2
Codificação de Voz

Um fator crítico para o desempenho de codificadores de voz que operam a baixas taxas é a quantidade de bits transmitidos em cada quadro. Tal relação é definida no esquema de codificação adotado. Em geral, quanto maior o número de bits utilizado, melhor a qualidade da voz sintetizada no decodificador. Entretanto, é preciso determinar um ponto de equilíbrio de modo que se obtenha a qualidade de voz desejada sem se usar muitos bits. É comum os esquemas de codificação reservarem a maior parte dos bits para os parâmetros LSF quantizados. Uma maneira de se obter um ganho na qualidade do sinal decodificado é buscar obter LSF's mais precisos.

Esta dissertação analisa e propõe melhorias em um processo de ajuste de coeficientes LSF apresentado em [5]. Nessa análise será utilizado o codificador a baixas taxas proposto em [9]. A Figura 2.1 dá uma visão geral do ponto em que processo de ajuste se encaixa no codificador.

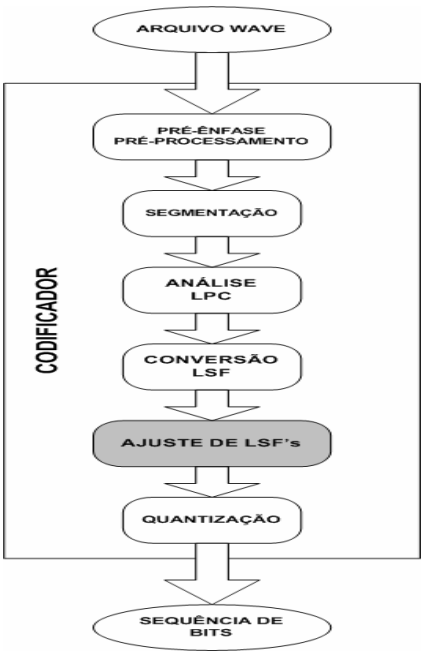

Figura 2.1 – Posição do ajuste de LSFs no codificador.

PUC-Rio - Certificação Digital Nº 0321227/CA

2.1
Análise LPC

A análise LPC consiste na obtenção de coeficientes preditores através de um filtro correlator que processa segmentos curtos de voz, de 20 ms no caso do codificador utilizado. Pela sua simplicidade, é a técnica mais usada atualmente na codificação de voz. Tal análise se baseia na idéia de estabelecer um modelo linear para o mecanismo de produção da fala, ilustrado na Figura 2.2. Neste modelo, a excitação, *u(n)*, é a entrada do sistema, e *s(n)*, a fala sintetizada, é a saída.

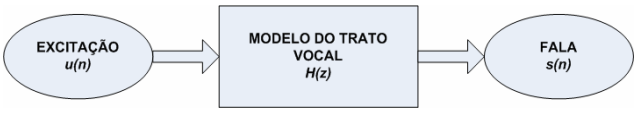

Figura 2.2 – Modelo de produção da fala

PUC-Rio - Certificação Digital Nº 0321227/CA

O filtro *H(z)*, linear e só de pólos, é caracterizado pela Equação (2-1).

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{1}{A(z)} \tag{2-1}$$

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \tag{2-2}$$

Esse filtro varia com o tempo de modo a acompanhar a envoltória espectral do sinal de voz, de onde são extraídos seus parâmetros. Tais parâmetros buscam representar uma aproximação da envoltória espectral [2], chamada de envoltória espectral de base segmentar curta (*eebsc*).

Os parâmetros do filtro correlator são quantizados e codificados, assim como os parâmetros de excitação, para serem transmitidos na forma de uma seqüência de bits. No decodificador essa seqüência de bits é convertida novamente em parâmetros que são usados no processo de síntese da fala.

O vetor $\{\boldsymbol{a}_k=[a_1 \ldots a_p]^T\}$ contém os coeficientes preditores ou LPC (linear prediction coding). A variável *p* é a ordem do filtro e *k* é o índice do quadro. Em

geral, para o padrão de freqüência de amostragem de 8 kHz das redes de telefonia, um filtro de décima ordem é capaz de representar de modo satisfatório a *eebsc*

De modo a reduzir a taxa de bits, sem, entretanto, prejudicar a qualidade da voz sintetizada no decodificador, é empregado o processo de interpolação. Tal processo permite a amostragem de parâmetros LPC em uma taxa menor no codificador, que depois é aumentada no decodificador. A Figura 2.3 ilustra a interpolação de parâmetros LSF entre os últimos sub-quadros do quadro atual e do anterior, definido pela Equação (2-3), onde f_{ij} é o parâmetro LSF de i-ésima ($1 \le i \le p$) ordem do j-ésimo ($1 \le j \le 4$) sub-quadro. Os parâmetros $f_{i_{anterior}}$ e $f_{i_{atual}}$ são os vetores LSF de i-ésima ordem do último sub-quadro do quadro anterior e do atual, respectivamente. A interpolação tem o efeito de produzir transições mais suaves na *eebsc* [9].

