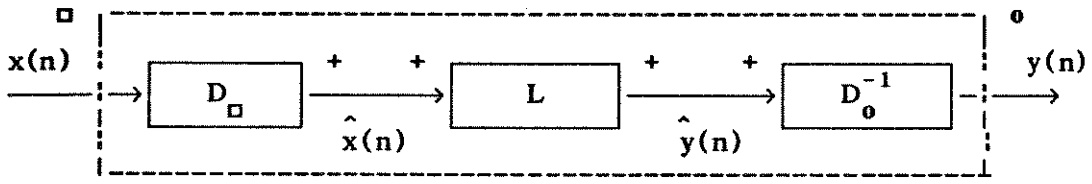


Figura 2.2 - Representação Canônica de Sistemas Homomórficos

Esta representação é denominada *Representação Canônica de Sistemas Homomórficos* [1].

O primeiro sistema, $D_{\square}$, atende às seguintes propriedades:

$$D_{\square}\,[\,x_1(n) \,\square\, x_2(n)\,] = D_{\square}[x_1(n)] + D_{\square}[x_2(n)] \qquad (2.5)$$

$$= \hat{x}_1(n) + \hat{x}_2(n)$$

$$D_{\square}[C : x_1(n)] = C \,.\, D_{\square}[x_1(n)] \qquad (2.6)$$

$$= C \,.\, \hat{x}_1(n)$$

Ou seja, $D_{\square}$ obedece ao Princípio da Superposição Generalizado onde a operação entre as entradas é do tipo "□" e a operação entre as saídas é a adição. O efeito do sistema $D_{\square}$ é a transformação da combinação dos sinais $x_1(n)$ e $x_2(n)$ de acordo com a regra "□" em uma combinação linear convencional dos sinais correspondentes, $D_{\square}[x_1(n)]$ e $D_{\square}[x_2(n)]$.

O sistema L é uma sistema linear convencional que obedece às seguintes equações:

$$L\,[\,\hat{x}_1(n) + \hat{x}_2(n)\,] = L\,[\hat{x}_1(n)] + L\,[\hat{x}_2(n)] \qquad (2.7)$$

$$= \hat{y}_1(n) + \hat{y}_2(n)$$

$$L\,[\,C\,\hat{x}_1(n)\,] = C \,.\, L\,[\hat{x}_1(n)] \qquad (2.8)$$

$$= C \,.\, \hat{y}_1(n)$$

Finalmente, o sistema D_0^{-1} realiza a transformação da operação de adição para a operação representada pelo símbolo "o". Tem-se então:

$$D_0^{-1}\,[\,\hat{y}_1(n) + \hat{y}_2(n)\,] = D_0^{-1}[\hat{y}_1(n)] \text{ o } D_0^{-1}[\hat{y}_2(n)] \qquad (2.9)$$

$$= y_1(n) \text{ o } y_2(n)$$

$$D_0^{-1}[C\,\hat{y}_1(n)] = C \rfloor D_0^{-1}[\hat{y}_1(n)] \qquad (2.10)$$

$$= C \rfloor y_1(n)$$

Desde que o sistema $D_\square$ é completamente determinado pelas operações "□" e ":", ele é denominado *Sistema Característico* para a operação "□" . Da mesma forma,conclui-se que todos os Sistemas Homomórficos com mesmas operações de entrada e saída diferem apenas na parte linear. Este resultado é de fundamental importância, pois uma vez determinadas as características do sistema relativas às operações de entrada e saída , o problema a ser resolvido torna-se puramente linear.

2.3 Sistemas Homomórficos Convolucionais

Existe uma variedade de problemas de processamento de sinais onde as entradas do sistema encontram-se combinadas por convolução. É de particular interesse o processamento do sinal de voz, onde frequentemente se deseja realizar a separação dos efeitos da resposta impulsiva do trato vocal e da excitação. A convolução destes dois sinais produz o sinal de voz [2,3].

2.3.1 O SISTEMA CANÔNICO

A forma canônica para os Sistemas Homomórficos Convolucionais é ilustrada na figura 2.3 :

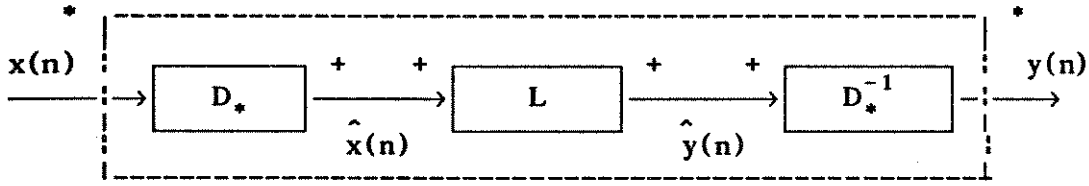


Figura 2.3 - Forma Canônica dos Sistemas Homomórficos Convolucionais

O sistema característico D_* obedece às seguintes propriedades :

$$D_* [x_1(n) * x_2(n)] = D_*[x_1(n)] + D_*[x_2(n)] \qquad (2.11)$$

$$= \hat{x}_1(n) + \hat{x}_2(n)$$

$$D_*[C \cdot x_1(n)] = C \cdot D_*[x_1(n)] \qquad (2.12)$$

$$= C \cdot \hat{x}_1(n)$$

O sistema **L** é um sistema linear convencional e D_*^{-1} é o sistema inverso ao sistema característico D_*.

2.3.2 REPRESENTAÇÃO MATEMÁTICA DO SISTEMA CARACTERÍSTICO D_*

Seja :

$$y(n) = x_1(n) * x_2(n)$$

Da teoria de processamento de sinais sabe-se que:

$$Y(z) = X_1(z) \cdot X_2(z) \qquad (2.13)$$

Esta propriedade serve de base para a representação matemática do sistema característico D_*. Tem-se então:

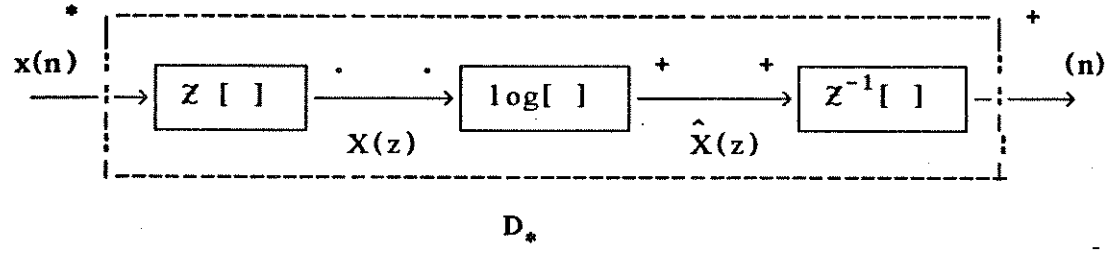


Figura 2.4 - Sistema Característico D_*

Como o sinal X(z) é normalmente complexo, deve ser empregado o logaritmo complexo na definição do sistema, ou seja :

$$\log [\ X(z)\ ] = \hat{X}(z) = \ln|X(z)| + j\ \arg[X(z)] \tag{2.14}$$

O sinal $\hat{X}(z)$ deve representar uma transformada $\mathcal{Z}$ válida e $\hat{x}(n)$ deve ser unicamente definido.

Assumindo que tanto x(n) quanto $\hat{x}(n)$ são sequências absolutamente somáveis, as regiões de convergência de X(z) e $\hat{X}(z)$ devem incluir a circunferência de raio unitário.

Se $\hat{X}(z)$ representa uma transformada $\mathcal{Z}$ válida, então ela pode ser expandida em série de Laurent:

$$\hat{X}(z) = \log[X(z)] = \sum_{n=-\infty}^{\infty} (n)z^{-n} \tag{2.15}$$

A região de convergência desta série deve incluir a circunferência de raio unitário, ou seja, $\hat{X}(z)$ deve ser analítica na região que inclui a circunferência de raio unitário. Expressando $\hat{X}(z)$ nesta circunferência tem-se:

$$\hat{X}(e^{j\omega}) = \hat{X}_R(e^{j\omega}) + j\ \hat{X}_I(e^{j\omega}) \tag{2.16}$$

Para que x(n) e $\hat{x}(n)$ sejam sequências reais, $\hat{X}_R(e^{j\omega})$ deve ser uma função par de ω e $\hat{X}_I(e^{j\omega})$ deve ser uma função ímpar de ω. Além disso $\hat{X}(e^{j\omega})$ deve ser uma função periódica em ω com período 2π.

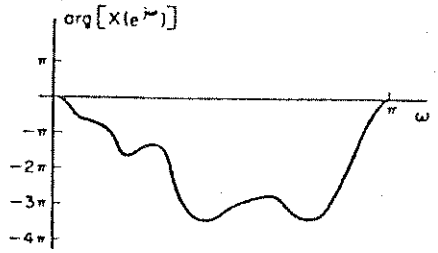
A analiticidade de $\hat{X}(z)$ na circunferência de raio unitário implica que $\hat{X}(e^{j\omega})$ deve ser uma função contínua de ω. Lembrando a equação (2.14), pode-se escrever :

$$\hat{X}_R(e^{j\omega}) = \ln |X(e^{j\omega})| \tag{2.17}$$

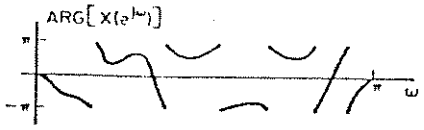
$$\hat{X}_I(e^{j\omega}) = \arg [X(e^{j\omega})] \tag{2.18}$$

Levando-se em consideração as afirmações anteriores, conclui-se que as funções definidas pelas equações (2.17) e (2.18) devem ser contínuas em ω. Supondo que X(z) não possui zeros sôbre a circunferência de raio unitário, a continuidade de $\hat{X}_R(e^{j\omega})$ é garantida pelo fato de que X(z) é analítica na circunferência de raio unitário. A continuidade de $\hat{X}_I(e^{j\omega})$ porém, é dependente da definição do logaritmo complexo. Dessa forma, o problema da validade da transformada $\mathcal{Z}$, $\hat{X}(z)$, está intimamente relacionado com uma definição não ambígua para o logaritmo complexo expresso na equação (2.14).

O problema da unicidade do logaritmo complexo é ilustrado pela figura 2.5 [1]:



a)



b)

Figura 2.5

Na figura 2.5.a tem-se uma curva de fase típica para uma transformada $\mathcal{Z}$, X(z), avaliada na circunferência de raio unitário. Se X(z) representa o produto de duas transformadas $\mathcal{Z}$, $X_1(z)$ e $X_2(z)$, então esta curva pode ser vista como sendo o resultado da soma das duas curvas de fase contínuas de $X_1(z)$ e $X_2(z)$. A figura 2.5.b mostra o valor principal da fase de X(z). As duas curvas da figura 2.5 são

representações válidas da fase de X(z), desde que :

$$e^{j\arg[X(z)]} = e^{j\,ARG[X(z)]} \tag{2.19}$$

Contudo, pode-se ver que a curva do valor principal da fase não pode, em geral, corresponder à soma das duas curvas de fase principal das transformadas $X_1(z)$ e $X_2(z)$. Além disso, a função Arg [X(z)] é descontínua e, portanto, não satisfaz às exigências de continuidade que resultam do fato de que $\hat{X}(z)$ deve ser analítica na circunferência de raio unitário.

Para contornar este problema, o logaritmo complexo contínuo é obtido a partir da integração de sua derivada. Dessa maneira :

$$\frac{d}{dz}\left[\log[X(z)]\right] = \frac{d}{dz}\left[\hat{X}(z)\right] = \frac{1}{X(z)}\,\frac{d}{dz}\left[X(z)\right] \tag{2.20}$$

Se esta derivada for avaliada sôbre a circunferência de raio unitário, obtem-se:

$$\frac{d}{d\omega}\left[\hat{X}(e^{j\omega})\right] = \frac{d}{d\omega}\left[\hat{X}_R(e^{j\omega})\right] + j\,\frac{d}{d\omega}\left[\hat{X}_I(e^{j\omega})\right] = \frac{d}{d\omega}\left[X(e^{j\omega})\right]\,\frac{1}{X(e^{j\omega})} \tag{2.21}$$

Lembrando que :

$$X(e^{j\omega}) = X_R(e^{j\omega}) + j\,X_I(e^{j\omega}) \tag{2.22}$$

e usando-se a equação (2.21), tem-se:

$$\hat{X}'(e^{j\omega}) = \frac{X_R'(e^{j\omega}) + j\,X_I'(e^{j\omega})}{X_R(e^{j\omega}) + j\,X_I(e^{j\omega})} \tag{2.23}$$

$$\hat{X}'(e^{j\omega}) = \frac{[X_R'(e^{j\omega})X_R(e^{j\omega})+X_I'(e^{j\omega})X_I(e^{j\omega})]}{X_R^2(e^{j\omega}) + X_I^2(e^{j\omega})} +$$

$$j \frac{[X_R(e^{j\omega})X_I'(e^{j\omega}) - X_R'(e^{j\omega})X_I(e^{j\omega})]}{X_R^2(e^{j\omega}) + X_I^2(e^{j\omega})}$$

onde o símbolo "'" denota a operação de derivada. Assim :

$$\frac{d}{d\omega}\left[\hat{X}_I(e^{j\omega})\right] = \frac{d}{d\omega}\left[\arg[X(e^{j\omega})]\right] = \frac{[X_R(e^{j\omega})X_I'(e^{j\omega}) - X_R'(e^{j\omega})X_I(e^{j\omega})]}{X_R^2(e^{j\omega}) + X_I^2(e^{j\omega})} \qquad (2.24)$$

Integrando a equação (2.24) em relação a ω e aplicando-se a condição : $\arg [X(e^{j\omega})]_{w=0} = 0$, pode-se garantir que $\arg [X(e^{j\omega})]$ será uma função ímpar e contínua em ω.

O problema da analiticidade da função $\hat{X}(z)$ se manifestará principalmente quando da implementação computacional da operação de logaritmo complexo .

2.3.2.1 - Cálculo de $\hat{x}(n)$

Voltando à equação (2.20), demonstra-se que :

$$z\,\hat{X}'(z) = \sum_{n=-\infty}^{\infty} [-n\,\hat{x}(n)]\,z^{-n} = z\,\frac{X'(z)}{X(z)} \qquad (2.25)$$

Dessa maneira :

$$-n\,\hat{x}(n) = \frac{1}{2\pi j}\int_c \frac{z\,X'(z)}{X(z)}\,z^{n-1}\,dz \qquad (2.26)$$

onde c representa um contorno fechado na região de convergência de $\hat{X}(z)$. Resolvendo a equação (2.26) para $\hat{x}(n)$, obtem-se:

$$\hat{x}(n) = \frac{-1}{2\pi jn}\int_c z\,\frac{X'(z)\,z^{n-1}}{X(z)}\,dz \qquad n \neq 0 \tag{2.27}$$

e

$$\hat{x}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{X}(e^{j\omega})d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}\hat{X}_R(e^{j\omega})d\omega + \frac{j}{2\pi}\int_{-\pi}^{\pi}\hat{X}_I(e^{j\omega})d\omega$$

Como $\hat{X}_I(e^{j\omega})$ é uma função ímpar, tem-se :

$$\hat{x}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln|X(e^{j\omega})|\,d\omega \tag{2.28}$$

As equações (2.27) e (2.28) podem ser utilizadas no cálculo de $\hat{x}(n)$ quando se dispõe da função X(z).

Partindo-se novamente da equação (2.20), tem-se:

$$z\,X'(z) = z\,\hat{X}'(z)\,X(z)$$

Aplicando-se a transformada Z inversa em ambos os lados desta equação, obtem-se :

$$n\,x(n) = \sum_{k=-\infty}^{\infty} k\,\hat{x}(k)\,x(n-k) \tag{2.29}$$

Dividindo ambos os membros por n :

$$x(n) = \sum_{k=-\infty}^{\infty} (k/n)\,\hat{x}(k)\,x(n-k) \qquad n \neq 0 \tag{2.30}$$

Neste caso, a equação (2.30) pode ser rearrumada num formato recursivo, permitindo o cálculo de $\hat{x}(n)$ diretamente a partir de x(n) em algumas implementações computacionais [1].

2.3.3 O SISTEMA LINEAR L

Para os Sistemas Homomórficos Convolucionais é usual realizar-se uma

abordagem ao sistema linear, considerando-o do tipo invariante com a frequência ao invés de invariante com o deslocamento, como normalmente é feito. Dessa maneira se está interessado na representação deste sistema no domínio do tempo, ou seja:

$$\frac{\hat{y}(n)}{\hat{x}(n)} = l(n) \qquad (2.31)$$

onde $l(n)$ faz o papel de uma função de transferência no domínio do tempo, é real e, em geral, estável. A utilidade desta classe de sistemas, bem como seus critérios de projeto, são analisados levando-se em consideração as propriedades gerais e específicas da sequência $\hat{x}(n)$, resultante de cada aplicação.

2.3.4 O SISTEMA CARACTERÍSTICO INVERSO

O sistema característico inverso D_*^{-1} está ilustrado na figura 2.6:

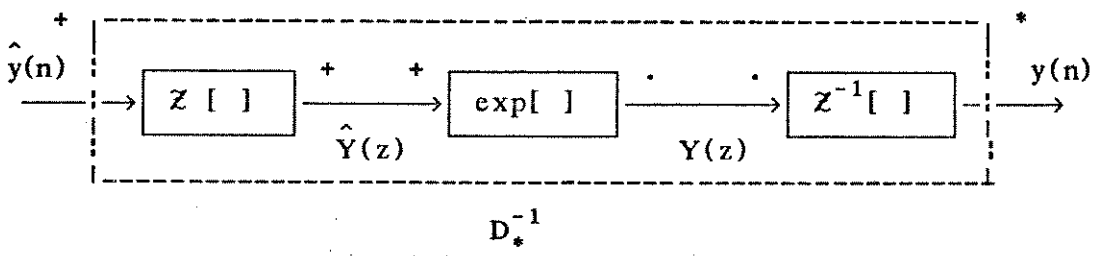


Figura 2.6 - Sistema Característico Inverso D_*^{-1}

Supondo-se que $\hat{x}(n)$ é uma sequência absolutamente somável, $\hat{y}(n)$ representa também uma sequência absolutamente somável. Dessa maneira, a região de convergência de $\hat{Y}(z)$ deve incluir a circunferência de raio unitário.

Como a função exponencial não apresenta problemas de unicidade e considerando que $\hat{Y}(z)$ é analítica sobre a circunferência de raio unitário, então $\exp[\hat{Y}(z)] = Y(z)$ também apresenta esta característica. Conclui-se então que $y(n) = \mathcal{Z}^{-1}[Y(z)]$, representa uma sequência absolutamente somável.

2.3.5 TERMINOLOGIA

A partir deste ponto a sequência $\hat{x}(n)$ será denominada de *cepstrum complexo*

da sequência x(n) segundo a terminologia estabelecida por Bogert, Healy e Tukey [4].

O termo *cepstrum* vem de uma inversão das primeiras letras da palavra inglesa *spectrum*. Esta terminologia é utilizada porque a sequência $\hat{x}(n)$ representa o "espectro" no domínio do tempo do logaritmo do espectro do sinal x(n). Esta inversão de domínios inspirou o termo *cepstrum*.

2.4 Propriedades do Cepstrum Complexo

Seja x(n) uma sequência de entrada em um Sistema Homomórfico Convolucional. Supondo X(z) do tipo racional, sua forma genérica é :

$$X(z) = \frac{A\, z^{r} \prod_{k=1}^{mi}(1 - a_k z^{-1}) \prod_{k=1}^{mo}(1 - b_k z)}{\prod_{k=1}^{pi}(1 - c_k z^{-1}) \prod_{k=1}^{po}(1 - d_k z)} \qquad (2.32)$$

Onde $|a_k|$, $|b_k|$, $|c_k|$ e $|d_k|$ são menores do que a unidade . Dessa forma, os fatores do tipo $(1 - a_k z^{-1})$ e $(1 - c_k z^{-1})$ correspondem a zeros e pólos dentro da circunferência de raio unitário e os fatores do tipo $(1 - b_k z)$ e $(1 - d_k z)$ correspondem a zeros e pólos localizados no exterior da circunferência de raio unitário. Em particular,quando não existem pólos fora da origem, isto é , o denominador da equação (2.32) é unitário, temos uma sequência de comprimento finito.

