2
Modelo do Sistema

A expansão do uso das tecnologias de redes tem possibilitado uma ampla utilização dos serviços de comunicações para o transporte de dados, voz e imagens com taxas de transmissão cada vez mais elevadas. Essa evolução das redes levou ao aparecimento de tecnologias para o fornecimento de serviços de telefonia utilizando a rede de pacotes no estabelecimento de chamadas e comunicação de voz.

A primeira idéia para transmitir voz através de uma rede de pacotes é digitalizar (codificar) o sinal de voz, produzindo um arquivo de dados e enviá-los utilizando um protocolo de transferência de arquivos. Ao chegar no destino, usamos um decodificador para reproduzir a voz novamente.

Um dos problemas na codificação da voz é o fato de alguns canais de comunicação, como internet por exemplo, terem capacidade de transmissão limitada, sendo necessário codificadores com taxas de bits mais baixas do que a suportada pela rede telefônica convencional. Além disso, a transmissão de voz em redes de pacotes exige que os codificadores sejam robustos a elevadas taxas de erro de bits e perda de pacotes na transmissão, em função das características hostis das redes de pacotes para transmissão de voz. Outro problema enfrentado na codificação é o ruído ambiente, que pode afetar a qualidade do sinal codificado. Para codificação da voz, usaremos um codificador a baixas taxas de bits proposto por de Lamare e Alcaim [7] nas simulações de transmissão de voz em uma rede de pacotes.

A Figura 2.1 mostra um diagrama ilustrativo do sistema estudado neste trabalho, onde o sinal de voz é captado, codificado, encapsulado, transmitido em uma rede de pacotes e finalmente decodificado. A qualidade do sinal de voz decodificado será avaliada nessa dissertação para diversos esquemas de quantização de LSFs e para condições de perda de quadros em redes de pacotes decritas em um trabalho recente [8].

Este capítulo apresenta uma descrição dos diferentes esquemas de quantização das LSF e do codec utilizado nas simulações. Também são discutidos os problemas de transmissão de LSFs em redes IP e apresentadas

as condições de rede em que ocorrem as simulações.

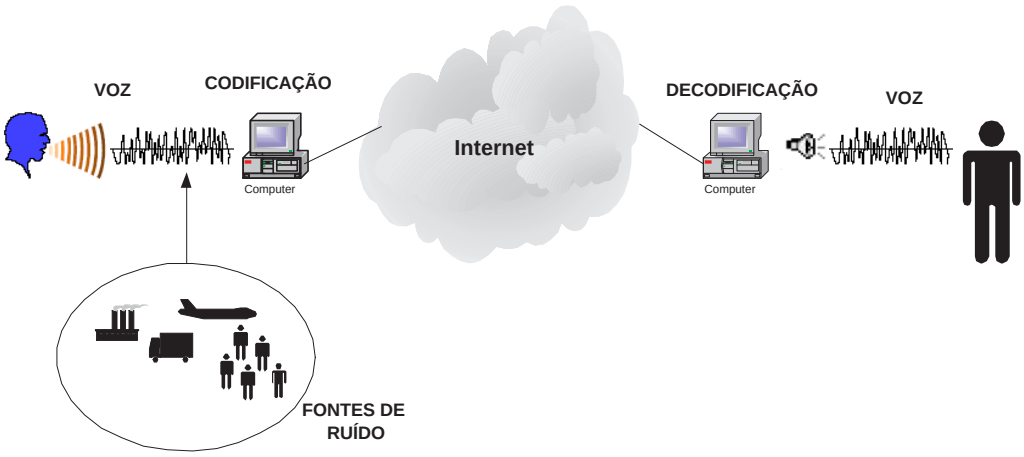

Figura 2.1: Diagrama ilustrativo do sistema estudado.

2.1 Análise LPC e Diferentes Esquemas de Quantização das LSFs

2.1.1 Análise LPC

Com base nas características do mecanismo vocal humano, foi desenvolvido um modelo linear de produção da fala, onde a fonte de excitação e o aparelho vocal são considerados como dois sistemas separados. De acordo com esse modelo, o sinal de voz em tempo discreto $s(n)$ é a resposta do sistema de filtragem do aparelho vocal a uma ou mais fontes de som, e suas propriedades podem ser especificadas em termos de características de fonte e filtro. Assim o sinal $s(n)$ é a convolução discreta no tempo da forma de onda que caracteriza a excitação $e(n)$ com a resposta impulsional do filtro $h(n)$ que caracteriza o aparelho vocal.