$$f_{ij} = (1 - 0{,}25\,j)\, f_{i_{anterior}} + 0{,}25\, j f_{i_{atual}} \tag{2-3}$$

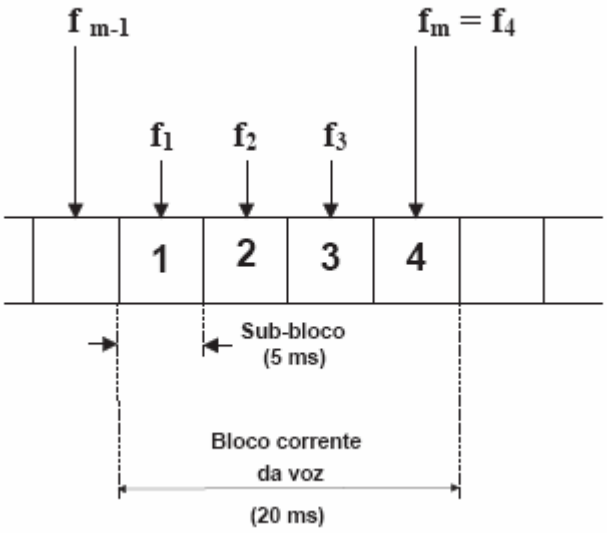

Figura 2.3 – Interpolação de parâmetros LSF.

PUC-Rio - Certificação Digital Nº 0321227/CA

2.2
Parâmetros LSF

Devido à característica de possuírem uma grande faixa dinâmica, os coeficientes LPC não devem ser quantizados e interpolados diretamente. Tal procedimento pode causar instabilidade no filtro de síntese. Em função disso, são usados os parâmetros LSF [4], que correspondem a uma transformação dos parâmetros LPC.

Os parâmetros LSF são obtidos através de dois polinômios, um simétrico, *P(z)*, e outro anti-simétrico, *Q(z)*, representados pelas Equações (2-4) e (2-5), respectivamente, obtidos aumentando-se a ordem do filtro inverso *A(z)*, Equação (2-2), para *p+1*

$$P(z) = A(z) + z^{-(p+1)} A(z^{-1}) \tag{2-4}$$

$$Q(z) = A(z) - z^{-(p+1)} A(z^{-1}) \tag{2-5}$$

Os polinômios simétricos se relacionam segundo

$$A(z) = \frac{1}{2}\left[P(z) + Q(z)\right] \tag{2-6}$$

Os parâmetros LSF são obtidos através das posições angulares das raízes dos polinômios. As LSF's apresentam como característica principal o ordenamento, o que garante estabilidade ao filtro de síntese. Esse ordenamento se mantém mesmo após a interpolação. A sensibilidade das LSF's a erros é apenas local, isto é, um erro de quantização afeta a *eebsc* somente nas vizinhanças. Um agrupamento de LSF's indica a localização de formantes sendo, portanto, uma região importante no espectro.

PUC-Rio - Certificação Digital Nº 0321227/CA

2.3
Descrição do codificador

Nessa dissertação é utilizado um codificador a baixas taxas [9] proposto por De Lamare e Alcaim, que opera a uma taxa média de 1,2 kbps. O codificador utiliza janelas deslizantes (que permitem uma maior precisão) para a detecção do período fundamental, o que é importante para a distinção entre quadros surdos e sonoros. O emprego de janelas deslizantes possibilita a redução do ruído artificial normalmente encontrado em segmentos não-estacionários que contém vogais e proporciona decisões sonoras e períodos fundamentais mais precisos. Na codificação dos sons sonoros é usada a excitação mista em multibandas (EMM). Sons fricativos e oclusivos surdos são detectados utilizando-se a taxa de cruzamentos por zero e o sinal residual da análise LPC, respectivamente. O sinal de voz é segmentado em quadros de 20 ms que passam pela predição linear para depois serem transformados em parâmetros LSF. Assim como o período fundamental, a excitação e o ganho, as LSF são aplicadas a um quantizador vetorial (QV). A Tabela 2.1 apresenta a alocação de bits pra cada tipo de parâmetro.

PUC-Rio - Certificação Digital Nº 0321227/CA

Tabela 2.1 – Alocação de bits.