Calculando o logaritmo complexo como definido anteriormente tem-se :

$$\log [X(z)] = \hat{X}(z) = \ln A + \log [z^{r}] + \sum_{k=1}^{mi} \log [(1-a_k z^{-1})] \; + \sum_{k=1}^{mo} \log [(1-b_k z)] - \sum_{k=1}^{pi} \log [(1-c_k z^{-1})] - \sum_{k=1}^{po} \log [(1-d_k z)] \qquad (2.33)$$

Para sequências reais, A é real. Se o termo A for positivo, o fator Ln A contribue apenas para $\hat{x}(0)$. Se A é negativo,torna-se mais difícil determinar a contribuição ao cepstrum complexo relativa ao termo Ln A. Similarmente, o termo z^{r} na equação (2.32) corresponde apenas a um atraso ou avanço da sequência x(n). Se

$r = 0$, o termo $\log [z^r]$ se anula. Porém, se $r \neq 0$, haverá uma contribuição não nula para o cepstrum complexo. Apesar de ser perfeitamente possível levar-se em consideração os casos nos quais tem-se A negativo e/ou $r \neq 0$, isto não traz nenhuma vantagem real, pois, se duas funções da forma da equação (2.32) são multiplicadas (convolvidas no domínio do tempo), não é possível determinar qual a contribuição de cada uma delas para os valores A e r. Na prática, estes problemas são evitados usando-se o módulo do valor A e retirando-se o termo responsável pela fase linear, z^r. Estas informações são armazenadas e posteriormente recuperadas no final do processamento. Dessa forma:

$$X(z) = \frac{|A| \prod_{k=1}^{mi}(1 - a_k z^{-1}) \prod_{k=1}^{mo}(1 - b_k z)}{\prod_{k=1}^{pi}(1 - c_k z^{-1}) \prod_{k=1}^{po}(1 - d_k z)} \qquad (2.34)$$

E assim :

$$\log [X(z)] = \hat{X}(z) = \ln |A| + \sum_{k=1}^{mi} \log[(1 - a_k z^{-1})] + \sum_{k=1}^{mo} \log[(1-b_k z)] - \sum_{k=1}^{pi} \log[(1- c_k z^{-1})] - \sum_{k=1}^{po} \log[(1-d_k z)] \qquad (2.35)$$

Lembrando que :

$$\log(1 - \alpha z^{-1}) = -\sum_{n=1}^{\infty} \frac{\alpha^n}{n} z^{-n} \quad ; |z| > |\alpha|$$

$$\log(1 - \beta z) = -\sum_{n=1}^{\infty} \frac{\beta^n}{n} z^{n} \quad ; |z| < \frac{1}{|\beta|}$$

e considerando que cada um dos fatores na equação (2.35) deve representar uma transformada $\mathcal{Z}$ com região de convergência incluindo a circunferência de raio unitário, tem-se:

$$\hat{x}(n) = \ln |A| \qquad ; n = 0 \tag{2.36.1}$$

$$= -\sum_{k=1}^{mi} \frac{a_k^n}{n} + \sum_{k=1}^{pi} \frac{c_k^n}{n} \qquad ; n > 0 \tag{2.36.2}$$

$$= \sum_{k=1}^{mo} \frac{b_k^{-n}}{n} - \sum_{k=1}^{po} \frac{d_k^{-n}}{n} \qquad ; n < 0 \tag{2.36.3}$$

No caso de sequências de comprimento finito,o segundo termo do lado direito das equações (2.36.2) e (2.36.3) se anula.

Analisando-se as equações (2.36), podem ser atribuídas as seguintes propriedades ao cepstrum complexo:

P.1 - O cepstrum complexo cai, pelo menos, com 1/n ;

$$|\hat{x}(n)| < C \left| \frac{\alpha^{|n|}}{n} \right| \qquad -\infty < n < \infty$$

onde α é o valor máximo dentre os valores $|a_k|$, $|b_k|$, $|c_k|$ e $|d_k|$.

P.2 - Se x(n) é de fase mínima (X(z) não possui nenhum pólo ou zero fora da circunferência de raio unitário), então:

$$\hat{x}(n) = 0 \qquad ; n < 0$$

P.3 - Se x(n) é de fase máxima (X(z) não possui nenhum pólo ou zero no interior da circunferência de raio unitário), então:

$$\hat{x}(n) = 0 \qquad ; n > 0$$

P.4 - Mesmo que x(n) seja de duração finita, $\hat{x}(n)$ tem duração infinita.

2.4.1 SEQUÊNCIAS DE FASE MÍNIMA E FASE MÁXIMA

Seja inicialmente uma sequência de fase mínima x(n), cuja transformada $\mathcal{Z}$ é do tipo:

$$X(z) = \frac{|A| \prod_{k=1}^{mi} (1 - a_k z^{-1})}{\prod_{k=1}^{pi} (1 - c_k z^{-1})}$$

Analisando a equação anterior nota-se que:

$$x(n) = 0 \qquad ; n < 0$$

e pela propriedade **P.2** :

$$\hat{x}(n) = 0 \qquad ; n < 0$$

Assim, o cepstrum complexo de uma sequência de fase mínima é causal. Da teoria de *Transformada de Hilbert* [1], tem-se que a transformada $\mathcal{Z}$ de uma sequência real, causal e absolutamente somável é completamente determinada pela parte real de sua transformada de Fourier. Como $\hat{x}(n)$ é causal, o cálculo de $\hat{X}_R(e^{j\omega}) = \mathbf{Ln}\ [|X(e^{j\omega})|]$ é suficiente para obter-se $\hat{x}(n)$. A transformada inversa de $\hat{X}_R(e^{j\omega})$ é igual a componente par da sequência $\hat{x}(n)$ [1] e assim:

$$\mathcal{F}^{-1}[\hat{X}_R(e^{j\omega})] = \hat{x}_P(n) = \frac{\hat{x}(n) + \hat{x}(-n)}{2} \qquad (2.37)$$

A sequência $\hat{x}_P(n)$ é denominada *Cepstrum* da sequência x(n). Como $\hat{x}(n)$ é nulo para *n* negativo, pode-se escrever :

$$\hat{x}(n) = \hat{x}_P(n) \cdot u_+(n) \qquad (2.38)$$

onde :

$$u_+(n) = \begin{cases} 0 & ; n < 0 \\ 1 & ; n = 0 \\ 2 & ; n > 0 \end{cases}$$

Esta exposição demonstra que para sequências de fase mínima, o cepstrum complexo pode ser obtido a partir apenas do cálculo do logaritmo do módulo da transformada de Fourier $X(e^{j\omega})$. Este procedimento é denominado *Cálculo do Cepstrum Complexo a partir de uma Realização de Fase Mínima.* A figura 2.7 ilustra o procedimento a ser seguido [1].

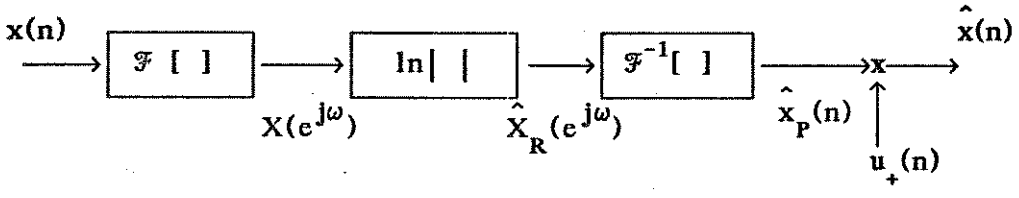


Figura 2.7- Cálculo do Cepstrum Complexo em Realizações de Fase Mínima

No caso de sequências de fase máxima,pode-se obter um resultado semelhante. Para este tipo de sequências tem-se:

$$x(n) = \hat{x}(n) = 0 \qquad ; n > 0$$

Como no caso anterior, apenas o logaritmo do módulo da transformada de Fourier, $X(e^{j\omega})$, é necessário para se computar o cepstrum complexo $\hat{x}(n)$:

$$\hat{x}(n) = u_-(n) \cdot \hat{x}_P(n) \qquad (2.39)$$

onde: $\hat{x}_P(n)$ é dado pela equação (2.37); e

$$u_-(n) = \begin{cases} 2 & ; n < 0 \\ 1 & ; n = 0 \\ 0 & ; n > 0 \end{cases}$$

Finalmente, considere o procedimento adotado para sequências de fase mínima e ilustrado na figura 2.7, aplicado a sequências de fase mista. Uma sequência de fase mista , a menos de uma componente linear de fase, pode ser representada da seguinte maneira:

$$x(n) = x_{min}(n) * x_{max}(n) \tag{2.40}$$

onde: $x_{min}(n)$ é uma sequência de fase mínima;
$x_{max}(n)$ é uma sequência de fase máxima.

Calculando a transformada de Fourier de ambos os membros da equação (2.40), obtem-se :

$$X(e^{j\omega}) = X_{min}(e^{j\omega}).\ X_{max}(e^{j\omega}) \tag{2.41}$$

Calculando o logaritmo do módulo, tem-se:

$$\ln [|X(e^{j\omega})|] = \ln [|X_{min}(e^{j\omega}).\ X_{max}(e^{j\omega})|]$$

$$= \ln [|X_{min}(e^{j\omega})|] + \ln [|\ X_{max}(e^{j\omega})|]$$

$$\hat{X}_R(e^{j\omega}) = \hat{X}_{R\ min}(e^{j\omega}) + \hat{X}_{R\ max}(e^{j\omega}) \tag{2.42}$$

Fazendo a transformada inversa,

$$\hat{x}_P(n) = \hat{x}_{P\ min}(n) + \hat{x}_{P\ max}(n) \tag{2.43}$$

onde : $\hat{x}_{P\ min}(n) = \dfrac{\hat{x}_{min}(n) + \hat{x}_{min}(-n)}{2}$

$$\hat{x}_{P\ max}(n) = \frac{\hat{x}_{max}(n) + \hat{x}_{max}(-n)}{2}$$

Multiplicando a sequência $\hat{x}_P(n)$ pela sequência $u_+(n)$ e levando-se em conta que :

$$\hat{x}_{min}(-n) = 0 \qquad p/\ n > 0$$

e

$$\hat{x}_{max}(n) = 0 \qquad p/\ n > 0$$

tem-se:

$$\hat{y}(n) = \hat{x}_p(n).u_+(n) = \hat{x}_{min}(n) + \hat{x}_{max}(-n) \qquad ; n \geq 0 \qquad (2.44)$$

$$= 0 \qquad ; n < 0$$

O processamento da sequência $\hat{y}(n)$ no sistema inverso àquele ilustrado na figura 2.7 produz os seguintes resultados:

$$\hat{Y}(e^{j\omega}) = \hat{X}_{min}(e^{j\omega}) + \hat{X}_{max}(e^{-j\omega})$$

$$Y(e^{j\omega}) = \exp[\hat{Y}(e^{j\omega})] = X_{min}(e^{j\omega}) \cdot X_{max}(e^{-j\omega}) \qquad (2.45)$$

Como $|X_{max}(e^{j\omega})| = |X_{max}(e^{-j\omega})|$, conclui-se que este procedimento resulta em uma sequência de fase mínima (Ver equação 2.44 e propriedade **P.2**) com espectro de amplitude idêntico ao espectro de amplitude da sequência x(n) original.

Como será visto posteriormente, este tipo de procedimento é bastante útil no processamento de sinais de voz, pois o ouvido humano é pouco sensível às variações de fase [1,3].

2.5 Aspectos Computacionais

2.5.1 CÁLCULO DO CEPSTRUM COMPLEXO

Na maioria das implementações computacionais do Sistema Homomórfico Convolucional, utiliza-se a Transformada Discreta de Fourier (DFT).

Como o cepstrum complexo possui duração infinita, o uso da DFT resulta numa versão do cepstrum complexo que apresenta o fenômeno do *aliasing*:

$$\hat{x}_a(n) = \sum_{k=-\infty}^{\infty} \hat{x}(n + kN) \qquad (2.46)$$

onde N é o tamanho da DFT utilizada. Porém, as amplitudes das amostras do cepstrum complexo diminuem rapidamente com o deslocamento n (Propriedade **P.1**), o que garante que o efeito do *aliasing* pode ser considerado desprezível para N suficientemente grande.

Outro ponto a ser analisado é o cálculo da fase da transformada de Fourier, $X(e^{j\omega})$. A maioria dos computadores de uso geral trabalha com funções do tipo **arctg**, que calculam o valor principal de um dado argumento, ou seja (Ver Figura5.b):

$$-\Pi < \text{ARG}\,[X(e^{j\omega})] \leq \Pi$$

O problema da não ambiguidade da operação do logaritmo complexo, bem como o da existência de $\hat{X}(e^{j\omega})$, requerem que a fase da transformada de Fourier $X(e^{j\omega})$ seja uma função contínua. Dessa maneira, após a computação do argumento de $X(e^{j\omega})$ deve ser realizada uma operação de correção das descontinuidades obtidas, conhecida como *"Unwrapping" de Fase* [5]. Esta operação, a grosso modo, detecta a existência das descontinuidades devidas à função **arctg** e corrige-as através da soma ou subtração de múltiplos inteiros de 2Π.

Neste ponto deve-se ressaltar que o tamanho da DFT utilizada também é importante para um bom desempenho do algoritmo de correção de fase, pois, quanto maior for a dimensão da DFT, mais próximas entre si estarão as amostras da função $\text{ARG}[X(e^{j\omega})]$. Assim o trabalho de detecção das descontinuidades devidas à função **arctg** será realizado de modo mais preciso, já que amostras muito afastadas poderiam induzir o algoritmo a detectar falsas descontinuidades no argumento.

O procedimento convencional de cálculo do cepstrum complexo, isto é, levando-se em conta a informação do espectro de fase da sequência x(n), é denominado *Cálculo do Cepstrum Complexo a Partir de uma Realização de Fase Mista.*

2.5.2 CÁLCULO DO CEPSTRUM

Da mesma forma, o uso da DFT introduz o fenômeno do *aliasing* na sequência obtida:

$$\hat{x}_{Pa}(n) = \sum_{k=-\infty}^{\infty} \hat{x}_p(n + kN) \qquad (2.47)$$

Como no caso anterior, o uso de uma DFT com dimensão N suficientemente grande reduz os efeitos do *aliasing* a níveis desprezíveis.

2.5.3 CÁLCULO DO CEPSTRUM COMPLEXO EM REALIZAÇÕES DE FASE MÍNIMA

No cálculo do cepstrum complexo em realizações de fase mínima utiliza-se a seguinte expressão :

$$\begin{aligned} \hat{c}_a(n) &= \hat{x}_{Pa}(n) && ; n = 0, N/2 \\ &= 2.\hat{x}_{Pa}(n) && ; 1 \leq n < N/2 \\ &= 0 && ; N/2 < n \leq N-1 \end{aligned} \qquad (2.48)$$

onde : $\hat{x}_{Pa}(n)$ é dado pela equação (2.47) e N é a dimensão da DFT utilizada.

Deve-se notar que $\hat{c}_a(n) \neq \hat{x}_a(n)$, pois neste caso é a componente par do cepstrum complexo que apresenta o fenômeno do *aliasing* e não o próprio cepstrum complexo como descrito no ítem 2.4.1. Convencionou-se denominar o cepstrum complexo calculado a partir de uma realização de fase mínima, apenas de *cepstrum* de x(n), enquanto que o termo *cepstrum complexo* caracteriza uma realização de fase mista.

2.6 REFERÊNCIAS

[1] A. V. Oppenheim and R. W. Schafer, *"Digital Signal Processing"*. Englewood Cliffs, NJ,Prentice - Hall, 1975.

[2] A. V. Oppenheim and R. W. Schafer, *" Homomorphic Analysis of Speech"*, IEEE Trans. Audio Electroacoust., vol AU-16 pp 221-226 June 1968.

[3] L. R. Rabiner and R. W. Schafer, *"Digital Processing of Speech Signals"*, Englewood Cliffs, NJ, Prentice-Hall, 1978.

[4] B. P. Bogert, M Healy and J. Tukey, *"The Quefrency Analysis of Time Series for Echoes : Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum and Shape Cracking"*, in Proc. Symp on Time Series Analysis, M. Rosemblatt, Ed. New York : Wiley, 1963, Ch 15 pp 209-243

[5] J. M. Tribolet, " *A New Phase Unwrapping Algorithm*", IEEE Trans.Acoust. Speech, Signal Processing, Vol ASSP-25, pp 170-177, April 1977

CAPÍTULO 3

APLICAÇÃO DO PROCESSAMENTO HOMOMÓRFICO NA ANÁLISE-SÍNTESE DE SINAIS DE VOZ

CONTEÚDO

3.1 INTRODUÇÃO

Neste capítulo são analisados os aspectos específicos do Processamento Homomórfico aplicado à Análise-Síntese de sinais de voz. O estudo é realizado com base na Representação Paramétrica utilizando o modelo tradicional de produção de sinais de voz.

3.2 REPRESENTAÇÃO PARAMÉTRICA DO SINAL DE VOZ

A Codificação de Voz a Baixas Taxas utiliza um procedimento de Análise-Síntese baseado numa Representação Paramétrica da Voz. A Representação

Paramétrica, por sua vez, está baseada num modelo de producão de voz a partir do qual é possível representar o sinal de voz como sendo o resultado da filtragem de um sinal de excitação conveniente através de um sistema digital linear que reproduz os efeitos do Aparelho Fonador humano.

Como visto anteriormente, no modelo tradicional de produção de voz, reproduzido na figura 1.1, os efeitos do *Pulso Glótico,* do *Trato Vocal* e da *Impedância de Irradiação* são combinados em um filtro digital H(z), cujos parâmetros característicos são considerados fixos para um intervalo de observação T_0 suficientemente pequeno. O sinal de voz é produzido a partir da convolução da resposta impulsiva h(n) do filtro digital H(z) com o sinal de excitação p(n).

Os sons sonoros são gerados excitando-se o filtro H(z) com um sinal formado por um trem de impulsos unitários periódico com período N_0. Este período é conhecido como *Período de Pitch.* Os sons não sonoros são produzidos usando-se como excitação um sinal do tipo ruído branco. As duas fontes de excitação são consideradas independentes entre si. A informação de amplitude dos sinais de voz produzidos está embutida no parâmetro G.

Para cada intervalo de tempo T_0, que caracteriza um *quadro* do sinal de voz, devem ser definidos :

- a natureza do som produzido : sonoro ou não sonoro;
- o período de Pitch, no caso de sons sonoros e
- os parâmetros característicos do filtro digital H(z).

Estes parâmetros constituem os *Parâmetros do Modelo.* Uma vez definido, este conjunto de parâmetros pode ser transmitido, o que permite a reconstrucão do sinal de voz original na recepção com alto grau de inteligibilidade, desde que os parâmetros tenham sido obtidos de maneira adequada.

Na representação do filtro H(z), os parâmetros característicos normalmente são os pólos e os zeros, ou de outro modo, os coeficientes do denominador e do numerador , supondo o sistema H(z) racional. A técnica da Análise LPC tradicional utiliza esta estratégia, porém considerando o filtro H(z) formado apenas por pólos.

Como visto no capítulo anterior, as amostras do cepstrum complexo de uma sequência genérica x(n) diminuem rapidamente em amplitude com o tempo. Esta propriedade pode ser explorada para se obter um pequeno conjunto de amostras do cepstrum complexo da resposta impulsiva do filtro H(z) que , ao ser transmitido, seja suficiente para a realização de uma boa estimativa da resposta impulsiva na recepção.

A seguir será analisada a técnica da *Desconvolução Homomórfica*. A partir desta técnica é possível obter o cepstrum complexo da resposta impulsiva do filtro H(z). Esta técnica não impõe nenhuma restrição quanto ao formato do filtro H(z), como realizado na Análise LPC tradicional. Isto permite a estimativa de um filtro genérico para simular o trato vocal.

3.3 Deconvolução Homomórfica da Voz

A técnica da *Desconvolução Homomórfica* pode ser aplicada na estimação do cepstrum complexo da resposta impulsiva do filtro H(z) do modelo tradicional de produção de sinais de voz, considerando-se este modelo como sendo válido em um curto intervalo de tempo T_0. Como o sinal de voz é o resultado da convolução da resposta impulsiva do filtro H(z) com um sinal de excitação adequado, a Desconvolução Homomórfica permite separar os cepstruns complexos da excitação e da resposta impulsiva.

Seja inicialmente um segmento de som sonoro s(n) :

$$s(n) = p(n) * h(n) \qquad 0 \le n \le L - 1 \qquad (3.1)$$

onde p(n) é a excitação;

h(n) é a resposta impulsiva do filtro H(z) e

L tamanho do segmento de voz.

Para minimizar os efeitos das descontinuidades nas extremidades do segmento, multiplica-se o sinal s(n) por uma janela w(n) com transição gradual nos extremos, como por exemplo, a janela de Hamming. Assim :

$$\begin{aligned} x(n) &= s(n) \cdot w(n) \\ x(n) &= [p(n) * h(n)] \cdot w(n) \end{aligned} \qquad (3.2)$$

Se a janela w(n) varia lentamente em relação ao termo h(n) dentro do intervalo L, pode-se escrever [1,2]:

$$x(n) \cong p_w(n) * h(n) \qquad (3.3)$$

onde :

$$p_w(n) = p(n) \cdot w(n) \tag{3.4}$$

A nível de cepstrum complexo tem-se:

$$\hat{x}(n) = \hat{p}_w(n) + \hat{h}(n) \tag{3.5}$$

A excitação p(n) pode ser escrita como :

$$p(n) = \sum_{k=0}^{M-1} \delta(n - kN_o) \tag{3.6}$$

e assim :

$$p_w(n) = \sum_{k=0}^{M-1} w(kN_o).\delta(n - kN_o) \tag{3.7}$$

Nas equações (3.6) e (3.7) assume-se que M impulsos ocorrem na janela w(n).

A transformada $\mathcal{Z}$ da sequência $p_w(n)$ pode ser escrita como:

$$P_w(z) = \sum_{k=0}^{M-1} w(kN_o)\, z^{-kNo} \tag{3.8}$$

Como será demonstrado adiante, o período de Pitch, N_0, determina os pontos de ocorrência de amostras não-nulas do cepstrum da sequência $p_w(n)$.

Da equação (3.8) conclui-se que $P_w(z)$ é um polinômio na variável z^{-No}. Assim , $P_w(z)$ pode ser expresso como o produto de fatores do tipo $(1 - \alpha z^{-No})$ e $(1 - \beta z^{No})$, ou seja :

$$P_w(z) = A \prod_{k=1}^{mi}(1 - \alpha_k z^{-No}) \prod_{k=1}^{mo}(1 - \beta_k z^{No}) \tag{3.9}$$

onde $|\alpha_k|$, $|\beta_k| < 1$ e os fatores $(1 - \alpha_k z^{-No})$ e $(1 - \beta_k z^{No})$ representam os zeros no interior e exterior da circunferência de raio unitário. Além disso, A = w(0), pois w(0) é o coeficiente da potência nula de z. Entretanto, deve ser ressaltado que o

termo **A** não corresponde sempre à primeira amostra da janela , mas sim a amplitude da janela na posição do primeiro impulso de excitação de um quadro. Este fato será abordado com mais detalhes no ítem 4.2.