Os parâmetros obtidos a partir de uma análise apropriada do sinal de voz caracterizam os mecanismos utilizados para a geração dos sons da fala: a fonte de excitação e o aparelho vocal. Para transmissão, esse parâmetros são digitalizados através de um sistema eficiente de quantização ou codificação. Um sistema de codificação paramétrica de voz é composto basicamente, de um sistema de análise e um de síntese. O sistema de análise tem a função de extrair, a partir do sinal de voz original, dois conjuntos de parâmetros: um representativo da excitação e um representativo do aparelho vocal.

Os parâmetros representativos do aparelho vocal são usualmente obtidos do sinal de voz original através da análise por predição linear (ou

análise LPC - "Linear Predictive Coding"), e por esse motivo são também chamados de parâmetros LPC.

Através de uma análise ou processamento de um segmento curto de 20 a 30 ms é possível obter uma estimativa da envoltória espectral [2]. Um nome mais apropriado seria envoltória espectral de base segmentar curta (*eebsc*), que na língua inglesa é *short-time spectral envelope*. O modelo escolhido para representar *eebsc* é o filtro $H(z)$, que é a transformada-z do filtro $h(n)$ que caracteriza o aparelho vocal.

O filtro $H(z)$, utiliza uma estrutura só de pólos. Essa estrutura, na qual os efeitos de zeros são aproximados por múltiplos pólos, pode ser representada por:

$$H(z) = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \tag{2-1}$$

onde a_i , com $i = 1, 2, \cdots, p$, são os coeficientes de predição linear ou coeficientes LPC.

Para uma freqüência de amostragem igual a 8 kHz, padrão em telefonia, sabe-se que um modelo $H(z)$ de ordem $p = 10$ é capaz de propiciar uma boa aproximação da *eebsc* para a grande maioria dos sons da fala e locutores.

A interpolação dos coeficientes de predição linear é um ponto importante na melhoria da qualidade da voz, uma vez que ao calcularmos os coeficientes LPC em um dos subquadros e interpolarmos esses coeficientes linearmente, a interpolação proporciona transições espectrais mais suaves e uma voz decodificada de melhor qualidade [9]. Neste trabalho os quadros de voz têm duração de 20 ms e são divididos em 4 subquadros de 5 ms, como mostrado na Figura 2.2.

Não é recomendável quantizar diretamente os coeficientes preditores ou parâmetros LPC (a_i) ou mesmo interpolá-los, devido a algumas de suas características. Eles têm por exemplo, grandes faixas dinâmicas e, além disso, pequenas mudanças em seus valores podem levar a grandes mudanças nas posições dos pólos gerando instabilidade ao filtro. Entretanto, os coeficientes preditores podem ser transformados em outros parâmetros que representam igualmente a envoltória espectral em intervalo curto do sinal de voz e que têm propriedades mais atraentes que os a_i. Uma dessas transformações fornece um conjunto de parâmetros, chamados parâmetros LSF (Line Spectrum Frequencies) [4]. As LSFs asseguram maior estabilidade ao filtro de síntese quando quantizadas e também são mais indicadas para

interpolação linear do que os coeficientes de predição obtidos na análise LPC.

A Figura 2.2 ilustra a interpolação dos parâmetros LSF, entre o quarto subquadro do quadro atual e do quadro passado de voz, definido da seguinte forma:

$$f_{i,j} = (1 - 0,5j) f_{i_{anterior}} + 0,5j f_{i_{atual}} \tag{2-2}$$

onde $f_{i,j}$ é o parâmetro LSF de i-ésima ordem para o j-ésimo subquadro ($1 \leq i \leq p \quad e \quad 1 \leq j \leq 2$) e $f_{i_{anterior}}$ e $f_{i_{atual}}$ são os parâmetros LSF de i-ésima ordem calculados no subquadro anterior e atual, respectivamente.

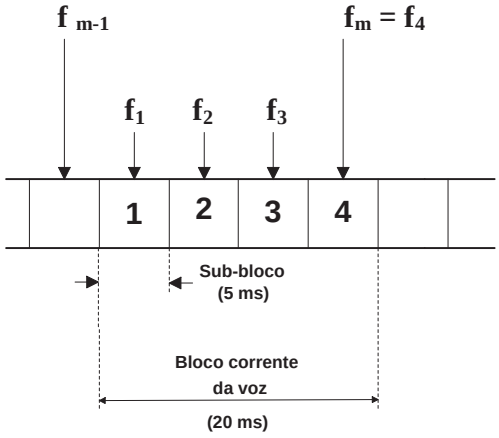

Figura 2.2: Diagrama ilustrando a interpolação dos parâmetros LSF.