Parâmetro	Quadro sonoro	Quadro surdo
Fundamental	6	0
Excitação	3	3
Ganho	5	5
LSF’s	21	0

No decodificador, os sons sonoros são filtrados dando origem a excitação. Para os sons surdos a excitação é totalmente surda, isto é, o sinal de excitação não é aplicado ao banco de filtros. O decodificador recebe as LSF’s, interpola e converte em coeficientes LPC que formarão o filtro de síntese. Nesse filtro são aplicados a excitação e o ganho, dando origem ao sinal sintetizado. Finalmente, esse sinal é aplicado a um banco de filtro de modo a melhorar a qualidade da voz

decodificada. A Figura 2.4 apresenta diagramas em blocos para o codificador e o decodificador.

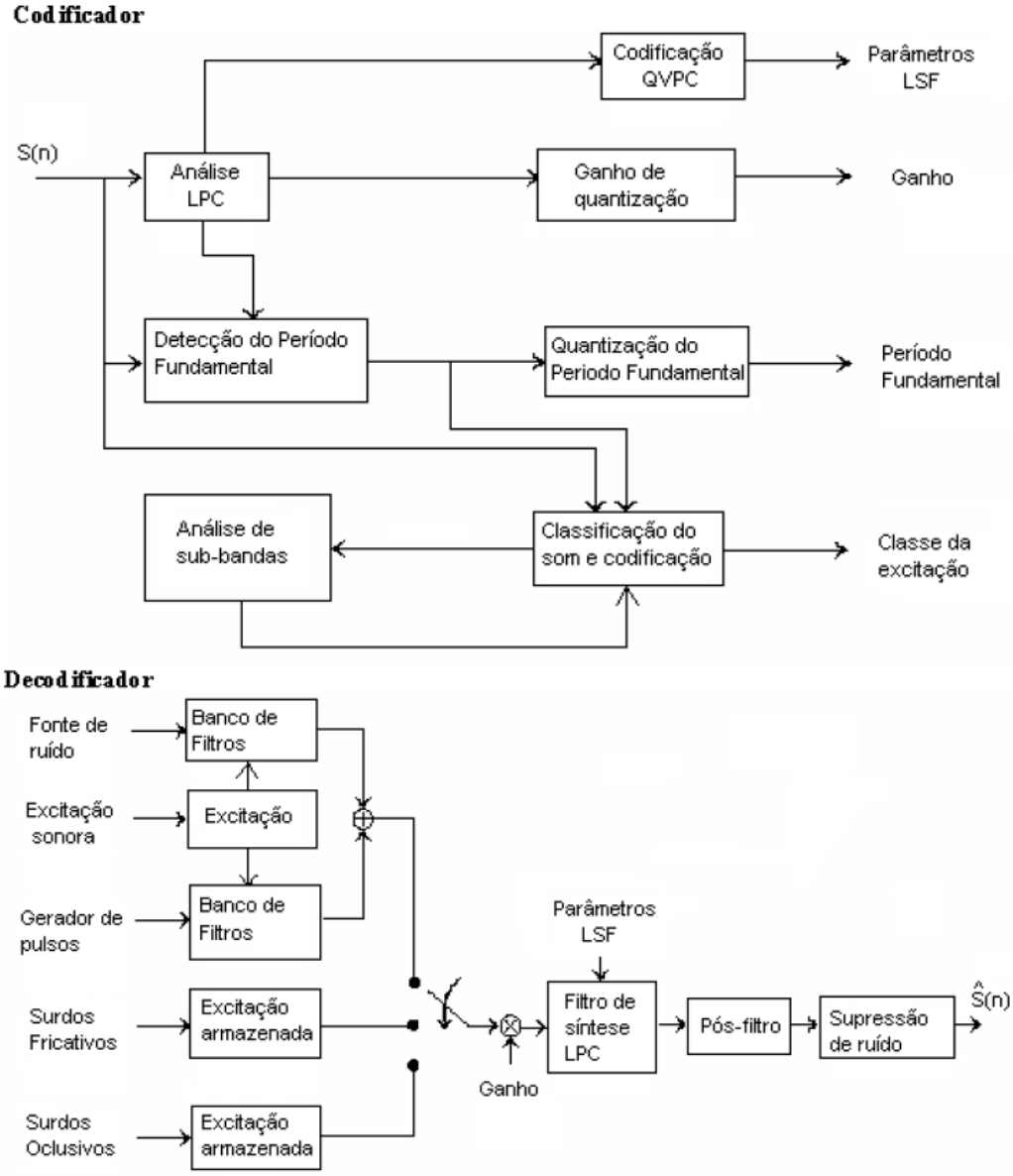

Figura 2.4 – Diagrama em blocos do codificador e do decodificador.

PUC-Rio - Certificação Digital Nº 0321227/CA