Calculando o logaritmo complexo da equação (3.9) e lembrando que :

$$\log\,(\,1 - \alpha z^{-No}) = -\sum_{n=1}^{\infty} \frac{\alpha^n}{n}\; z^{-nNo} \quad ;\; |z| > |\alpha|$$

$$\log\,(\,1 - \beta z^{No}) = -\sum_{n=1}^{\infty} \frac{\beta^n}{n}\; z^{nNo} \quad ;\; |z| < \frac{1}{|\beta|}$$

obtem-se:

$$\hat{p}_w(n) = \ln A \quad ;\; n = 0 \qquad (3.10.1)$$

$$= -\sum_{k=1}^{mi} \frac{\alpha_k^{n/No}}{n/No} \quad ;\; n = N_o,\, 2N_o,\, 3N_o.. \qquad (3.10.2)$$

$$= \sum_{k=1}^{mo} \frac{\beta_k^{n/No}}{n/No} \quad ;\; n = -N_o,-2N_o,-3N_o.. \qquad (3.10.3)$$

Este conjunto de equações permite concluir que $\hat{p}_w(n)$ será diferente de zero apenas em múltiplos inteiros de N_o. Além disso, a taxa de queda da amplitude do cepstrum complexo da sequência $p_w(n)$ com o deslocamento é inversamente proporcional a N_o, o que torna $\hat{p}_w(n)$ mais sensível aos problemas de *aliasing* . As equações (3.10.1 - 3.10.3) mostram também que a contribuição de $\hat{p}_w(n)$ ao cepstrum complexo $\hat{x}(n)$ está presente na região $|n| \geq N_o$, a menos de uma amostra na origem, $\hat{p}_w(0)$.

O cepstrum complexo da resposta impulsiva, $\hat{h}(n)$, pode ser obtido supondo H(z) do tipo [3]:

$$H(z) = G\; \frac{z^{-M}\; \prod_{k=1}^{mi}(1 - a_k z^{-1})\; \prod_{k=1}^{mo}(1 - b_k z)}{\prod_{k=1}^{p}(1 - c_k z^{-1})} \qquad (3.11)$$

onde: $|a_k|, |b_k|$ e $|c_k| < 1$; e $G > 0$.

Para sons sonoros, exceto os nasalados, pode-se supor $a_k = b_k = 0$ para qualquer valor de k [3]. Para sons não-sonoros e sonoros nasalados, é necessário incluir em H(z) a contribuição tanto de pólos quanto de zeros. Alguns zeros podem estar situados no exterior da circunferência de raio unitário, enquanto que todos os pólos , c_k, devem estar localizados no interior da circunferência de raio unitário , para que H(z) represente um sistema estável. Além disso, como a resposta impulsiva , h(n), deve ser real , todos os pólos e zeros complexos devem ocorrer em pares conjugados.

Retirando o têrmo responsável pela fase linear, z^{-M}, e calculando o logarítmo complexo da equação (3.11), tem-se :

$$\hat{H}(z) = \ln G + \sum_{k=1}^{mi} \log[(1 - a_k z^{-1})] + \sum_{k=1}^{mo} \log[(1 - b_k z^{-1})] - \sum_{k=1}^{P} \log[(1 - c_k z^{-1})] \qquad (3.12)$$

Dessa maneira, tem-se :

$$\hat{h}(n) = \sum_{k=1}^{mo} \frac{b_k^n}{n} \qquad ; n < 0 \qquad (3.13)$$

$$= \ln G \qquad ; n = 0$$

$$= \sum_{k=1}^{P} \frac{c_k^n}{n} - \sum_{k=1}^{mi} \frac{a_k^n}{n} \qquad ; n > 0$$

De um modo geral,o cepstrum complexo $\hat{h}(n)$ decai rapidamente. Assim, para valores de N_o razoavelmente grandes,o cepstrum complexo da resposta impulsiva não se superpõe de forma significativa com o cepstrum complexo $\hat{p}_w(n)$, a menos de uma componente do cepstrum complexo da excitação na origem.

A discussão anterior sugere um método para se obter uma separação aproximada dos cepstrums complexos relativos à resposta impulsiva h(n) e à excitação p(n), a partir do cepstrum complexo $\hat{x}(n)$, ou seja, pode-se aplicar um filtro passa-baixos tempos ao sinal $\hat{x}(n)$ para se obter $\hat{h}(n)$ e um filtro passa-altos tempos para se obter $\hat{p}_w(n)$. Assim :

$$\hat{x}(n) = \hat{y}(n) \,.\, l(n)$$

onde :

$$\begin{aligned} l(n) &= 1 \quad ; |n| < N_o \\ &= 0 \quad ; |n| \geq N_o \end{aligned} \tag{3.14}$$

Esta sequência é denominada *Janela Cepstral*. Tem-se então:

$$\begin{aligned} \hat{y}(n) &= \hat{h}(n) && ; n < 0 \\ &= \hat{h}(0) + \hat{p}_w(0) && ; n = 0 \\ &= \hat{h}(n) && ; n > 0 \end{aligned}$$

onde $\hat{p}_w(0) = \ln A$.

Processando o sinal $\hat{y}(n)$ pelo sistema característico inverso tem-se :

$$\hat{Y}(z) = \hat{H}(z) + \hat{p}_w(0)$$

$$y(n) = \mathcal{Z}^{-1}\{\exp[\hat{Y}(z)]\} = p_w(0).h(n) \tag{3.15}$$

Dessa maneira, conclui-se que a amostra do cepstrum complexo da excitação na origem representa apenas um fator de escala que aparece multiplicado à resposta impulsiva desejada.

Seja agora um segmento de som não-sonoro s(n) :

$$s(n) = r(n) * h(n) \qquad 0 \leq n \leq L - 1 \tag{3.16}$$

onde r(n) é uma realização de um ruído branco de média zero e variância unitária. Aplicando-se uma janela w(n), tem-se:

$$x(n) = [r(n) * h(n)] \,.\, w(n) \tag{3.17}$$

ou ainda:

$$\begin{aligned} x(n) &\cong [r(n) \,.\, w(n)] * h(n) \\ x(n) &\cong r_w(n) * h(n) \end{aligned} \tag{3.18}$$

Passando a equação (3.18) para o domínio da frequência e calculando o logaritmo complexo em ambos os membros, tem-se :

$$\log[X(e^{j\omega})] = \log[R(e^{j\omega}) * W(e^{j\omega})] + \log[H(e^{j\omega})]$$

$$\hat{X}(e^{j\omega}) = \log[R(e^{j\omega}) * W(e^{j\omega})] + \hat{H}(e^{j\omega}) \qquad (3.19)$$

Como o ruído é branco e possui variância unitária, o espectro de uma realização num dado quadro oscila em torno de uma constante igual à unidade. Assim, a convolução $[R(e^{j\omega}) * W(e^{j\omega})]$ é aproximadamente igual a $1/2\pi$ vezes a área de $W(e^{j\omega})$ no intervalo de $-\Pi$ a Π. Consequentemente tem-se :

$$\hat{x}(n) \cong k\,\delta(n) + \hat{h}(n) \qquad (3.20)$$

onde $k \cong \ln [1/2\pi(\text{área de } W(e^{j\omega}))]$, ou seja, o cepstrum do ruído branco é igual a um impulso na origem cuja amplitude está vinculada à janela w(n). Esta análise é inteiramente confirmada por observações práticas dos cepstruns complexos correspondentes a sequências de ruído branco janeladas.

Como no caso do segmento de som sonoro, pode ser utilizado um filtro passa-baixos tempos para obter-se as amostras mais significativas do cepstrum complexo da resposta impulsiva. O processamento pelo sistema característico inverso retornará a resposta impulsiva h(n), ponderada por um fator de escala devido ao impulso na origem referente ao cepstrum da excitação.

Uma vez obtida a resposta impulsiva do filtro H(z), o sinal de voz pode ser reconstruído a partir da convolução entre a resposta impulsiva e o sinal de excitação apropriado.

A análise desenvolvida até este ponto é perfeitamente válida para uma Realização de Fase Mínima. A única diferença básica é que as estimativas para as respostas impulsivas obtidas possuem espectros de amplitude idênticos aos espectros de amplitude das respostas impulsivas reais, porém com fase mínima. Como o ouvido humano é pouco sensível às variações de fase, esta realização é perfeitamente viável [3].

3.4 Imperfeições do Modelo Convolucional

O modelo convolucional, expresso na equação (3.3), apresenta algumas imperfeições quando utilizado na Desconvolução Homomórfica de sons sonoros, principalmente quando o período de Pitch é pequeno em relação ao comprimento da janela w(n).

Voltando à equação (3.2) e passando ao domínio da frequência tem-se :

$$X(\omega) = [\ H(\omega)\ .\ P(\omega)] * W(\omega) \qquad (3.21)$$

onde:

$$P(\omega) = 2\pi/No \sum_{k} \delta(\omega - k\ \frac{2\pi}{No}\ )$$

Fazendo $2\pi/No = \omega_0$ e $k\omega_0 = \omega_k$, tem-se:

$$X(\omega) = 2\pi/No[\ H(\omega)\ .\ \sum_{k} \delta(\ \omega - \omega_k)\ ] * W(\omega) \qquad (3.22)$$

ou :

$$X(\omega) = [\ \sum_{k}\ H(\omega_k)\ \delta(\ \omega - \omega_k)\ ] * \bar{W}(\omega) \qquad (3.23)$$

onde : $\bar{W}(\omega) = W(\omega)\ .\ 2\pi/No$. Assim :

$$X(\omega) = \sum_{k}\ H(\omega_k)\ \bar{W}(\omega - \omega_k) \qquad (3.24)$$

Ou seja, $X(\omega)$ é representado por uma sequência de espectros da janela w(n), deslocados de ω_k e ponderados por amostras do espectro do filtro H(z) nas frequências ω_k.

Usando o modelo convolucional da equação (3.3) tem-se:

$$X(\omega) \cong X_1(\omega) = H(\omega)\ .\ [P(\omega) * W(\omega)] \qquad (3.25)$$

Repetindo-se o procedimento anterior chega-se a seguinte equação:

$$X(\omega) \cong X_1(\omega) = H(\omega) \cdot \sum_k \bar{W}(\omega - \omega_k) \tag{3.26}$$

Dessa maneira, pode-se definir um sinal $E(\omega)$ que representa o erro entre o espectro $X(\omega)$ verdadeiro e o espectro $X_1(\omega)$ resultante do modelo convolucional:

$$X(\omega) = X_1(\omega).\ E(\omega)$$

onde:

$$E(\omega) = \frac{\sum_k H(\omega_k) \cdot \bar{W}(\omega - \omega_k)}{H(\omega) \sum_k \bar{W}(\omega - \omega_k)} \tag{3.27}$$

O desenvolvimento da equação (3.27) foi realizado supondo uma fase inicial nula no sinal de excitação p(n). Este procedimento pode ser generalizado considerando-se uma fase linear não-nula , resultando na equação:

$$E(\omega) = \frac{\sum_k H(\omega_k).\exp(jn_0\omega_k).\ \bar{W}(\omega - \omega_k)}{H(\omega) \sum_k \exp(jn_0\omega_k).\bar{W}(\omega - \omega_k)} \tag{3.28}$$

Neste caso a fase linear de $H(\omega)$ poderá ser incluída em n_0.

3.4.1 ANÁLISE DO SINAL DE ERRO NUMA REALIZAÇÃO DE FASE MÍNIMA

Supondo uma realização de fase mínima, o processamento homomórfico convolucional produz na saída do sistema um sinal de erro $E_h(\omega)$ dado por :

$$E_h(\omega) = \frac{\text{Exp}\ \{\ \ln\ |\ X(\omega)|\ *\ L(\omega)\ \}}{\text{Exp}\ \{\ \ln\ |X_1(\omega)|\ *\ L(\omega)\ \}} \tag{3.29}$$

onde $L(\omega)$ é a transformada de Fourier da janela cepstral $l(n)$, a qual, neste caso, inclui a janela $u_+(n)$. Pode-se escrever também :

$$E_h(\omega) = \frac{H'(\omega)}{H(\omega)} \qquad (3.30)$$

A equação (3.30) reflete alguns aspectos bastante importantes que precisam ser ressaltados. Em primeiro lugar, a Desconvolução Homomórfica só é aplicada em uma sequência $x(n)$ que possa ser expressa por uma equação de formato idêntico àquele da equação (3.3). Dessa forma, para que a aplicação da Desconvolução Homomórfica resulte na verdadeira resposta em frequência do trato vocal, $H(\omega)$, é necessário que a equação (3.3) seja perfeitamente válida

Na realidade, dadas as diferenças entre as equações (3.2) e (3.3), as operações desenvolvidas na Desconvolução Homomórfica resultam numa estimativa $H'(\omega)$ que deve ser a mais próxima possível da resposta impulsiva desejada $H(\omega)$.

O modelo convolucional da equação (3.3) é utilizado supondo-se que a janela $w(n)$ varie lentamente durante o comprimento efetivo da resposta impulsiva $h(n)$, ou , no domínio da frequência, $H(\omega)$ varie lentamente durante o comprimento efetivo de $W(\omega)$. Analisando a equação (3.27), a validade desta hipótese resulta :

$$E(\omega) \cong 1$$

e então :

$$E_h(\omega) \cong 1 \quad \Rightarrow \quad H'(\omega) \cong H(\omega)$$

Neste caso, a aplicação da Desconvolução Homomórfica resulta numa resposta $H'(\omega)$ que é uma boa aproximação para $H(\omega)$.

Nos casos em que a hipótese não é válida, o erro $E_h(\omega)$ dependerá de uma forma bastante complicada de $H(\omega)$ e da janela $w(n)$. Entretanto, uma análise da equação (3.27) mostra que o sinal erro $E_h(\omega)$ atinge valores elevados próximo às regiões onde $H(\omega) \cong 0$ e nos nulos periódicos do espectro $\sum_k W(\omega - \omega_k)$.

O espectro $\sum_k \bar{W}(\omega - \omega_k)$ é uma repetição periódica do espectro da janela $w(n)$ com período ω_o. Para valores pequenos do período de Pitch, correspondentes a voz feminina, o espalhamento dos espectros $W(\omega)$ resulta em extensas regiões onde o espectro $\sum_k \bar{W}(\omega - \omega_k)$ praticamente se anula.e, consequentemente, em valores elevados

para o sinal erro $E_h(\omega)$ nestas regiões. Estes erros podem ser minimizados com o preenchimento destas regiões a partir do alargamento do espectro da janela w(n). Esta operação é equivalente ao encurtamento da janela w(n) no domínio do tempo.

Entretanto, deve ser ressaltado que o encurtamento excessivo da janela w(n), apesar de contribuir para a validade do modelo convolucional de um ponto de vista matemático, tende a degradar a estimativa de $H(\omega)$. Isto se deve principalmente à perda de informação disponível no segmento de voz em análise, ou seja, a estimativa da resposta em frequência do trato vocal para um dado segmento de voz é realizada a partir de uma quantidade de amostras bastante inferior ao comprimento deste segmento. Assim, existe um comprimento ótimo para a janela w(n) que garante uma boa estimativa para $H(\omega)$ com o menor erro $E_h(\omega)$. Este valor é de 2 a 2,5 vezes o período de Pitch para uma janela do tipo Hamming [5]. Este comprimento ótimo , no entanto, é limitado pelo tamanho do segmento original de voz.

Vale ressaltar que, mesmo sob condições ótimas, o janelamento adaptado ao período de Pitch torna indispensável, na reconstrução do sinal de voz, a aplicação de alguma técnica que proporcione uma variação mais suave da resposta impulsiva do trato vocal de um dado segmento para o segmento seguinte, com o objetivo de compensar as variações bruscas resultantes do estreitamento da janela w(n).

Apesar de ser possível uma minimização dos erros devidos aos nulos do espectro $\sum_k \bar{W}(\omega - \omega_k)$, os erros devidos aos zeros da resposta em frequência do trato vocal ainda persistem na estimativa final de $H(\omega)$. Dessa forma, é de se esperar que para segmentos sonoros, nos quais a função $H(\omega)$ correspondente contenha zeros, a estimativa $H'(\omega)$ deve apresentar maiores imperfeições. Este é o caso de alguns sons sonoros nasalados.

3.4.2 ANÁLISE DO SINAL DE ERRO NUMA REALIZAÇÃO DE FASE MISTA

Numa realização de fase mista, isto é , utilizando o cepstrum complexo, a Desconvolução Homomórfica produz na saída do sistema um sinal de erro $E_h(\omega)$ dado por:

$$E_h(\omega) = \frac{\text{Exp } \{ \log [X(\omega)] * L(\omega) \}}{\text{Exp } \{ \log [X_1(\omega)] * L(\omega) \}} \qquad (3.31)$$

onde $L(\omega)$ é a transformada de Fourier da janela cepstral $l(n)$. Sabe-se que o cepstrum complexo de uma sequência $x(n)$ é dado por:

$$\hat{x}(n) = \mathcal{F}^{-1}[\log[X(\omega)]] \tag{3.32}$$

Reescrevendo a equação anterior tem-se:

$$\hat{x}(n) = \mathcal{F}^{-1}[\ \ln|X(\omega)| + j\ \theta(\omega)] \tag{3.33}$$

onde $\theta(\omega)$ representa a fase do espectro $\log[X(\omega)]$ após a operação de *"Unwrapping" de Fase* descrita no ítem 2.5.1.

Até aqui ficou demonstrado que regiões de baixa energia no logaritmo do módulo do espectro de $H(\omega)$ exibem uma grande sensibilidade a desvios do modelo convolucional , produzindo erros elevados na estimativa final da resposta em frequência do filtro $H(z)$.

Este problema também aparece na estimativa da fase $\theta(\omega)$, pois o valor principal da fase depende do quociente entre as partes imaginária e real do espectro complexo de $H(\omega)$. Dessa maneira, a utilização de um janelamento adaptado ao período de Pitch é também necessária para se minimizar os erros devidos ao modelo convolucional em realizações de fase mista.

3.4.2.1 Os Efeitos da Fase Linear

Além dos erros devidos às regiões de baixa energia no espectro, a estimativa de fase obtida em realizações de fase mista é fortemente influenciada pela presença da fase linear n_0, a qual, como foi visto, inclui as informações de fase linear devidas ao sinal de excitação e ao filtro $H(z)$.

Ao se aplicar a operação de *"Unwrapping" de Fase* sobre o sinal $\theta(\omega)$, a sua fase linear é removida, pois isto contribui para uma melhor estimativa de fase do filtro $H(z)$ [4,5]. A parcela de fase linear presente em n_0, devida ao sinal de excitação, contribui em alguns casos com uma inclinação muito elevada no espectro de fase, o que dificulta a sua remoção pela operação de *Unwrapping* , provocando uma degradação na estimativa de fase do filtro $H(z)$. Este problema pode ser minimizado

realizando um alinhamento entre a janela **w(n)** e o sinal de excitação **p(n)**. Este alinhamento remove a componente mais significativa da inclinação linear facilitando a operação de *Unwrapping* e levando a uma estimativa de fase mais acurada.

Na prática porém, não se dispõe do sinal de excitação isoladamente e, assim, deve-se lançar mão de alguns artifícios para se tentar um alinhamento . No capítulo seguinte será descrita uma técnica que permite um razoável grau de alinhamento e que produz resultados bastante satisfatórios.

3.4.2.2 O Efeito do "Jitter" de Pitch

Como foi visto no ítem anterior, a operação de *"Unwrapping" de Fase* remove a componente linear de fase n_0 que corresponde à fase linear do filtro **H(z)**, supondo um perfeito alinhamento entre a janela e o sinal de excitação.

Do ponto de vista relativo ao processamento de um único quadro de voz , a remoção da fase linear não apresenta consequências mais graves. Infelizmente porém, a remoção da fase linear em quadros sucessivos provoca um desalinhamento entre respostas impulsivas de quadros adjacentes que distorce o sinal de voz reconstruído na recepção. Este desalinhamento é conhecido como *"Jitter" de Pitch* e é particularmente sentido quando tem-se uma resposta impulsiva de variação lenta entre quadros sucessivos. A figura 3.1 ilustra os efeitos provocados pelo *"Jitter de Pitch"* :

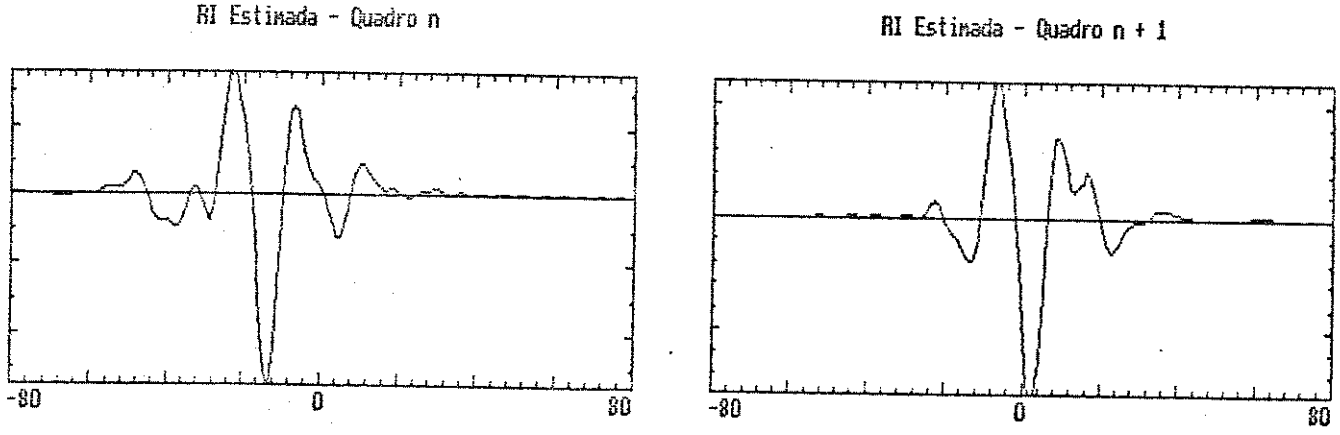


Figura 3.1 a)

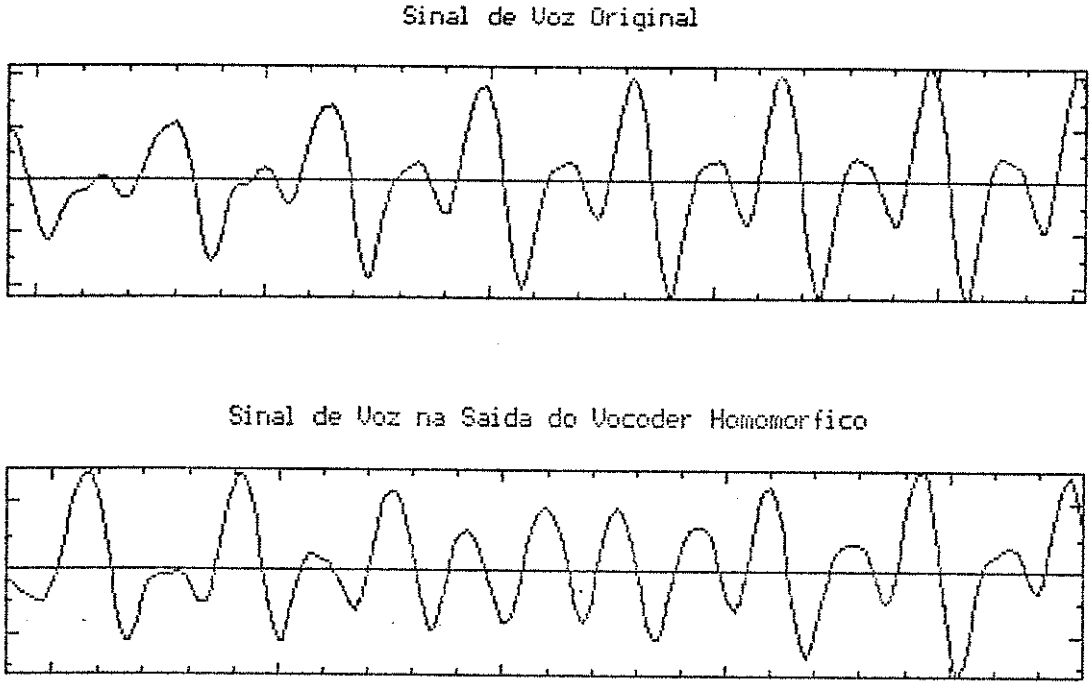


Figura 3.1 b)

Figura 3.1 - Efeito do Jitter de Pitch

A figura 3.1 a) mostra as respostas impulsivas de quadros adjacentes na presença do *"Jitter" de Pitch* . A figura 3.1.b) mostra a forma de onda original e o sinal de voz reconstruído sujeito ao *"Jitter" de Pitch.*

O efeito do *"Jitter" de Pitch* pode ser compensado através da reintegração da fase linear na estimativa do filtro H(z) já que a operação de *"Unwrapping" de Fase* retorna o parâmetro n_0. Este procedimento porém, exige a transmissão deste parâmetro para o receptor,o que demanda uma quantidade adicional de bits.

Um procedimento que dispensa a transmissão da informação de fase linear é a realização de uma correlação cruzada entre respostas impulsivas estimadas sucessivas de variação lenta. A correlação cruzada será máxima para um deslocamento igual ao oposto do deslocamento devido ao *"Jitter" de Pitch.*

3.4.3 EFEITOS DAS IMPERFEIÇÕES DO MODELO CONVOLUCIONAL NA DECONVOLUÇÃO HOMOMÓRFICA DE QUADROS NÃO-SONOROS

Em relação à Desconvolução Homomórfica de segmentos de som não-sonoros,

simulações práticas mostram não haver nenhum efeito relevante do comprimento da janela w(n) sobre a estimativa de $H(\omega)$ resultante, e que a aplicação do modelo convolucional não resulta em erros tão significativos quanto aqueles verificados para os segmentos de som sonoro.

3.5 REFERÊNCIAS

[1] A. V. Oppenheim and R. W. Schafer, *"Digital Signal Processing"*. Englewood Cliffs, NJ,Prentice - Hall, 1975.

[2] A. V. Oppenheim and R. W. Schafer, *" Homomorphic Analysis of Speech"*, IEEE Trans. Audio Electroacoust., vol AU-16 pp 221-226 June 1968.

[3] L. R. Rabiner and R. W. Schafer, *"Digital Processing of Speech Signals"*, Englewood Cliffs, NJ, Prentice-Hall, 1978.

[4] J. M. Tribolet, *" A New Phase Unwrapping Algorithm"*, IEEE Trans. Acoust. Speech, Signal Processing, Vol ASSP-25, pp 170-177, April 1977

[5] T. F. Quatieri,*"Minimun and Mixed Phase Speech Analysis-Synthesis by Adaptive Homomorphic Deconvolution"*, IEEE Trans. Acoust., Speech, Signal Processing, vol ASSP-27, pp 328-335, Aug. 1979.

CAPÍTULO 4

O VOCODER HOMOMÓRFICO

CONTEÚDO

4.1 INTRODUÇÃO

Baseado na análise do capítulo anterior, este capítulo traz uma descrição detalhada de um Vocoder Homomórfico simulado em computador. A partir de realizações de fase mínima e fase mista, são descritas as operações que compõem um sistema de codificação de voz a baixas taxas baseado na Desconvolução Homomórfica.

Este capítulo traz também uma descrição dos resultados obtidos com o Vocoder Homomórfico em testes subjetivos informais e testes subjetivos comparativos utilizando um Vocoder LPC tradicional.

4.2 Simulação de um Sistema Homomórfico para Análise-Síntese de Sinais de Voz - O Vocoder Homomórfico

No capítulo anterior foi descrito o procedimento que permite a separação dos efeitos do sinal de excitação e do filtro **H(z)** representativo do conjunto formado pelo Pulso Glótico, Trato Vocal e Impedância de Irradiação segundo o modelo tradicional de produção da voz. Esta separação, realizada a nível de cepstrum, proporciona uma representação do sinal de voz sob a forma de um conjunto de parâmetros que compreende um número finito de amostras do cepstrum da resposta impulsiva **h(n)** e o período de Pitch. Este conjunto de parâmetros pode ser transmitido, permitindo a reconstrução do sinal de voz original na recepção com alto grau de inteligibilidade. O sistema de codificação, transmissão e recepção de sinais de voz baseado na Desconvolução Homomórfica é denominado *Vocoder Homomórfico*[1].

Apesar de ser perfeitamente viável a obtenção do período de Pitch utilizando técnicas de processamento homomórfico [7], os sistemas aqui propostos empregam um procedimento baseado no *Algoritmo de Kurt-Schäfer Vincent*[5], o qual tem proporcionado excelentes resultados em outros vocoders [6]. Neste caso, compete à Desconvolução Homomórfica apenas a obtenção dos parâmetros do filtro **H(z)**.

Nesta simulação foram implementados dois tipos de Vocoders Homomórficos:
- um baseado numa Realização de Fase Mínima, que utiliza o cepstrum;
- outro baseado numa Realização de Fase Mista, que utiliza o cepstrum complexo.

Apesar de serem constituídos, basicamente, pelas mesmas operações, o Vocoder Homomórfico baseado no cepstrum complexo possui algumas funções específicas que devem ser analisadas separadamente.

4.2.1 - O VOCODER HOMOMÓRFICO BASEADO NUMA REALIZACÃO DE FASE MÍNIMA

Como visto no capítulo 2, a utilizacão do cepstrum complexo obtido a partir de uma Realização de Fase Mínima (ou simplesmente cepstrum), é perfeitamente viável quando a informação de fase é suposta irrelevante ou dispensável. Este é o caso quando se tratando de sinais de voz, pois o ouvido humano é pouco sensível às variações de fase. Neste ítem é descrito um Vocoder Homomórfico baseado numa Realização de Fase Mínima [11].

4.2.1.1 Detalhes da Configuração da Etapa de Análise - Transmissão

A etapa de análise consiste na obtenção do cepstrum complexo a partir de uma realização de fase mínima e uma caracterização do sinal de excitação através de uma decisão do tipo sonoro/não-sonoro, além do cálculo do período de Pitch para sons sonoros. A figura 4.1 ilustra o esquema utilizado:

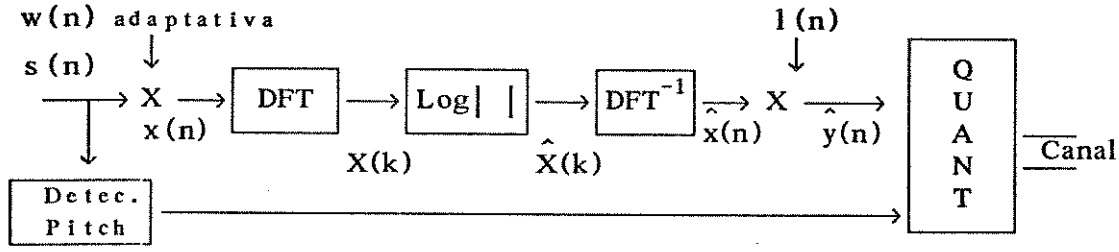


Figura 4.1 - Vocoder Homomórfico Baseado numa Realização de Fase Mínima : Análise - Transmissão

A seguir será dada uma descrição detalhada de cada uma das operações efetuadas .

a) Detecção de Pitch

A detecção do período de Pitch é realizada a partir do método desenvolvido por Kurt Schäfer-Vincent. [5]

b) Janelamento Adaptativo

O sinal de entrada é ponderado por uma janela de Hamming de comprimento efetivo igual a 2 (duas) vezes o período de Pitch para quadros sonoros e comprimento igual a 160 amostras para quadros não-sonoros. A análise do sinal de voz é feita a cada 160 amostras (cerca de 20 ms para uma taxa de amostragem de 8 KHz). Quando o comprimento da janela é inferior ao tamanho do quadro (160 amostras), a janela **w(n)** é do tipo :

$$w(n) = \begin{cases} 0.08 & ; 0 \le n < \frac{(L - 2N_o)}{2} \\ (0.54-0.46\cos\left[\frac{2\pi(n - (L - 2N_o)/2)}{(2N_o - 1)}\right] & ; \frac{(L - 2N_o)}{2} \le n < \frac{(L + 2N_o)}{2} \\ 0.08 & ; \frac{(L + 2N_o)}{2} \le n < L \end{cases} \quad (4.1)$$

onde: **L** é o comprimento do quadro - 160 amostras;

No é o período de Pitch.

Ou seja, a janela **w(n)** é centralizada no quadro sonoro e não se anula fora dos limites do janelamento. Daí o termo "comprimento efetivo". Este procedimento permite um aproveitamento um pouco maior da informação contida no segmento de voz sob análise.

Nos casos em que o comprimento da janela ultrapassa o comprimento do quadro, ou seja, na ocorrência de um Período de Pitch maior ou igual a 80, utiliza-se uma janela de Hamming convencional de comprimento igual a 160 amostras.

c) DFT e DFT^{-1}

O cálculo da Transformada Discreta de Fourier e da Transformada Discreta de Fourier Inversa é realizado através de um algoritimo de transformada rápida (FFT) de 512 pontos. Como foi discutido no ítem 2.4 o tamanho da DFT é de fundamental importância para se evitar o fenômeno do *aliasing* no cepstrum.

d) Janela Cepstral

A separação entre o cepstrum da excitação e o cepstrum da resposta impulsiva é realizada utilizando-se uma janela cepstral de 20 pontos, ou seja:

$$\begin{aligned} l(n) &= 1 \quad ; 0 \le n < 20 \\ &= 0 \quad ; n \ge 20 \end{aligned}$$

Simulações mostraram que este é o menor comprimento da janela cepstral que ainda

permite uma estimativa satifatória da resposta impulsiva para os sinais de teste utilizados.

Entretanto, quando o Período de Pitch for inferior a 20, a separação entre o cepstrum da excitação e o cepstrum da resposta impulsiva não é perfeita, levando a uma degradação da resposta impulsiva estimada. Felizmente, esta situação ocorre muito raramente.

4.2.1.2 DETALHES DA CONFIGURAÇÃO DA ETAPA DE SÍNTESE - RECEPÇÃO

A etapa de síntese consiste na convolução entre a resposta impulsiva, estimada a partir da análise cepstral, e o sinal de excitação, que é gerado com base na decisão sonoro/não-sonoro e na estimativa do período de pitch no caso de sons sonoros. A figura 4.2 ilustra o esquema utilizado :

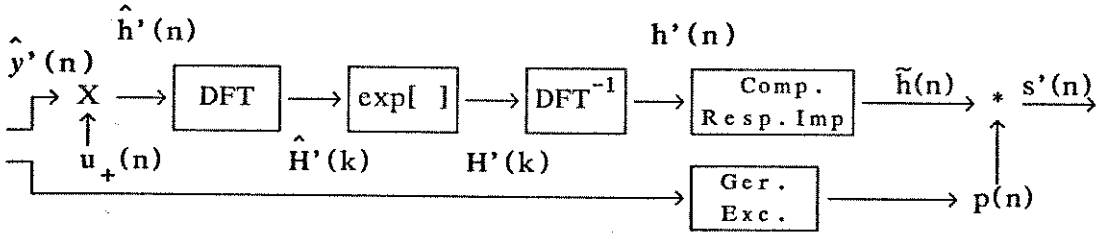


Figura 4.2 - Vocoder Homomórfico Baseado numa Realização de Fase Mínima : Síntese - Recepção

A seguir será dada uma descrição detalhada de cada uma das operações efetuadas.

a) Compensação de Amplitude da Resposta Impulsiva

Esta operação é o resultado de um novo procedimento, proposto neste trabalho, para se levar em consideração o efeito da amostra na origem do cepstrum da excitação sobre o cepstrum da resposta impulsiva [12].

Como foi discutido no ítem 3.3, a amostra do cepstrum da excitação que aparece somada ao cepstrum da resposta impulsiva, representa um fator de escala que será multiplicado à resposta impulsiva estimada.

No caso de uma excitação correspondente a um segmento de som sonoro, este fator de escala é menor que a unidade, pois é resultante de uma amostra da janela **w(n)** aplicada ao sinal de excitação. Dessa maneira a resposta impulsiva associada a um quadro de som sonoro aparecerá atenuada na saída da etapa de análise.

Para o caso de uma excitação correspondente a um quadro de som não-sonoro, este fator de escala é maior que a unidade, pois é resultante do efeito do janelamento sobre a variância do ruído branco. Assim, a resposta impulsiva associada a um quadro de som não-sonoro aparecerá amplificada na saída da etapa de análise.

Dois aspectos, porém, devem ser ressaltados:

- O impulso inicial de uma excitação correspondente a um quadro sonoro não ocorre sempre na mesma posição em relação à janela w(n). Isto faz com que o valor da atenuação sofrida pela resposta impulsiva seja distinta para cada quadro de som sonoro. A figura 4.3 ilustra este efeito sobre dois quadros de uma excitação sonora com período de Pitch igual a 50 e janela de *Hamming* de comprimento igual a 200.

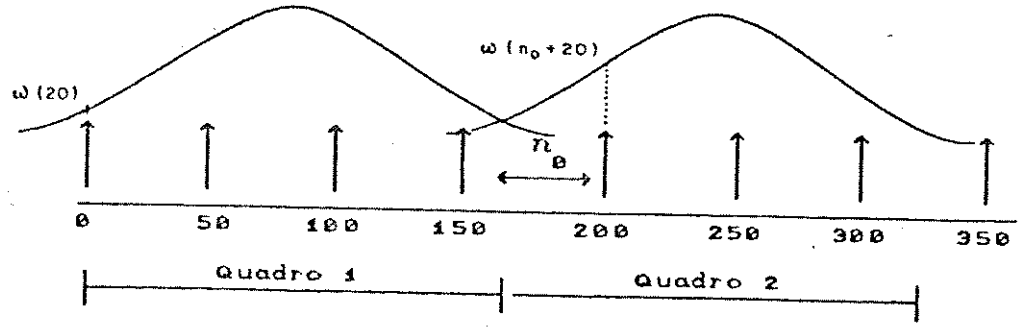


Figura 4.3 - Efeito da posição relativa entre a janela e o sinal de excitação no fator de escala para h'(n)

Vê-se claramente que a amplitude do primeiro impulso de excitação no quadro 1 será diferente daquela correspondente no quadro 2, uma vez que a posição relativa entre a janela w(n) e os impulsos associados muda a cada quadro. Assim o valor de A na expressão (3.10) varia a cada quadro;

- O ruído branco possui uma variância unitária, podendo no entanto, sofrer alterações de uma realização para outra, isto é, de um quadro para outro. Assim, o

valor do ganho aplicado à resposta impulsiva também é distinto para cada quadro de som não-sonoro;

Valores típicos para a relação entre os fatores de escala correspondente a quadros não-sonoro e sonoro, considerando variância unitária para a excitação não-sonora, estão entre 20 e 50, o que leva à conclusão de que este efeito deve ser compensado sob pena de haver uma significativa distorção no sinal de voz na saída do sistema.

Felizmente, as faixas de variação destas grandezas permitem estimar valores médios de compromisso para que se possa compensar estes efeitos. Dessa maneira, as respostas impulsivas relativas à análise cepstral de quadros de som sonoro são multiplicadas por um fator igual a 1,5 e as respostas impulsivas resultantes da análise cepstral de quadros de som não-sonoro são multiplicadas por um fator igual a 0,3. Estes valores são resultado de simulações com arquivos de voz feminina e masculina, usando ruído branco de variância 1/8 na excitação não-sonora (Ver ítem 4.2.1.2 b). Utilizando ruído branco de variância unitária, o fator correspondente aos quadros não-sonoros seria de 0.3/8.

A compensação de amplitude da resposta impulsiva por quadros permite a aplicação de um tratamento diferenciado aos segmentos de voz situados em regiões de transição entre sons sonoros e não-sonoros.

Os quadros de som sonoro ocorrendo imediatamente antes ou imediatamente depois de quadros de som não-sonoro, representam regiões de transição para as quais os dois efeitos aparecem combinados de alguma maneira. No sistema proposto é utilizada uma ponderação de 0.8 e de 0.6 (ao invés de 1.5) para as respostas impulsivas resultantes da análise cepstral de quadros de som sonoro imediatamente antes e imediatamente depois de sequências de quadros de som não-sonoro, respectivamente. Estes valores foram obtidos a partir de inúmeras simulações práticas e a diferença entre eles é um fato bastante razoável quando se verifica que quadros de som sonoro imediatamente depois de segmentos de som não-sonoro devem sofrer uma influência maior do efeito relativo aos quadros de som não-sonoro, ou seja, um ganho na resposta impulsiva estimada, o que justifica o menor fator de ponderação utilizado.

A compensação dos efeitos da amostra na origem do cepstrum da excitação sobre a estimativa da resposta impulsiva é também realizada em outros Vocoders Homomórficos [2,4], porém através de fatores de escala diferenciados aplicados aos sinais de excitação sonoro e não-sonoro. A origem destes fatores de escala não é relacionada à questão da superposição dos cepstruns na origem, sendo atribuída a

problemas de balanço energético entre sons sonoro e não-sonoro. A abordagem aqui proposta, além de ser fisicamente consistente, permitiu inferir a possibilidade de um tratamento específico para os quadros de transição, o que não é realizado nos outros Vocoders Homomórficos analisados.

b) Geração do Sinal de Excitação - p(n)

Com base na decisão sonoro/não-sonoro e na estimativa do período de Pitch, é gerado o sinal de excitação correspondente à cada quadro de voz.

O sinal de excitação correspondente a um quadro de som sonoro é um trem de impulsos unitários com espaçamento dado pelo período de Pitch.

O sinal de excitação correspondente a um quadro de som não-sonoro é um trem de impulsos unitários com polaridade aleatória e espaçados no tempo de uma quantidade maior que a unidade para se obter uma redução do trabalho computacional na convolução entre a resposta impulsiva e a excitação. Neste caso é adotado um espaçamento de 8 amostras, o que corresponde a 5% do tamanho do quadro. Este procedimento resultou também numa redução da variância do ruído utilizado e condicionou o valor do fator de compensação dos quadros não-sonoros.

c) Convolução

A convolução entre o sinal de excitação e a resposta impulsiva de cada quadro é realizada reproduzindo-se a resposta impulsiva a cada pulso de excitação. Com o objetivo de se obter uma redução do esforço computacional, a resposta impulsiva relativa a cada segmento de voz é truncada em 160 amostras, ou seja, o tamanho de um quadro, o que exige um atraso na convolução igual a um quadro para que se possa levar em consideração o efeito da "cauda" da resposta impulsiva na passagem de um quadro para o quadro seguinte.

Na convolução é também realizada uma interpolação progressiva entre as respostas impulsivas do quadro atual e a do quadro seguinte [2], ou seja, a resposta impulsiva reproduzida por um impulso da excitação no instante **k** dentro do quadro atual é computada como sendo :

$$h(k) = p(k).\ [\ h_1(k) + [(\ h_2(k) - h_1(k))\ k/160]] \qquad (4.2)$$

onde:

- **p(k)** é a amostra do sinal de excitação ocorrendo no instante **k** do quadro atual;
- $h_1(k)$ é a resposta impulsiva estimada para o quadro atual;
- $h_2(k)$ é a resposta impulsiva estimada para o quadro seguinte;
- 160 é o tamanho do quadro.

Este procedimento serve para garantir uma variação mais suave para a resposta impulsiva entre um quadro e o quadro seguinte, o que é de fundamental importância quando se utiliza o janelamento adaptativo, como ressaltado anteriormente.

4.2.1.3 Detalhes da Quantização dos Parâmetros de Transmissão

Após a etapa de análise, cada quadro de voz é representado por um conjunto de parâmetros que compreende as 20 amostras do cepstrum da resposta impulsiva do trato vocal e o período de Pitch. Estes parâmetros devem ser quantizados para que possam ser transmitidos em um meio digital.

Considerando uma faixa de variação típica entre 20 e 100, o período de Pitch é quantizado utilizando-se 7 bits com uma lei de quantização do tipo linear.

A amostra inicial do cepstrum representa o logaritmo do ganho do filtro H(z) em cada quadro, e assim foi quantizada de maneira particular utilizando-se 6 bits ,sendo 1 bit de sinal, com uma lei de quantização do tipo linear.

Para a quantização das outras 19 amostras do cepstrum foi utilizada uma quantização do tipo BC-PCM (Block Companded PCM) [8] com 6 bits para a informação lateral, correspondente à amplitude da maior amostra (em módulo) dentre as 19, e 4 bits para as amostras . Para garantir uma faixa de excursão das amplitudes das amostras do cepstrum que facilite a quantização, a função $u_+(n)$ é introduzida na recepção e não na transmissão (ver figura 4.2).

Considerando 1 bit de sincronismo, a taxa de transmissão resultante é da ordem de 4,8 kbits/s.

4.2.1.4. - Uma Versão Alternativa [10]

Na seção 3.4.1 foi demostrado que a Desconvolução Homomórfica quando aplicada a quadros de som sonoro, resulta numa estimativa para a resposta em frequência $H(\omega)$ dada por:

$$H'(\omega) = H(\omega).\ E_h(\omega)$$

onde $E_h(\omega)$ é o erro entre a estimativa , $H'(\omega)$, e a resposta ideal, $H(\omega)$.

Dessa maneira, o sinal de voz reconstruído na recepção é dado por:

$$S'(\omega) = H'(\omega)\ .\ P(\omega)$$
$$S'(\omega) = H(\omega)\ .\ E_h(\omega).\ P(\omega)$$

onde $P(\omega)$ é o espectro do sinal de excitação.

Simulações práticas mostram que a distorção na resposta em frequência $H(\omega)$, dada pelo sinal de erro, $E_h(\omega)$, é caracterizada principalmente por uma enfatização das componentes de alta frequência de $H(\omega)$. A figura 4.4 mostra o efeito do sinal de erro na estimativa da resposta em frequência do trato vocal, $H(\omega)$, para um quadro de som sonoro com período de Pitch igual a 38 e comprimento da janela w(n) igual a 200 amostras.

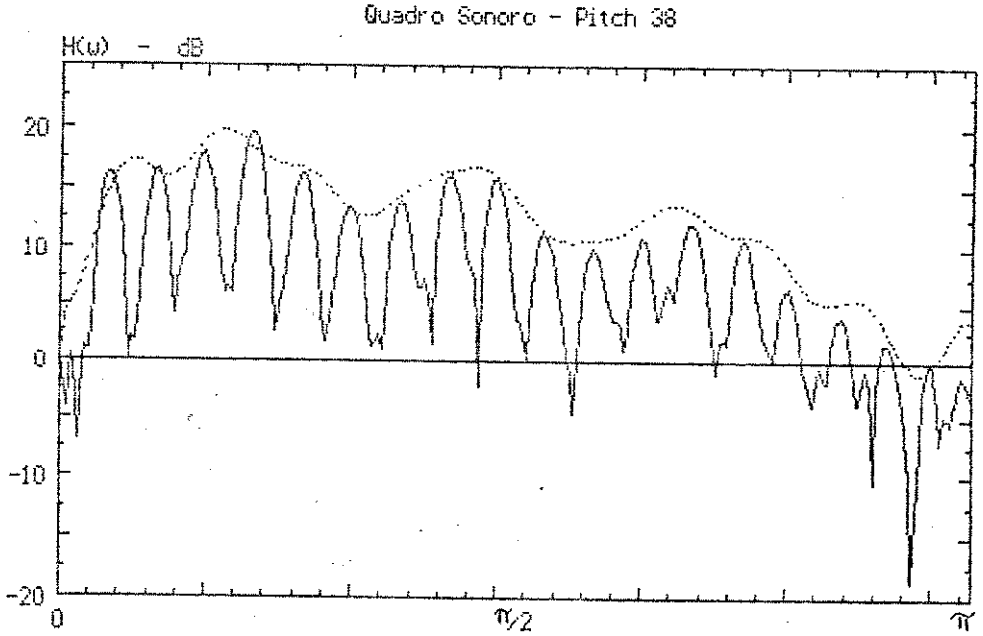


Figura 4.4 - Efeito do erro $E_h(\omega)$ na estimativa da resposta em frequência, $H(\omega)$ para um quadro de som sonoro com Pitch 38

Na figura 4.4 a curva cheia representa o espectro de Fourier do sinal de voz janelado. A curva pontilhada é a estimativa para a resposta em frequência do trato vocal obtida através da Desconvolução Homomórfica. Considerando que a resposta em frequência ideal seria representada por uma envoltória imaginária passando pelos picos do espectro periódico, nota-se claramente a enfatização das componentes de alta frequência na envoltória estimada pela Desconvolução Homomórfica.

Com o objetivo de explorar esta evidência prática, foi implementada uma versão alternativa do sistema proposto:

- O janelamento de Hamming adaptativo é substituído por uma janela de Hamming de comprimento fixo igual a 160 amostras.

- A mudança no janelamento do sinal de entrada provoca uma alteração dos valores de ponderação para a operação de compensação de amplitude da resposta impulsiva:

Quadros Sonoros:	2.0
Quadros Não-Sonoros:	0.4
Quadros Sonoros Antes de Segmentos Não-Sonoros:	1.1
Quadros Sonoros Após Segmentos Não-Sonoros:	0.9

- Para compensar o efeito do sinal de erro $E_h(\omega)$, é realizada , na etapa de síntese, uma operação de de-ênfase nas componentes de alta frequência das estimativas $H'(\omega)$ dos quadros de som sonoro [10]. Esta operação procura modelar, na média, o efeito inverso do sinal erro $E_h(\omega)$:

$$H(\omega) \cong H'(\omega) \cdot D(\omega)$$

onde $D(\omega) \cong 1/E_h(\omega)$.

A operação de de-ênfase é realizada para quadros de som sonoro com período de Pitch menor que 70. Este número foi definido levando-se em consideração os valores típicos encontrados para o período de Pitch (entre 20 e 100) e o comprimento da janela w(n) utilizada (160 amostras). Uma análise destes valores mostra que quadros de som sonoro com período de Pitch acima de 70 se aproximam da condição ótima que produz um erro mínimo (comprimento da janela = 2 x período de Pitch), e assim a operação de de-ênfase poderia corresponder à introdução de uma distorção adicional.

A utilização de um filtro de de-ênfase, além de ser mais simples que o uso

do janelamento adaptativo discutido no ítem 3.3.1, proporciona um total aproveitamento da informação contida no quadro de som sonoro . Isto permite, portanto, supor que a estimativa para a resposta impulsiva do trato vocal será mais representativa do segmento de voz sob análise, e resultará em variações menos bruscas entre as respostas impulsivas estimadas de quadros adjacentes.

O filtro de de-ênfase utilizado é do tipo :

$$D(z) = \frac{0.6}{1 - 0.4z^{-1}}$$

Esta configuração foi obtida após inúmeros testes auditivos.

A figura 4.5 ilustra a etapa de síntese desta versão alternativa do Vocoder Homomórfico.

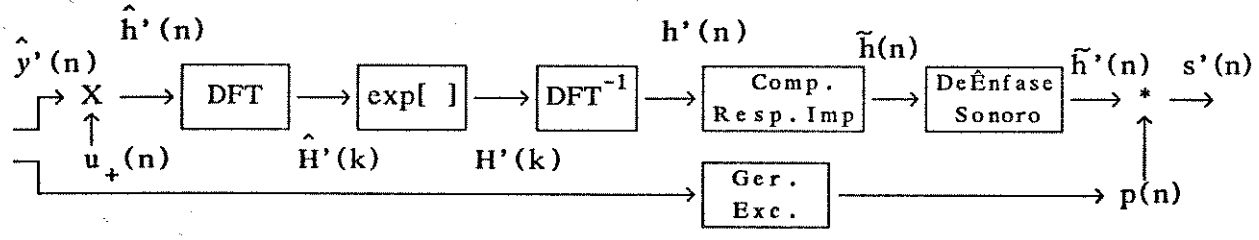


Figura 4.5 - Vocoder Homomórfico Baseado numa Realização de Fase Mínima (Versão Alternativa): Síntese

4.2.2 - O VOCODER HOMOMÓRFICO BASEADO NUMA REALIZACÃO DE FASE MISTA

O sinal de voz resultante na saída do Vocoder Homomórfico, baseado numa Realização de Fase Mínima, é diferente quanto à forma de onda em relação ao sinal original. Como já discutido anteriormente , a Realização de Fase Mínima garante que o espectro de amplitude do sinal obtido na saída do sistema seja idêntico ao espectro de amplitude do sinal original. A perda da informação de fase, embora altere a forma de onda, não é relevante já que se considera que o ouvido humano é pouco sensível às variações de fase.

Utilizando um Vocoder Homomórfico baseado numa Realização de Fase Mista e cepstrum complexo, é possível levar a informação de fase do sinal original até a

saída do sistema. Dessa forma é esperado um sinal de voz sintetizado cuja forma de onda reproduza de maneira bastante satisfatória a forma de onda do sinal original. Este ítem descreve este sistema, a partir do qual é possível avaliar a real importância da informação de fase para o sistema auditivo humano.

4.2.2.1 Detalhes da Configuração da Etapa de Análise - Transmissão

Em linhas gerais, a etapa de análise é muito semelhante àquela do Vocoder Homomórfico de Fase Mínima. As diferenças principais estão localizadas na operação do Logaritmo Complexo que inclui o *"Unwrapping" de Fase* e na janela cepstral que neste caso, é de 40 pontos, já que o cepstrum complexo é não-causal (Ítem 2.4).

Outra diferença significativa é a realização do alinhamento da janela adaptativa w(n). Como discutido anteriormente, o alinhamento é de fundamental importância para garantir uma boa estimativa de fase resultante da Desconvolução Homomórfica, pois remove a parcela mais significativa da inclinação linear , a qual é devida ao sinal de excitação.

Como não se dispõe do sinal de excitação isoladamente, é utilizado um artifício para se obter o melhor alinhamento com a janela. O procedimento consiste em detectar o pico do sinal de voz dentre as primeiras 2 N_0 amostras do quadro (N_0 é o Período de Pitch). A posição de referência para o alinhamento da janela é marcada como sendo o ponto de máximo determinado anteriormente menos $N_0/10$ amostras [3]. Com isto o alinhamento é feito sempre em relação ao segundo impulso de excitação e a janela tem um posicionamento variável dentro do quadro. Neste caso , dependendo do período de Pitch, o janelamento pode envolver amostras do quadro atual e do quadro seguinte.

Infelizmente, este procedimento não garante o perfeito alinhamento entre a janela de entrada e o sinal de excitação . Este fato será responsável por erros na estimativa de fase resultante da Desconvolução Homomórfica.

A figura 4.6 ilustra a etapa de análise do Vocoder Homomórfico de Fase Mista.

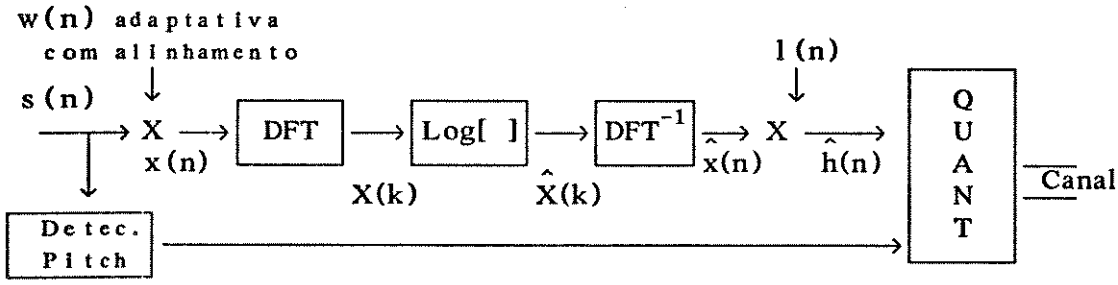


Figura 4.6 - Vocoder Homomórfico Baseado numa Realização de Fase Mista : Análise - Transmissão

O janelamento é realizado utilizando-se a janela de Hamming convencional. A operação de alinhamento anula o efeito do maior aproveitamento da informação obtido com a janela dada pela equação 4.1. Na verdade, dependendo da posição de referência para o alinhamento e do Período de Pitch, a utilização desta janela pode até mesmo levar a uma degradação dos resultados obtidos com a Desconvolução Homomórfica de Fase Mínima para o quadro de voz sob análise.

4.2.2.2 - Detalhes da Configuração da Etapa de Síntese - Recepção

Como ocorrido com a etapa de análise, a etapa de síntese é basicamente idêntica à etapa de síntese do Vocoder Homomórfico de Fase Mínima. A única diferença entre os dois sistemas está na presença da operação de Compensação de Jitter de Pitch de Quadros Sonoros. Simulações mostraram que a operação de alinhamento, apesar de alterar a evolução do posicionamento relativo entre a janela de entrada w(n) e o sinal de excitação, não provoca mudanças significativas nos valores dos fatores de compensação de amplitude das respostas impulsivas estimadas utilizados nos quadros sonoro , não-sonoro e de transição. Seu principal efeito é uma redução da faixa de variação do ganho ou da atenuação sofrida pelas respostas impulsivas dos quadros não-sonoro e sonoro, respectivamente.

A figura 4.7 ilustra a etapa de síntese do Vocoder Homomórfico de Fase Mista.

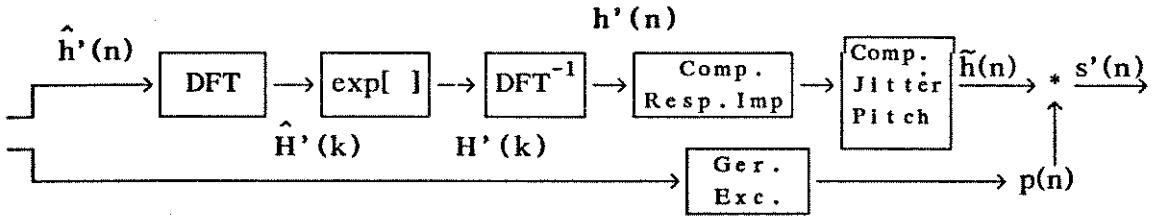


Figura 4.7 - Vocoder Homomórfico Baseado numa Realização de Fase Mista : Síntese - Recepção

De acordo com a discussão do ítem 3.4.2.1, o efeito do Jitter de Pitch pode ser compensado de duas maneiras : através da reintegração da fase linear estimada pela operação de "*Unwrapping*" na estimativa da resposta impulsiva do filtro **H(z)**; ou através do cálculo do pico de correlação cruzada entre respostas impulsivas sucessivas de variacão lenta.

A utilização da informação de fase linear resultante da operação de "*Unwrapping*" é a opção mais simples de ser implementada, porém exige um aumento da taxa de transmissão resultante do sistema, já que esta informação deve ser enviada ao receptor juntamente com o Período de Pitch e as amostras do cepstrum complexo. Este , a princípio, parece ser um preço bastante razoável a ser pago para se obter a virtual eliminação do problema do Jitter de Pitch. Entretanto, este resultado só é obtido se houver o perfeito alinhamento entre a janela de entrada adaptativa **w(n)** e o sinal de excitação e ,então, a informação de fase linear retornada pela operação de "*Unwrapping*" for devida exclusivamente ao filtro **H(z)**.

Como foi visto no ítem anterior, a operação de alinhamento está sujeita a erros e, na prática, é mais razoável supor que a informação de fase linear retornada pela operação de "*Unwrapping*" inclui uma componente devida ao sinal de excitação, a qual pode variar de um quadro para outro. Como o sinal de excitação é gerado independentemente da estimativa da resposta impulsiva, o sinal de voz sintetizado na saída do sistema apresentará o desalinhamento entre quadros sucessivos característico do Jitter de Pitch.

A utilização do pico de correlação cruzada, apesar de dispensar o envio de informação adicional, é uma opção bem mais complexa. Em primeiro lugar só faz sentido a pesquisa pelo pico de correlação cruzada entre respostas impulsivas de quadros adjacentes com variação lenta, pois o valor de pico para a correlação cruzada entre respostas impulsivas altamente descorrelacionadas normalmente não é equivalente ao deslocamento devido ao Jitter de Pitch.

Este problema requer a utilização de um limiar de correlação. Este limiar

serve para avaliar o nível de correlação entre as respostas impulsivas adjacentes e determina a convêniencia de se utilizar a informação de correlação cruzada na compensação do Jitter de Pitch. Ainda assim, vale ressaltar que a utilização de um limiar de correlação não evita falsas indicações e isto deve ser levado em conta na escolha de seu valor nominal.

Outro ponto importante que deve ser analisado é o fato de que a correlação cruzada funciona, neste caso, como uma medida do grau de identidade entre respostas impulsivas adjacentes visando a caracterização de uma variação lenta. Na prática, outras grandezas podem ser utilizadas para este mesmo fim produzindo, em alguns casos, resultados mais significativos. Assim foram testadas 3 tipos de medidas :

- O valor máximo da correlação cruzada propriamente dita :

$$r_{h1h2}(k) = \sum_{n} h_1(n)h_2(n - k)$$

- O valor mínimo da diferença absoluta entre respostas impulsivas adjacentes:

$$r_{h1h2}(k) = \sum_{n} | h_1(n) - h_2(n - k) |$$

- O valor mínimo da diferença quadrática entre respostas impulsivas adjacentes:

$$r_{h1h2}(k) = \sum_{n} [h_1(n) - h_2(n - k)]^2$$

Para garantir uma uniformidade dos valores calculados, as respostas impulsivas são previamente normalizadas.

Os testes mostraram que o valor mínimo do quadrado da diferença normalizada é a medida mais consistente e a partir da qual é possível escolher o limiar de compensação de Jitter de Pitch mais adequado.

O Vocoder Homomórfico de Fase Mista foi implementado em duas versões: uma utilizando a reintegração da fase linear estimada pela operação de *"Unwrapping"* ; e a outra utilizando o valor mínimo da diferença quadrática entre respostas impulsivas normalizadas de variação lenta.

4.2.2.3 - Detalhes da Quantização dos Parâmetros de Transmissão

O procedimento para quantização dos parâmetros de transmissão foi idêntico àquele utilizado no Vocoder Homomórfico de Fase Mínima.

Na simulação da versão que realiza a compensação de jitter de Pitch via reintegração da fase linear estimada pelo *"Unwrapping"*, utilizou-se 7 bits para quantizar a fase com uma lei de quantização do tipo linear.

Considerando que o número de amostras do cepstrum complexo é, neste caso, o dobro daquele utilizado no Vocoder de Fase Mínima, tem-se uma taxa final de transmissão de 9,15 kbits/s para a versão com reintegração da fase linear estimada pelo *"Unwrapping"* e 8,8 kbits/s para a versão com detecção do valor mínimo da diferença quadrática entre respostas impulsivas adjacentes normalizadas.

4.3 Medidas de Desempenho do Vocoder Homomórfico

Todas as versões do Vocoder Homomórfico descritas no ítem anterior foram simuladas usando a linguagem TURBO C™ num microcomputador compatível com o IBM-PC™ acoplado ao sistema SAPDV-A (Sistema de Análise e Processamento Digital de Voz) [9]. Este sistema permite a gravação e reprodução de arquivos de voz digitalizados. As simulações operam em tempo não-real.

O desempenho dos sistemas foi avaliado usando testes subjetivos informais. Estes testes foram realizados com um conjunto de 5 ouvintes de ambos os sexos. Os testes consistiram da avaliação de arquivos de voz feminina e masculina processados pelos Vocoders Homomórficos com e sem quantização dos parâmetros de transmissão, usando como referência os arquivos de voz originais, amostrados a 8 kHz e submetidos ao conversor A/D linear de 12 bits por amostra do SAPDV-A.

4.3.1 - RESULTADOS DOS TESTES SUBJETIVOS INFORMAIS

4.3.1.1 - O Vocoder Homomórfico de Fase Mínima

O Vocoder Homomórfico de Fase Mínima produziu um sinal de voz sintetizada de elevada inteligibilidade e boa qualidade com um grau de naturalidade bastante satisfatório. A versão utilizando o janelamento adaptado ao Período de Pitch

apresentou um desempenho superior ao desempenho obtido com a versão que realiza a de-ênfase nas altas frequências para sons sonoros.

4.3.1.2 - O Vocoder Homomórfico de Fase Mista

O desempenho do Vocoder Homomórfico de Fase Mista utilizando a técnica da compensação de jitter de Pitch através do valor mínimo da diferença quadrática entre as respostas impulsivas adjacentes foi inferior àquele obtido com o Vocoder Homomórfico de Fase Mínima. A versão utilizando a técnica da reintegração da fase linear estimada pelo *"Unwrapping"*, por sua vez, apresentou um desempenho superior para determinados arquivos de voz e inferior para outros.

4.3.1.3 - Desempenho dos Vocoders Homomórficos com Quantização

Os testes realizados com todas as versões do Vocoder Homomórfico para processamento com quantização demonstraram a elevada tolerância deste tipo de sistema à quantização dos parâmetros de transmissão, já que todos os resultados obtidos nos testes subjetivos informais para processamento sem quantização foram também obtidos nos testes realizados com quantização.

4.3.1.4 - Desempenho dos Vocoders a Nível de Forma de Onda

Para ilustrar os resultados descritos anteriormente são reproduzidas, a seguir, as formas de onda originais e sintetizadas pelos Vocoders Homomórficos referentes a um trecho de voz feminina. Nestas figuras utiliza-se a seguinte convenção :

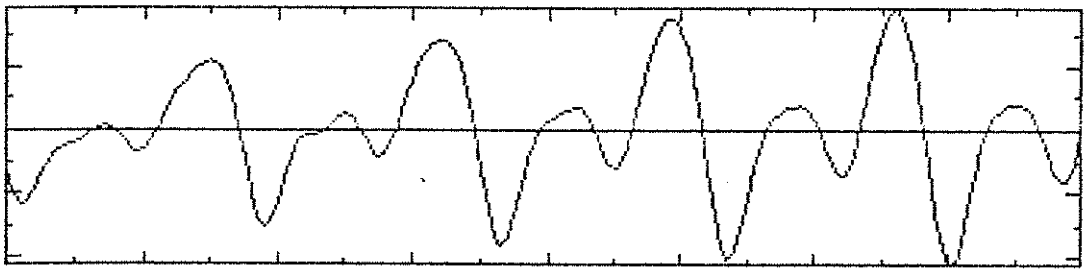
Vocoder Homomórfico de Fase Mínima :

- Versão 1 : Utiliza o janelamento adaptado ao Período de Pitch.
- Versão 2 : Utiliza a de-ênfase de altas frequências para sons sonoros.

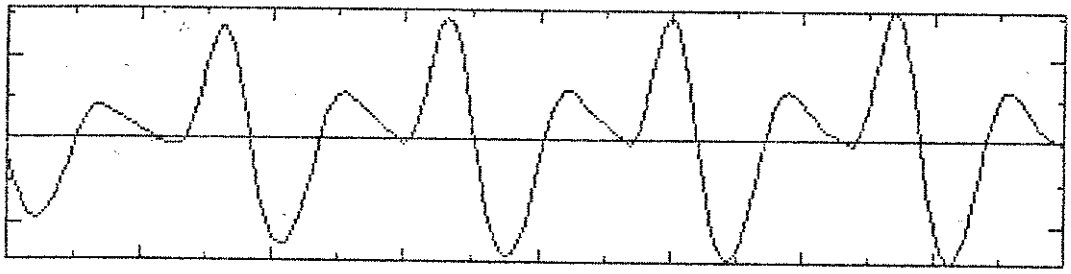
Vocoder Homomórfico de Fase Mista :

- Versão 1 : Utiliza a técnica de reintegração da fase linear estimada pelo "*Unwrapping*".
- Versão 2 : Utiliza o valor mínimo da diferença quadrática entre respostas impulsivas adjacentes normalizadas

a) Sinal Original



b) Sinal na Saída do Vocoder Homomórfico de Fase Mínima - Versão 1



c) Sinal na Saída do Vocoder Homomórfico de Fase Mínima - Versão 2

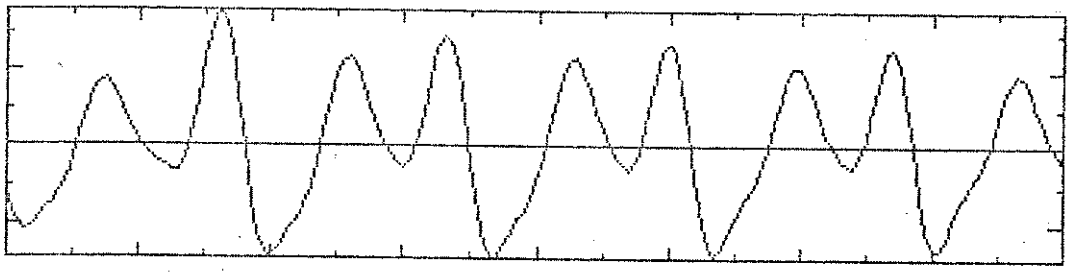
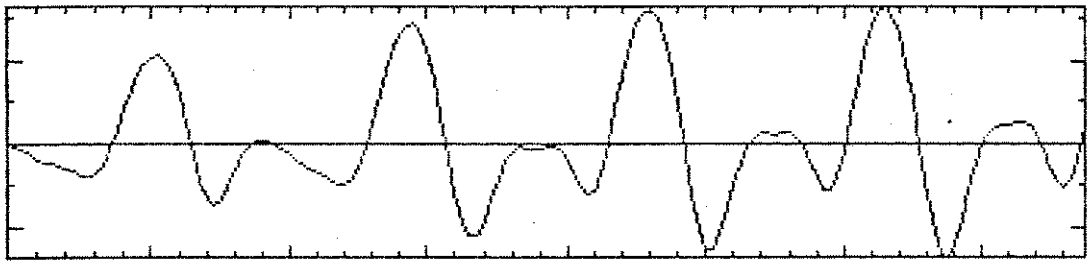
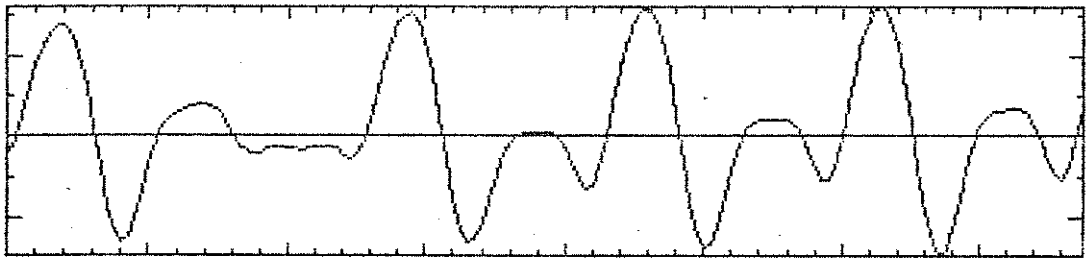


Figura 4.8 - a) Sinal Original ; b) Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mínima : Versão 1; c) Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mínima : Versão 2; d) Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mista : Versão 1; e) Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mista : Versão 2; f) Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mínima : Versão 1 c/ Quantização ; g) Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mista : Versão 1 c/ Quantização

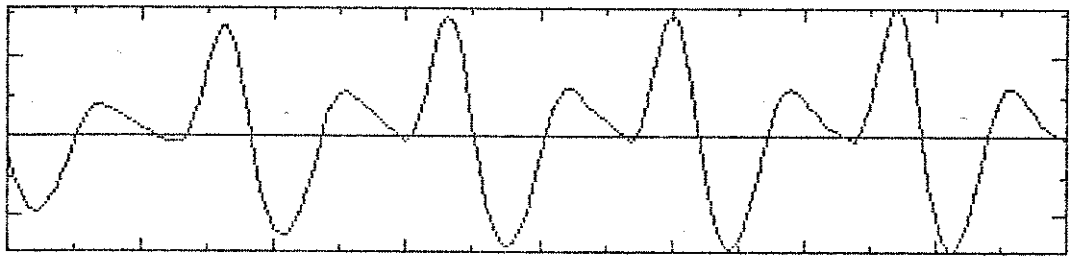
d) Sinal na Saída do Vocoder Homomórfico de Fase Mista - Versão 1



e) Sinal na Saída do Vocoder Homomórfico de Fase Mista - Versão 2



f) Sinal na Saída do Vocoder Homomórfico de Fase Mínima - Versão 1 com Quantização



g) Sinal na Saída do Vocoder Homomórfico de Fase Mista - Versão 1 com Quantização

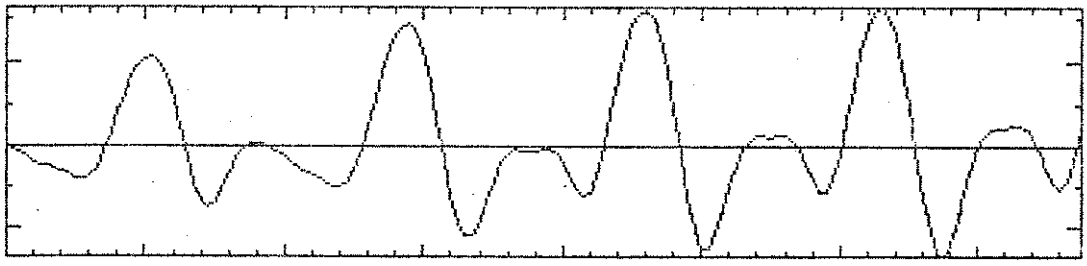


Figura 4.8 - Continuação

4.3.2 DISCUSSÃO DOS RESULTADOS

Os resultados obtidos nos testes subjetivos demonstraram que o Vocoder Homomórfico produz um sinal de voz sintetizado com elevado grau de inteligibilidade e de alta qualidade.

Em relação ao Vocoder Homomórfico de Fase Mínima, o desempenho superior da versão que utiliza o janelamento adaptado ao Período de Pitch, comparativamente à versão com de-ênfase de altas frequências para sons sonoros, pode ser explicado levando-se em consideração que um filtro de de-ênfase fixo compensa da mesma forma diferentes graus de distorção, o que provoca um desempenho significativamente inferior em alguns quadros de voz, em relação ao uso da janela adaptativa. Apesar disto, a de-ênfase para sons sonoros, de maneira geral, reduziu de forma marcante as distorções devidas ao sinal de erro $E_h(\omega)$.

Uma análise individual dos resultados obtidos com o Vocoder Homomórfico de Fase Mista mostra que, apesar de dispensar uma taxa adicional de bits, o uso do valor mínimo da diferença quadrática entre as respostas impulsivas adjacentes normalizadas para obter o deslocamento necessário para compensar o Jitter de Pitch , não traz resultados satisfatórios. Os erros cometidos na estimativa deste deslocamento afetam bastante a síntese correta do sinal de voz no receptor. Na figura 4.8 e) vê-se o efeito presente na voz sintetizada devido a um erro na compensação do Jitter de Pitch.

O uso da fase linear estimada pelo *"Unwrapping"* garante um grau de acerto mais elevado , permitindo a obtenção de um desempenho mais satisfatório. Entretanto, o seu desempenho inferior para alguns arquivos de voz, quando comparado ao desempenho do Vocoder Homomórfico de Fase Mínima, mostra que o nível de erro ainda possível é bastante significativo na qualidade da voz sintetizada na saída do sistema. Este erro, como discutido no ítem 4.2.2.2, é devido principalmente a um alinhamento imperfeito entre a excitação e a janela de entrada w(n). Estes resultados demonstram que o problema do Jitter de Pitch é bastante crítico para o Vocoder Homomórfico de Fase Mista [12].

4.4 - COMPARAÇÕES COM O VOCODER LPC

Usando os arquivos de voz originais como referência foram realizados

testes subjetivos para verificar o desempenho do Vocoder Homomórfico comparativamente a um Vocoder LPC baseado no modelo tradicional de produção de voz [1]. O Vocoder LPC utiliza um preditor de ordem 8 , modelagem AR pelo método da Autocorrelação e síntese do sinal de voz na recepção através da equação de diferenças definida pelos coeficientes do modelo AR referente a cada um dos quadros.

Os testes subjetivos mostraram que o Vocoder Homomórfico de Fase Mínima apresenta um desempenho superior ao Vocoder LPC convencional, principalmente para arquivos de voz masculina.

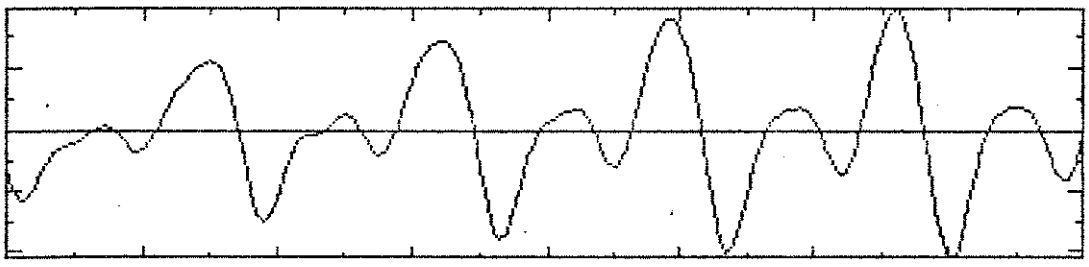
Em relação ao Vocoder Homomórfico de Fase Mista, a versão com compensação de jitter de Pitch via reintegração da fase linear estimada pelo "*Unwrapping*", apresentou um desempenho superior, enquanto que a versão que utiliza o valor mínimo do quadrado da diferença entre as respostas impulsivas adjacentes normalizadas não conseguiu superar o desempenho obtido com o Vocoder LPC convencional.

Como discutido no início deste trabalho, um dos objetivos deste estudo é confrontar a técnica da Desconvolução Homomórfica com a técnica da Análise LPC quando aplicadas na obtenção de um conjunto de parâmetros característicos do filtro digital **H(z)**, buscando um melhor desempenho do modelo tradicional na reprodução de sinais de voz. Conforme a descrição das características operacionais do Vocoder Homomórfico, a presença de uma interpolação entre as respostas impulsivas estimadas adjacentes na etapa de síntese representa uma vantagem adicional a qual, apesar de ser uma exigência da técnica da Desconvolução Homomórfica (Ver ítem 3.4), pode perfeitamente ser aplicada no Vocoder LPC tradicional. Este procedimento além de permitir uma avaliação da real influência da interpolação no melhor desempenho apresentado pelo Vocoder Homomórfico em relação ao Vocoder LPC, garante uma comparação mais imparcial entre as duas técnicas.

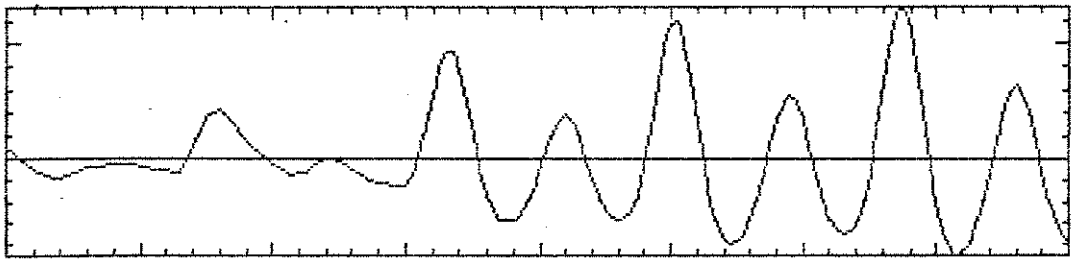
Neste sentido, os testes comparativos, após a modificação no Vocoder LPC, não apresentaram alterações significativas em relação àqueles descritos anteriomente. Dessa maneira, pode-se concluir que a utilização da técnica da Desconvolução Homomórfica, quando comparada a Análise LPC, efetivamente resulta num conjunto de parâmetros que melhor representam as características espectrais do filtro **H(z)** e proporcionam um desempenho superior do modelo tradicional na reprodução de sinais de voz.

A seguir são reproduzidas as formas de onda sintetizadas na saída dos dois sistemas para um mesmo quadro de voz.

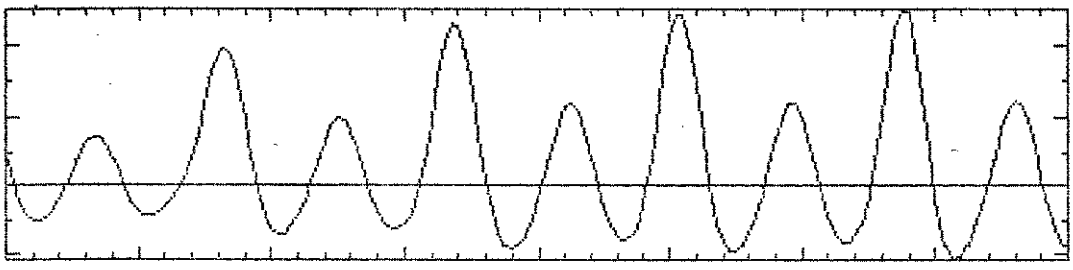
a) Sinal Original



b) Sinal de Voz na Saída do Vocoder LPC Convencional



c) Sinal de Voz na Saída do Vocoder LPC com Interpolação Progressiva entre Quadros



d) Sinal na Saída do Vocoder Homomórfico de Fase Mínima - Versão 1

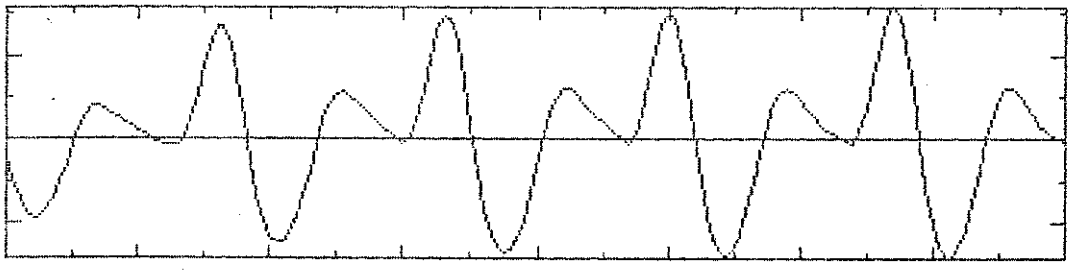


Figura 4.9 - a) Sinal Original ; b) Sinal de Voz na Saída do Vocoder LPC Convencional; c) Sinal de voz na Saída do Vocoder LPC com Interpolação Progressiva Entre Quadros ; d) Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mínima - Versão 1; e)Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mista - Versão 1.

e) Sinal na Saída do Vocoder Homomórfico de Fase Mista - Versão 1

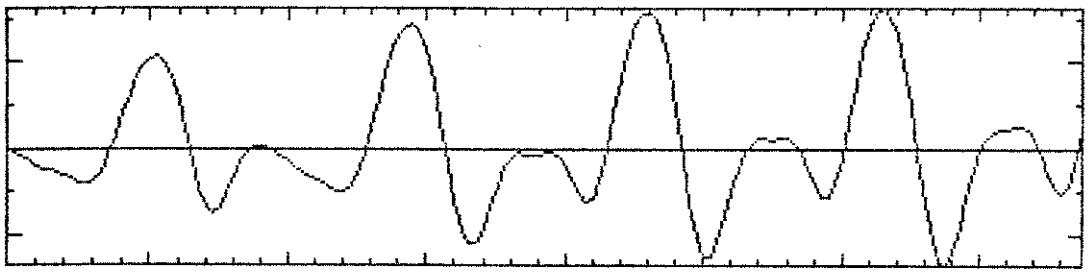


Figura 4.9 - Continuação

4.5 - REFERÊNCIAS

[1] L. R. Rabiner and R. W. Schafer, *"Digital Processing of Speech Signals"*, Englewood Cliffs, NJ, Prentice-Hall, 1978.

[2] A. V. Oppenheim, *" Speech Analysis-Synthesis System based on Homomorphic Filtering"*, J. Acoust. Soc. Amer., vol 45, pp 459- 462 1969.

[3] T. F. Quatieri,*"Minimun and Mixed Phase Speech Analysis-Synthesis by Adaptive Homomorphic Deconvolution"*, IEEE Trans. Acoust., Speech, Signal Processing, vol ASSP-27, pp 328-335, Aug. 1979.

[4] C. R. Patisaul and J. C. Hammett Jr.,*"Time Frequency Resolution Experiment in Speech Analysis and Synthesis"*, J. Acoust. Soc. Amer., vol 58, pp 1296-1307.

[5] Kurt Schäffer-Vincent, *"Pitch Period Detection and Chaining : Method and Evaluation"*, Phonetica 40, 1983, pp 177-202

[6] Fábio Violaro e José Antonio Martins, *"Low Bit Rate LPC Vocoders Using Vector Quantization and Interpolation"*, ICASSP 91, Toronto Ontário ,Canadá .

[7] A. M. Noll, *"Cepstrum Pitch Determination"*, The Journal of the Acoustical Society of America, Vol 41 Nº 2 pp 293-309, 1967

[8] P. Noll, *"Adaptive Quantizing in Speech Coding System"* , Proc. 1974 Zürich Seminar on Digital Communications, Zürich, March 1974 pp B3(1)-B3(6).

[9] Fábio Violaro, *"Nova Versão do Sistema de Análise e Processamento Digital de Voz SAPDV-A*, 7º Simp. Brasileiro de Telecomunicações Florianópolis, SC, Brasil, 1989, pp 50-53.

[10] P. C. D. Oliveira, Amauri Lopes e Fábio Violaro, *"Um Vocoder Baseado na Deconvolução Homomórfica"*, Anais do 9º Simp. Brasileiro de Telecomunicações - USP , São Paulo - SP, 1991.

[11] P. C. D. Oliveira, Amauri Lopes e Fábio Violaro, *"Nova Versão de um Vocoder Baseado na Deconvolução Homomórfica"*, Anais da IV Reunion del Trabajo en Procesamiento de la Informacion y Control - CNEA - Buenos Aires , Argentina, 1991

[12] P. C. D. Oliveira, Amauri Lopes e Fábio Violaro, " *Minimun and Mixed Homomorphic Vocoders : An analisys of the Design Problems"*, Submetido à comissão organizadora do EUSIPCO - 92.

CAPÍTULO 5

A PREDIÇÃO HOMOMÓRFICA

CONTEÚDO

5.1 INTRODUÇÃO

Este capítulo traz uma discussão sôbre a técnica da Predição Homomórfica que combina a Desconvolução Homomórfica com a Análise Preditiva Linear Generalizada. O objetivo principal deste estudo é buscar uma alternativa para uma redução da taxa de transmissão final do Vocoder Homomórfico de Fase Mista.

Este capítulo traz também uma investigação dos efeitos da Predição Homomórfica utilizando a Análise LPC convencional. O objetivo deste estudo é avaliar o desempenho da Análise LPC quando aplicada diretamente sôbre a resposta impulsiva do filtro H(z).

5.2 - Fundamentos da Predição Homomórfica

Como discutido no capítulo 3, o objetivo da representação paramétrica do sinal de voz é obter um conjunto de parâmetros que caracterize o filtro digital H(z) do modelo tradicional de produção de voz. Na Análise LPC convencional, supõe-se que este filtro seja do tipo racional e formado apenas por pólos. Esta limitação é notada principalmente na reprodução de sons nasalizados, onde a ausência dos zeros espectrais compromete a qualidade da voz sintetizda [3].

Algumas generalizações da Análise LPC tem sido propostas para a realização do modelagem do filtro H(z) com pólos e zeros (Modelo ARMA). Ocorre porém, que muito poucas destas técnicas podem ser aplicadas diretamente a segmentos contendo vários períodos de voz, devido à dificuldade que tais métodos enfrentam para distinguir entre os zeros do filtro H(z) e os os zeros introduzidos pela excitação. Dessa maneira, as técnicas de modelagem ARMA, de um modo geral, requerem algum tipo de deconvolução do sinal de voz. Um procedimento que pode ser utilizado é a Análise com Sincronismo de Pitch, na qual períodos de som sonoro são extraídos e analisados individualmente. Esta técnica envolve uma precisa localização de cada pulso de excitação, sem a qual os resultados obtidos não são muito satisfatórios.

Uma alternativa é a utilização da Desconvolução Homomórfica, a qual permite obter a resposta impulsiva do filtro H(z) sem apresentar o problema do sincronismo de Pitch.

Uma vez determinada a resposta impulsiva, pode-se utilizar qualquer um dos métodos conhecidos de modelagem ARMA e assim, através da transmissão dos coeficientes do numerador e denominador, reconstruir o sinal de voz no receptor.

Este procedimento que combina a Desconvolução Homomórfica baseada em Realizações de Fase Mínima ou de Fase Mista com a Análise Preditiva Linear Generalizada para modelagem ARMA, é conhecido como *Predição Homomórfica* [1,2].

5.3 - Aplicação da Predição Homomórfica na Redução da Taxa de Transmissão do Vocoder Homomórfico de Fase Mista

Conforme o item 4.3, a utilização da informação de fase na estimativa da resposta impulsiva do filtro H(z) pode trazer uma melhoria no desempenho do Vocoder

Homomórfico. Entretanto, o Vocoder Homomórfico de Fase Mista é baseado no cálculo do cepstrum complexo o qual, como visto no item 2.4, é não-causal. Dessa forma, a taxa de transmissão resultante é praticamente o dobro daquela obtida para o Vocoder Homomórfico de Fase Mínima, supondo um mesmo procedimento de quantização. O nível de melhoria de desempenho obtido com a versão de fase mista não compensa este aumento da taxa de transmissão e a maior complexidade computacional.

O Vocoder Homomórfico de Fase Mista proposto no ítem 4.2 utiliza 40 amostras do cepstrum complexo. A aplicação da técnica da Predição Homomórfica baseada num modelagem ARMA com cerca de uma dezena de pólos e uma dezena de zeros (ARMA(10,10)), poderia resultar numa redução significativa do número de parâmetros a serem transmitidos para o receptor e, consequentemente, da taxa de transmissão final do sistema.

Ocorre porém, que a não causalidade da resposta impulsiva exige modelos ARMA distintos para as porções causal e anti-causal da resposta impulsiva estimada. Além disso, neste caso estamos interessados num modelagem que reproduza o mais fielmente possível a resposta impulsiva e não o módulo da função de transferência do filtro **H(z)**. Dessa maneira, as técnicas de modelagem ARMA mais apropriadas para esta situação se baseiam em equações de erro envolvendo a resposta impulsiva [5]. Neste caso, a ordem do numerador (número de zeros) determina a quantidade de amostras da resposta impulsiva do modelo que reproduzirá de forma exata a resposta impulsiva desejada. Assim, é comum utilizar-se modelos ARMA com a ordem do numerador igual à ordem do denominador.

O procedimento utilizado inclui o modelagem ARMA da porção causal ($n \geq 0$) da resposta impulsiva estimada e o modelagem ARMA da porção não-causal ($n < 0$), após sua reversão no domínio do tempo.

Baseado na discussão anterior, a utilização da técnica de Predição Homomórfica, a princípio, não resultaria em uma redução significativa da taxa de transmissão final do Vocoder Homomórfico de Fase Mista. Entretanto, simulações mostram que existe um elevado nível de redundância entre os modelos ARMA das porções causal e anti-causal da resposta impulsiva estimada a partir do cepstrum complexo. Esta redundância está localizada principalmente nos pólos dos modelos próximos à circunferência de raio unitário (correspondentes aos formantes da voz)[4].

A figura 5.1 mostra a localização dos pólos do modelo causal, representados por "x" e os pólos do modelo anti-causal, representados por "+", referentes a dois quadros de voz.

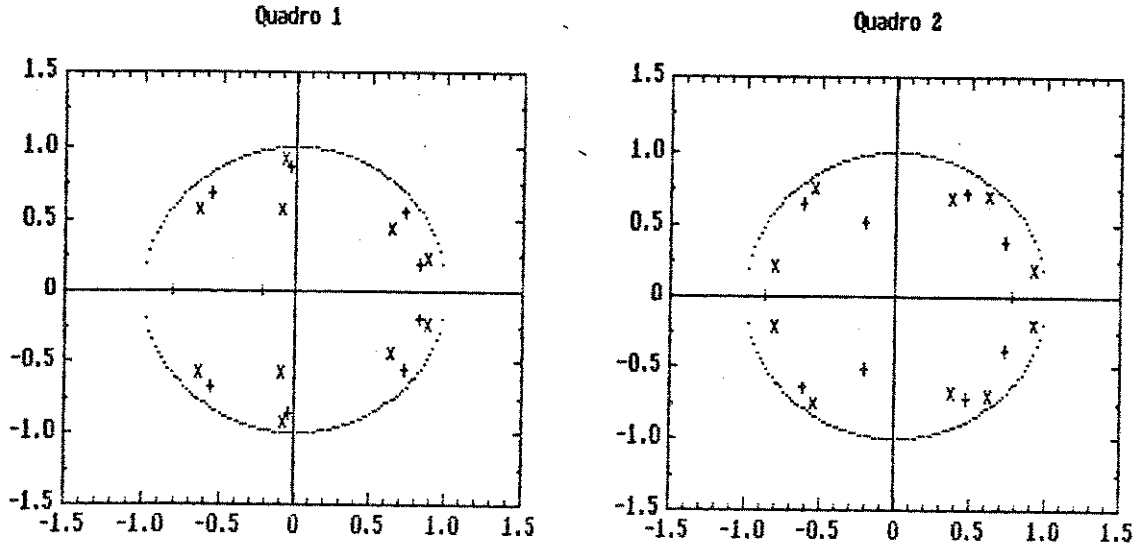


Figura 5.1

Nota-se que os pólos situados próximos à circunferência de raio unitário são muito similares para ambos os modelos. Os pólos restantes e as condições iniciais (ou zeros), as quais não são mostradas, são menos similares.

Esta redundância pode ser explorada para reduzir significativamente a taxa de transmissão final do Vocoder Homomórfico de Fase Mista.

5.3.1 SIMULAÇÃO DE UM VOCODER BASEADO NA PREDIÇÃO HOMOMÓRFICA DE FASE MISTA.

A técnica da Predição Homomórfica foi testada através da simulação de um Vocoder Homomórfico de Fase Mista empregando um modelagem ARMA. As porções causal e anti-causal da resposta impulsiva estimada foram representadas por modelos distintos, porém com mesma ordem. Os modelos utilizados foram do tipo ARMA(12,12), ARMA(10,10), ARMA(8,8) e ARMA(8,4).

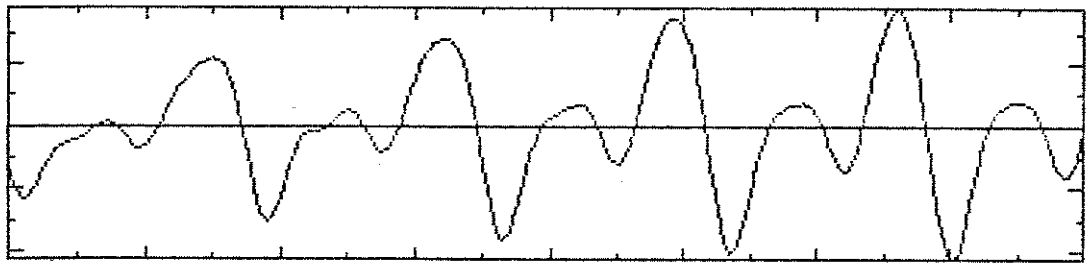
O desempenho destes sistemas foi avaliado através de testes subjetivos informais e os resultados obtidos mostraram que para os modelagens ARMA(8,8) e ARMA(8,4) o desempenho foi bastante inferior àquele do Vocoder Homomórfico de Fase Mista convencional. Um desempenho bastante satisfatório foi obtido utilizando modelos ARMA(10,10) e ARMA(12,12) , com uma leve superioridade para a versão com modelagem ARMA(12,12).

Vale ressaltar que os métodos de modelagem ARMA através das equações de erro padrão [5] não garantem a estabilidade do filtro digital obtido. Assim, alguns

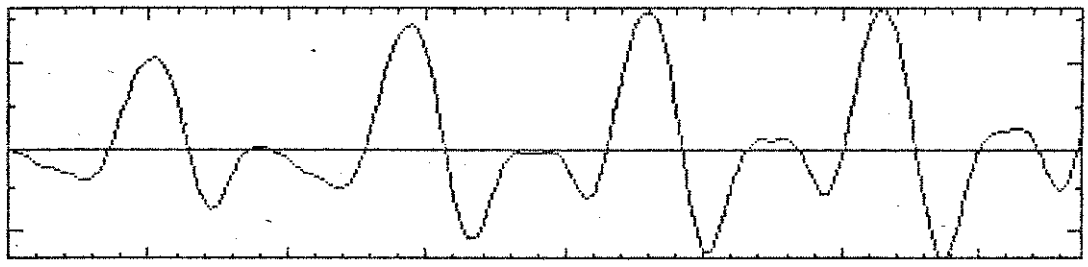
quadros de voz apresentaram problemas de divergência da resposta impulsiva estimada, comprometendo a qualidade do sinal de voz sintetizado. Nestes casos o sistema estava preparado para não realizar o modelagem e transmitir as amostras do cepstrum complexo, como no Vocoder Homomórfico de Fase Mista convencional.

Para ilustrar estes resultados, a figura 5.2 apresenta as formas de onda sintetizadas pelos Vocoders Homomórfico de Fase Mista, com e sem modelagem ARMA.

a) Sinal Original



b) Sinal na Saída do Vocoder Homomórfico de Fase Mista - Versão 1



c)Sinal na Saída do Vocoder com Predição Homomórfica - ARMA(12,12)

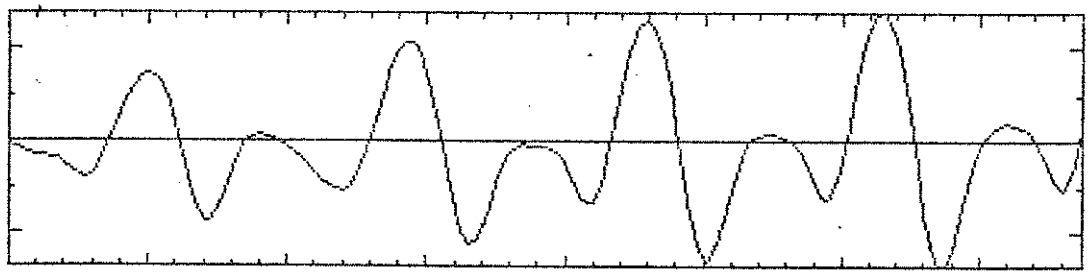
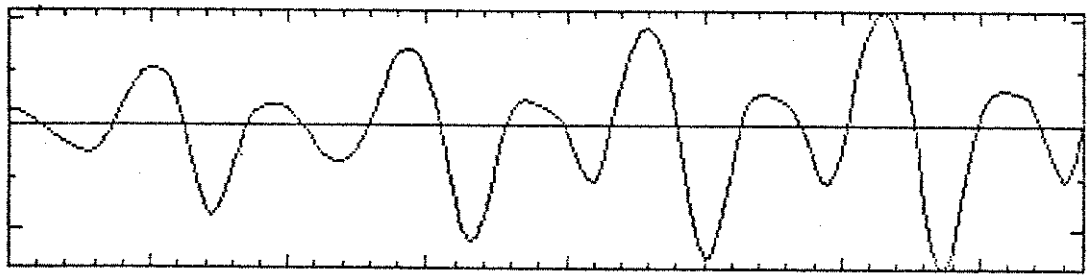
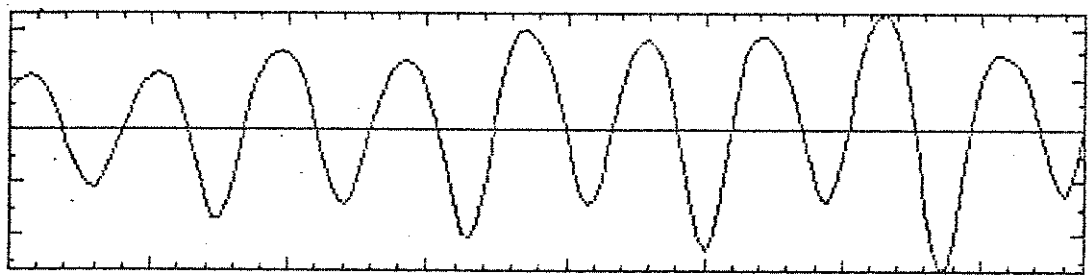


Figura 5.2 a) Sinal Original ; b) Sinal de Voz na Saída do Vocoder Homomórfico de Fase Mista - Versão 1 ; c) Sinal de Voz na Saída do Vocoder Baseado na Predição Homomórfica - ARMA (12,12) ; d) Sinal de Voz na Saída do Vocoder Baseado na Predição Homomórfica - ARMA(10,10); e) Sinal de Voz na Saída do Vocoder Baseado na Predição Homomórfica - ARMA(8,8); f) Sinal de Voz na Saída do Vocoder Baseado na Predição Homomórfica - ARMA (8,4).

d)Sinal na Saída do Vocoder com Predição Homomórfica - ARMA(10,10)



e)Sinal na Saída do Vocoder com Predição Homomórfica - ARMA(8,8)



f)Sinal na Saída do Vocoder com Predição Homomórfica - ARMA(8,4)

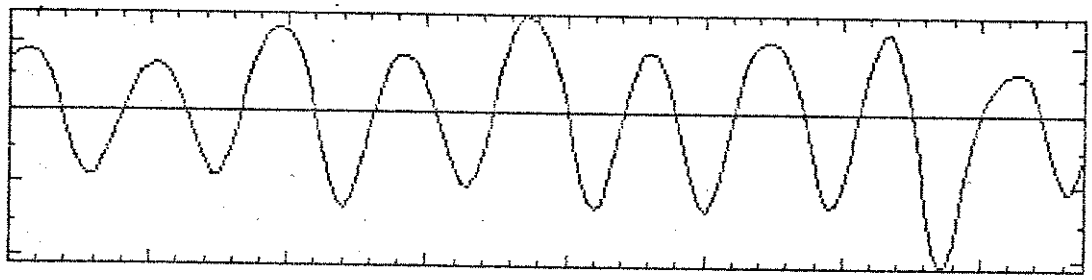


Figura 5.2 - Continuação

5.3.1.1 - Avaliação do Desempenho

Uma análise dos resultados obtidos nos testes subjetivos demonstram que o uso de um modelo ARMA(10,10) é a opção mais razoável para a implementação de um Vocoder baseado na Predição Homomórfica. O número de parâmetros característicos do filtro H(z) é, neste caso, igual ao número de amostras do cepstrum complexo transmitidas no Sistema Homomórfico de Fase Mista convencional. Além disso,

considerando a exploração da elevada redundância entre os pólos dos modelos das porções causal e anti-causal da resposta impulsiva, a taxa de transmissão resultante pode ser inferior àquela obtida com o Vocoder Homomórfico de Fase Mista convencional.

O nível de melhoria no desempenho obtido com um modelo ARMA(12,12) não justifica o aumento do número de parâmetros.Da mesma forma, o fraco desempenho dos Vocoders baseados nos modelos ARMA(8,8) e ARMA(8,4) inviabiliza a utilização destes sistemas.

Finalmente, vale ressaltar que a implementação de um Vocoder baseado na Predição Homomórfica requer o uso de algum procedimento para previnir os problemas de instabilidade que foram verificados em todas as versões simuladas.

5.4 - Combinando a Deconvolução Homomórfica com a Análise LPC Convencional

A realização de um sistema utilizando a técnica da Desconvolução Homomórfica baseada numa realização de Fase Mínima, combinada com uma Análise LPC convencional da resposta impulsiva estimada para o filtro H(z) e transmissão dos coeficientes do modelo AR obtido, é totalmente incoerente. A melhor representação das características espectrais do filtro H(z), obtida com a transmissão do cepstrum da resposta impulsiva , como realizado no Vocoder Homomórfico tradicional, seria anulada com a execução de um modelagem AR. Além disso, a complexidade computacional adicional de um sistema desta natureza o torna inviável.

Vale ressaltar porém, que a Análise LPC convencional é geralmente aplicada diretamente sôbre o sinal de voz [3], como realizado no Vocoder LPC . Assim, um sistema com Desconvolução Homomórfica e modelagem AR da resposta impulsiva estimada, permite estudar os efeitos da Análise LPC convencional quando aplicada diretamente sôbre a resposta impulsiva do filtro H(z).

Supondo um formato racional utilizando apenas pólos, o filtro H(z) será do tipo :

$$H(z) = \frac{G}{1 - \sum_{k=1}^{P} a_k z^{-k}} \qquad (5.1)$$

onde : **P** é a ordem do modelo. Dessa maneira, os parâmetros básicos do filtro **H(z)** a serem calculados são o ganho **G** e os coeficientes { a_k }. No domínio do tempo, a equação (5.1) pode ser escrita como:

$$h(n) = \sum_{k=1}^{P} a_k \, h(n-k) + G \, \delta(n) \tag{5.2}$$

Seja **P(z)** um preditor linear com coeficientes α_k definido como:

$$P(z) = \sum_{k=1}^{P} \alpha_k \, z^{-k} \tag{5.3}$$

Este preditor pode ser utilizado para realizar uma estimativa da resposta impulsiva **h(n)** do tipo :

$$\tilde{h}(n) = \sum_{k=1}^{P} \alpha_k \, h(n-k) \tag{5.4}$$

O erro de predição **e(n)** é definido como:

$$e(n) = h(n) - \tilde{h}(n) = h(n) - \sum_{k=1}^{P} \alpha_k \, h(n-k) \tag{5.5}$$

No domínio da frequência tem-se:

$$E(z) = H(z) \left[1 - \sum_{k=1}^{P} \alpha_k \, z^{-k} \right] \tag{5.6}$$

$$E(z) = H(z) \, . \, A(z) \tag{5.7}$$

onde $A(z) = 1 - P(z)$.

Comparando as equações (5.2) e (5.5) vê-se que se o sinal de voz obedece ao modelo da equação (5.2) e se $\alpha_k = a_k$, então $e(n) = G.\delta(n)$. Assim, o filtro de erro de predição **A(z)** será o filtro inverso ao sistema **H(z)**, a menos de um ganho **G**, ou seja:

$$H(z) = \frac{G}{A(z)} \tag{5.8}$$

Na análise LPC convencional, a determinação do conjunto de coeficientes α_k's é realizada diretamente a partir do sinal de voz [3]. Combinando a Desconvolução Homomórfica com a Análise LPC convencional, o conjunto de coeficientes α_k's pode ser obtido com base na resposta impulsiva **h(n)** estimada pela Desconvolução Homomórfica.

A abordagem utilizada consiste na busca dos coeficientes do preditor que minimizem o valor da soma dos erros quadráticos de predição dado pela equação (5.5). Os coeficientes assim obtidos são os parâmetros do filtro **H(z)** no modelo de produção da voz. Utilizando o critério do mínimo valor da soma dos erros quadráticos, obtém-se um conjunto de equações lineares que podem ser eficientemente resolvidas.

Dessa maneira, tem-se:

$$\overline{E}_h = \sum_n e^2(n) \tag{5.9}$$

$$= \sum_n [\, h(n) - \tilde{h}(n) \,]^2$$

$$= \sum_n [\, h(n) - \sum_{k=1}^{P} \alpha_k \, h(n-k) \,]^2 \tag{5.10}$$

Os limites dos somatórios nas equações (5.9) e (5.10) não serão especificados por enquanto, mas fica evidente que estes limites devem definir um somatório finito.

Nota-se também que foi omitida a constante que divide o somatório para se obter o valor médio $\overline{E}_h$. Isto foi feito porque esta constante se cancela para o conjunto de equações obtido.

Para calcular os coeficientes α_i's faz-se:

$$\frac{\partial \, \overline{E}_h}{\partial \, \alpha_i} = 0 \qquad \text{p/ } i = 1,2,.....P$$

Obtém-se então o seguinte conjunto de **P** equações a **P** incógnitas:

$$\sum_n h(n-i)\, h(n) = \sum_{k=1}^{P} \alpha_k \sum_n h(n-i)h(n-k) \qquad \text{p/ } 1 \le i \le P \tag{5.11}$$

Definindo:

$$\phi_h(i,k) = \sum_n h(n - i)h(n - k) \qquad \begin{matrix} p/\ i = 0,1,...P\ e \\ k = 0,1,...P \end{matrix} \tag{5.12}$$

pode-se escrever:

$$\phi_h(i,0) = \sum_{k=1}^{P} \alpha_k\ \phi_h(i,k) \qquad p/\ i = 1,2,...P \tag{5.13}$$

Este grupo de equações pode ser resolvido de maneira eficiente para se obter os α_k's ótimos para o preditor.

O erro mínimo para os coeficientes ótimos dados pelas equações (5.11) - (5.13) será:

$$\overline{E}_{h\ min} = \sum_n h^2(n) - \sum_{k=1}^{P} \alpha_{k\ ótimo} \sum_n h(n)\ h(n - k) \tag{5.14}$$

ou

$$\overline{E}_{h\ min} = \phi_h(0,0) - \sum_{k=1}^{P} \alpha_{k\ ótimo}\ \phi_h(0,k) \tag{5.15}$$

A solução das equações (5.13) requer inicialmente o cálculo dos valores de $\phi_h(i,k)$ para $1 \leq i \leq P$ e $0 \leq k \leq P$. Deve-se então definir claramente os limites dos somatórios nas equações (5.11) e (5.12).

5.4.1 O MÉTODO DA AUTOCORRELACÃO

Pelo método da Autocorrelação assume-se que a resposta impulsiva é nula fora do intervalo $0 \leq n \leq N - 1$. Isto pode ser expresso como:

$$h'(n) = h(n)\ .\ w(n) \tag{5.16}$$

onde w(n) é uma janela retangular de comprimento finito. Levando-se em conta que a Desconvolução Homomórfica é feita a partir de uma realização de fase mínima, baseada

no cálculo do cepstrum e não do cepstrum complexo, conclui-se que a resposta estimada pela Desconvolução Homomórfica é causal. Além disso, considerando que h(n) representa a resposta impulsiva de um filtro estável, sua amplitude decai com o tempo, tornando-se desprezível para um valor *n* maior que um certo N. Dessa maneira, a hipótese representada na equação (5.16) é perfeitamente válida.

O erro de predição e(n), para um preditor de ordem P, será, então, não-nulo no intervalo $0 \leq n \leq N - 1 + P$. Assim :

$$\overline{E}_h = \sum_{n=0}^{N+P-1} e^2(n) \qquad (5.17)$$

Neste ponto vale ressaltar uma importante diferença entre a Análise LPC tradicional e a Desconvolução Homomórfica com modelagem AR. Como na Análise LPC tradicional trabalha-se diretamente com o sinal de voz, é necessária a utilização de um janelamento w(n) do tipo *Hamming* ou *Hanning*, os quais minimizam os efeitos dos maiores erros de predição verificados nas extremidades do segmento de voz sob análise. A Desconvolução Homomórfica de Fase Mínima fornece uma resposta impulsiva h(n) do tipo causal e com decaimento rápido. Dessa maneira, a utilização de qualquer tipo de janela diferente da janela retangular traz consequências bastante desastrosas, já que o modelo resultante refletirá uma resposta impulsiva diferente da resposta estimada pela Deconvolucão Homomórfica.

Levando-se em conta a equação (5.16), tem-se:

$$\phi_h(i,k) = \sum_{n=0}^{N-1+P} h(n - i)\, h(n - k) \qquad 0 \leq i \leq P \text{ e } 0 \leq k \leq P \qquad (5.18)$$

Como $h(n) = 0$ fora do intervalo $0 \leq n \leq N - 1$, tem-se:

$$\phi_h(i,k) = \sum_{n=i}^{N-1+P} h(n - i)\, h(n - k)$$

Seja $n' = n - i$:

$$\phi_h(i,k) = \sum_{n'=0}^{N-1+P-i} h(n')\, h(n' + i - k)$$

Voltando à variável n e lembrando que:

$$h(n + i - k) = 0 \qquad p/\ n + i - k > N - 1$$

tem-se :

$$\phi_h(i,k) = \sum_{n=0}^{N-1-(i-k)} h(n)\ h(n + i - k) \tag{5.19}$$

Neste caso, a menos de um fator de normalização, $\phi_h(i,k)$ é igual autocorrelação do sinal h(n) calculada para um atraso de (i - k):

$$\phi_h(i\ ,\ k) = R_h(i - k) \tag{5.20}$$

Desde que $R_n(k)$ é uma função par, tem-se :

$$\phi_h(i\ ,\ k) = R_h(|i - k|) \quad p/\ \ i = 0,1....P \text{ e } k = 0,1,....P$$

Finalmente, o sistema a ser resolvido torna-se :

$$R_h(i) = \sum_{k=1}^{P} \alpha_k\ R_h(|\ i - k\ |) \qquad 1 \le i \le P$$

O erro mínimo toma a seguinte forma :

$$\overline{E}_{h\ min} = R_h(0) - \sum_{k=1}^{P} \alpha_{k\ ótimo} R_h(k) \tag{5.21}$$

O conjunto de equações (5.21) pode ser escrito em forma matricial :

$$\begin{bmatrix} R_h(0) & R_h(1) \ldots\ldots R_h(P-1) \\ R_h(1) & \vdots \\ \vdots & \vdots \\ R_h(P-1) \ldots\ldots\ldots\ldots & R_h(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_P \end{bmatrix} = \begin{bmatrix} R_h(1) \\ R_h(2) \\ \vdots \\ R_h(P) \end{bmatrix}$$

$$R_h \cdot \vec{\alpha_k} = \vec{R_P} \tag{5.22}$$

A matriz de Autocorrelação $\mathbf{R_h}$ é Toeplitz e este fato é utilizado na solução do sistema de equações (5.22) através do algoritmo de Levinson-Durbin.[3]

5.4.2 O MÉTODO DA COVARIÂNCIA

O método da Covariância é utilizado na análise LPC tradicional como uma alternativa ao método da Autocorrelação. Este método se caracteriza por considerar apenas os erros de predição e(n) do intervalo $\mathbf{P \leq n \leq N - 1}$. Assim :

$$\overline{E}_h = \sum_{n=P}^{N-1} e^2(n) \qquad (5.23)$$

Neste caso a predição linear é realizada com o filtro de predição "cheio". Este fato evita os erros de predição característicos das extremidades do segmento sob análise quando se aplica a predição linear diretamente ao sinal de voz. Por isto é possível dispensar o uso de janelas deformadoras como a de Hamming, por exemplo.

O método da Covariância resulta num sistema de equações que matricialmente tem uma forma semelhante à equação (5.22). Neste caso, porém, a matriz $\mathbf{R_h}$ não é Toeplitz e o sistema de equações resultante é resolvido através de um algoritmo conhecido como Decomposição de Cholesky [3].

Os resultados obtidos com o método da Covariância normalmente são superiores àqueles obtidos pelo método da Autocorrelação. Entretanto, o método da Covariância não garante que o conjunto de coeficientes resultante represente um filtro estável.

O fato do método da Covariância realizar a predição linear com filtro "cheio" é o grande responsável pelo seu desempenho superior, quando aplicado na Análise LPC do sinal de voz. Entretanto, é também o responsável por um desempenho francamente desfavorável quando utilizado em conjunto com a Desconvolução Homomórfica. Os erros de predição ocorridos no intervalo de 0 a P não pesam na estimativa dos coeficientes (α_k) :

$$\overline{E}_h = \sum_{n=P}^{N-1} \left[h(n) - \sum_{k=1}^{P} \alpha_k h(n-k) \right]^2 \qquad (5.24)$$

No entanto, a informação contida na porção inicial da resposta impulsiva é a mais

importante na definição do filtro H(z). Dessa maneira, o filtro estimado pelo método da Covariância não possui uma resposta impulsiva que se aproxime de modo satisfatório da resposta impulsiva desejada h(n). Uma tentativa de resolver este problema utilizando uma abordagem para a predição linear do tipo *backward*, não traz resultados satisfatórios devido ao agravamento dos problemas de instabilidade do filtro resultante.

5.4.3 GANHO DO MODELO

Na análise LPC tradicional o cálculo do ganho do modelo é realizado partindo-se da idéia de que a energia do erro de predição deve ser igual a energia da excitação [3]. Tem-se então:

$$G^2 = \overline{E}_{h\ min} = R_h(0) - \sum_{k=1}^{P} \alpha_{k\ ótimo} R_h(k) \qquad (5.25)$$

Com a realização de uma Desconvolução Homomórfica antes do modelagem AR, o ganho do modelo pode ser determinado de modo alternativo. Uma vez obtido o cepstrum da resposta impulsiva, a amostra inicial, $\hat{h}(0)$, representa o logaritmo natural do ganho *G* (Ver item 2.4). Desta forma, o valor de G será independente de estimativas para a função de autocorrelação do conjunto de amostras disponível, bem como do conjunto de coeficientes $\{\alpha_k\}$ resultante, como calculado a partir da equação 5.25. Além disso, a validade da equação 5.25 depende de uma escolha adequada para a ordem do preditor [3].

5.4.4 AVALIAÇÃO FINAL

Com base nos resultados discutidos nos itens anteriores, pode-se inferir algumas conclusões com respeito a aplicação da Análise LPC diretamente sobre a resposta impulsiva do trato vocal .

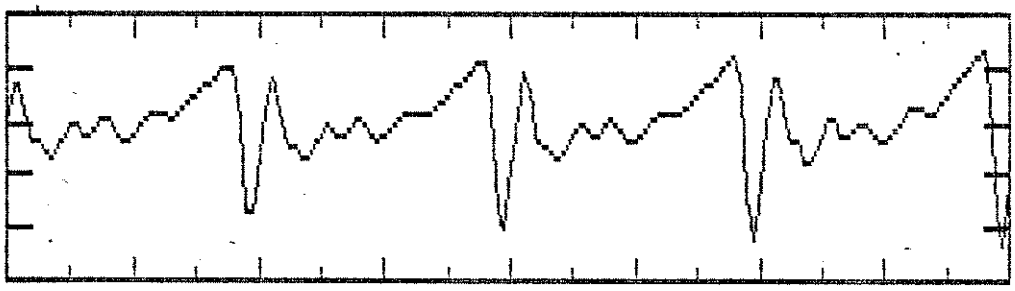
Em primeiro lugar, a utilização do Método da Covariância no cálculo dos coeficientes do filtro H(z) não traz bons resultados. Isto ocorre porque a predição linear realizada com filtro "cheio" não leva em consideração os erros de predição

ocorridos na porção inicial da resposta impulsiva , onde se concentra a parcela de informação mais importante para a definição do filtro H(z).

Em relação a utilização do Método da Autocorrelação, a única diferença está no fato de que o janelamento utilizado para minimizar os maiores erros de predição verificados nas extremidades do segmento de voz, não é necessário quando a Análise LPC é realizada sobre a estimativa da resposta impulsiva. Na verdade, a utilização de uma janela do tipo *Hamming* ou *Hanning* leva a uma distorção do filtro H(z) estimado.

Subjetivamente, o uso do método da Autocorrelação conduz a resultados bastantes semelhantes àqueles obtidos quando a Análise LPC é aplicada diretamente sobre o sinal de voz. Para ilustrar estes resultados, a figura 5.3 traz a forma de onda na saída de um sistema que combina a Desconvolução Homomórfica com a Análise LPC convencional.

a) Sinal Original



b) Sinal na Saída de um Vocoder LPC convencional - Ordem 8

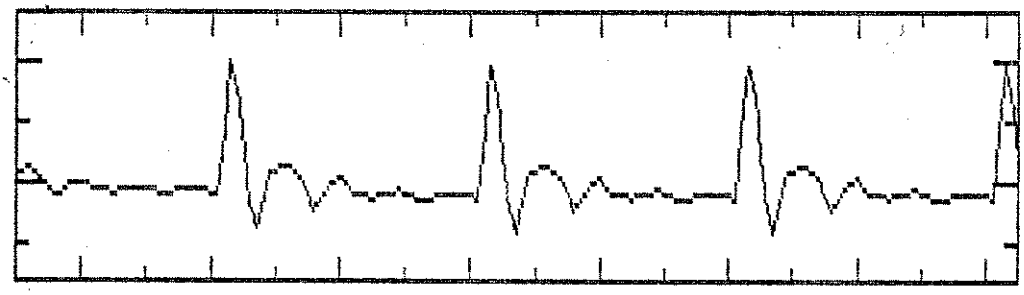
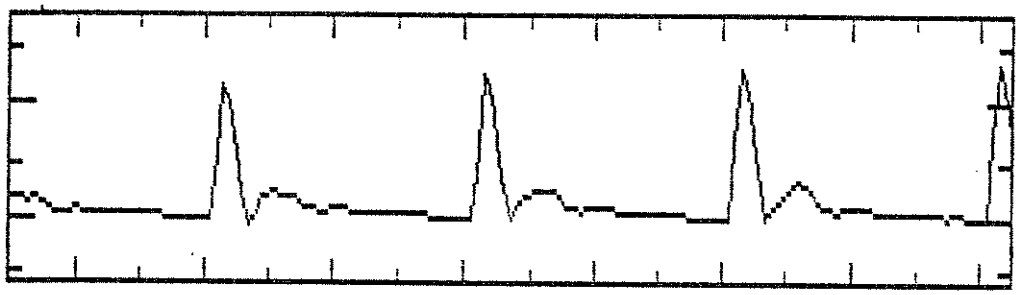


Figura 5.3 - a)Sinal Original; b) Sinal na Saída de um Vocoder LPC Convencional - Ordem 8 ; c) Sinal na Saída de um Sistema com Deconvolução Homomórfica e Análise LPC Convencional - Ordem 8 e Método da Autocorrelação; d) Sinal na Saída de um Sistema com Deconvolução Homomórfica e Análise LPC Convencional - Ordem 8 e Método da Covariância

c) Sinal na Saída de um Sistema com Deconvolução Homomórfica e Análise LPC convencional - Ordem 8 e Método da Autocorrelação



d) Sinal na Saída de um Sistema com Deconvolução Homomórfica e Análise LPC convencional - Ordem 8 e Método da Covariância

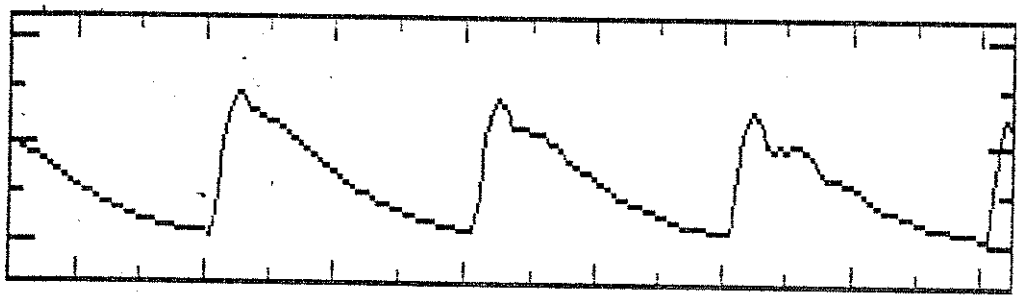


Figura 5.3 - Continuação

5.5 Referências

[1] A. V. Oppenheim, G. E. Kopec and J. M. Tribolet, *"Signal Analysis by Homomorphic Prediction"*, IEEE Trans. Acoust. Speech and Signal Processing Vol ASSP-24, Nº 4 ,pp 327 - 332

[2] G. E. Kopec, A. V. Oppenheim and J. M. Tribolet, *"Speech Analysis by Homomorphic Prediction"*, IEEE Trans. Acoust., Speech , Signal Processing, vol ASSP-25, pp 40-49, Feb. 1977.

[3] L. R. Rabiner and R. W. Schafer, *"Digital Processing of Speech Signals"*, Englewood Cliffs, NJ, Prentice - Hall , 1978.

[4] L. B. Jackson, *"Noncausal ARMA Modeling of Voiced Speech"*, IEEE Trans. Acoust. , Speech and Signal Processing, Vol 37 Nº 10 pp 1606 - 1608 , October 1989

[5] L. B. Jackson, *"Digital Filters and Signal Processing"*, Norwell, MA : Kluwer Academic, 1986

CAPÍTULO 6

CONCLUSÕES

A utilização de um modelo fonte-filtro associado à uma representação paramétrica já é um procedimento consagrado para a Codificação de Voz à Baixas Taxas. Há atualmente uma tendência para a sofisticação do sinal de excitação, o que tem estimulado as pesquisas de sistemas do tipo Multi-Pulse LPC, CELP e RELP. O objetivo é obter um desempenho superior ao do Vocoder LPC baseado no modelo tradicional de produção de voz que utiliza como sinais de excitação o trem de impulsos e o ruído branco.

Este trabalho demonstrou, entretanto, que é possível explorar melhor o potencial do modelo tradicional de produção de voz. Utilizando a Desconvolução Homomórfica, pode-se obter um desempenho superior àquele obtido com o Vocoder LPC tradicional. De fato, a Desconvolução Homomórfica mostrou-se uma melhor alternativa para a obtenção de parâmetros que caracterizem o filtro digital H(z), representativo dos efeitos combinados do Pulso Glótico, do Trato Vocal e da Impedância de Irradiação, segundo o modelo fonte-filtro tradicional.

O desempenho superior obtido com esta técnica é o resultado de dois fatores principais:

- a Desconvolução Homomórfica não pressupõe nenhuma restrição ao filtro H(z), diferentemente da Análise LPC, na qual o filtro H(z) é suposto ser constituído apenas por pólos.
- os parâmetros de transmissão do Vocoder Homomórfico, ou seja, as amostras do cepstrum da resposta impusiva do filtro H(z) são bem menos sensíveis à quantização do que os parâmetros LPC.

Este trabalho também propôs uma nova forma de compensação das distorções oriundas de alterações nos fatores de escala de quadros sonoros e não-sonoros. Mostrou-se que estas alterações são devidas à amostra na origem do cepstrum da excitação. Esta constatação permitiu inferir compensações típicas.A técnica proposta

também permitiu aplicar tratamentos diferenciados aos quadros de transição sonoro/não-sonoro e vice-versa, o que contribuiu para o melhor desempenho do Vocoder Homomórfico em relação ao Vocoder LPC tradicional, o qual não utiliza tais procedimentos.

Os resultados obtidos com o Vocoder Homomórfico de Fase Mista, no qual utiliza-se uma resposta impulsiva não-causal para sintetizar sinais de voz de boa qualidade, adicionam novos elementos à discussão sobre um modelo que represente o mais fielmente possível o processo de produção da fala humana. Desde que o trato vocal humano é certamente um sistema causal, o fato da análise cepstral de fase mista resultar numa resposta impulsiva não-causal sugere a possibilidade de aperfeiçoamento do modelo fonte-filtro tradicional, na busca de alternativas que assegurem a causalidade do filtro H(z) estimado. Talvez seja possível obter modelos que permitam maior fidelidade nos sinais de voz sintetizados e eventualmente possam simplificar a implementação do Vocoder Homomórfico de Fase Mista.

Finalmente, o uso da técnica da Predição Homomórfica proporciona uma alternativa para a redução da taxa de transmissão do Vocoder Homomórfico de Fase Mista, permitindo dessa forma, o aproveitamento do seu elevado desempenho a taxas compatíveis com as necessidades de compressão de alguns sistemas. Vale ressaltar que as taxas de transmissão obtidas neste trabalho foram decorrentes principalmente do uso da técnica da Desconvolução Homomórfica, associada a um procedimento extremamente simples de quantização. Nenhum esforço específico para a obtenção de valores nominais pré-determinados foi realizado.

SUGESTÕES PARA CONTINUAÇÃO DO TRABALHO

Este trabalho pode ser utilizado como ponto de partida para o desenvolvimento dos seguintes temas :

- um estudo mais elaborado do efeito do Jitter de Pitch, o qual se mostrou bastante crítico para desempenho do Vocoder Homomórfico de Fase Mista.

- o estudo da redundância entre os modelos das porções causal e anti-causal da resposta impulsiva estimada pelo Vocoder Homomórfico de Fase Mista e a análise de procedimentos que proporcionem uma exploração mais eficiente desta redundância.

- o estudo de técnicas alternativas de quantização , visando uma redução das taxas de transmissão obtidas neste trabalho

- a aplicação da Desconvolução Homomórfica na Análise de Sinais Sísmicos em Sistemas de Prospecção de Petróleo, no Processamento de Sinais de Sonar e Radar e também em Sistemas de Equalização Cega.

- aperfeiçoamento do modelo fonte-filtro tradicional aplicado à Desconvolução Homomórfica de Fase Mista.