Existem diferentes esquemas de quantização que podem ser usados para quantizar de forma eficiente os parâmetros LSF, esses esquemas serão apresentados na próxima seção. Inicialmente serão descritos alguns dos mais populares esquemas de quantização vetorial multiestágios com busca em árvore, onde inclui-se a quantização vetorial sem memória (QVSM), a quantização vetorial preditiva (QVP) e uma abordagem mais recente denominada quantização vetorial preditiva chaveada (QVPC).

2.1.2
Quantizadores Vetoriais de LSFs Utilizados nas Simulações

Esta seção, trata da codificação de forma eficiente dos parâmetros LSF que modelam a *eebsc* e na qual são empregadas técnicas de quantização vetorial. Em sistemas operando a baixas taxas de bits, a transmissão desta informação consome uma parcela significativa da taxa de bits total e, por isso, têm sido feitos grandes esforços em pesquisas, no sentido de desenvolver técnicas eficientes de codificação que utilizem o menor número de bits possível na transmissão desses parâmetros. Veremos aqui, alguns dos mais

populares esquemas de quantização vetorial multiestágios com busca em árvore. Esses esquemas são: quantização vetorial sem memória (QVSM), a quantização vetorial preditiva (QVP) e uma abordagem mais recente denominada quantização vetorial preditiva chaveada (QVPC) e algumas variações propostas por de Lamare e Alcaim [7].

Na quantização vetorial sem memória (QVSM), cada vetor de parâmetros LSF é quantizado de maneira independente de qualquer outro conjunto de LSFs [10]. Paliwal e Atal [11] demonstraram que um esquema de QV particionada sem memória é capaz de codificar de forma eficiente os parâmetros LSF com 24 bits por quadro. Um esquema de QV multiestágio sem memória com estrutura em árvore foi apresentado por Leblanc [12] e seu desempenho superou o esquema de QV particionada sem memória. No esquema de QV multiestágio com estrutura em árvore, um vetor LSF $\mathbf{f}$ é aproximado pelo vetor quantizado $\hat{\mathbf{f}}$, dado por:

$$\hat{\mathbf{f}} = \mathbf{c}_m = \mathbf{c}_{1i} + \mathbf{c}_{2j} + \ldots \mathbf{c}_{kl} \tag{2-3}$$

onde k é o número de estágios e $\mathbf{c}_{ki}$ é o i-ésimo vetor-código do dicionário do k-ésimo estágio, representado pelo conjunto $C_k = \{\mathbf{c}_{k,i}; i = 1, \ldots, I_k\}$, onde I_k é o número de vetores-código de LSFs armazenados em cada dicionário.

Uma técnica simples de melhorar o desempenho das estruturas em [12] e explorar a correlação existente entre vetores adjacentes de LSF, é usar quantização vetorial preditiva (QVP) [13]. Um preditor linear vetorial forma uma estimativa dos vetores de entrada como uma combinação linear de observações passadas. Nessa dissertação restringiu-se os experimentos para preditores de primeira ordem, onde o vetor residual de predição (δ_{j+1}) é quantizado vetorialmente e pode ser expresso por:

$$\delta_{j+1} = \mathbf{f}_{j+1} - \hat{\mathbf{f}}_j.\rho^t \tag{2-4}$$

onde ρ é o vetor com os coeficientes de correlação e $\hat{\mathbf{f}}_j$ é a versão quantizada de $\mathbf{f}_j$.

A correlação entre os vetores de LSF pode ser explorada pelo uso de QV com memória. Entretanto, existem situações de rápidas mudanças na envoltória espectral da voz e, portanto, baixas correlações entre os conjuntos de LSFs adjacentes. Essa observação motivou a combinação de técnicas de QVSM e QVP para codificar quadros de baixa correlação separadamente dos quadros com alta correlação, essa técnica é denominada quantização vetorial preditiva chaveada (QVPC) . Uma busca de ambos esquemas de QV é realizada para cada quadro e o melhor candidato, no que diz respeito

a um critério de distorção, é codificado e transmitido [13].

Dentre os esquemas de QVPC, destacamos aquele denominado *safety-net*, proposto por Eriksson *et al.* [14]. Este sistema chaveia entre uma estrutura QV sem memória e uma estrutura QV preditiva (QVPC2), conseguindo uma quantização eficiente dos parâmetros LSF com apenas 20 bits por quadro. Outra técnica de QVPC, introduzida posteriormente por McCree e De Martin [15], usa 2 estruturas multiestágios de QVP com busca em árvore (QVPCP2), onde cada preditor é projetado para um banco de dados de treinamento específico e opera com 21 bits por quadro.

Recentemente, foi desenvolvido um método de QVPC mais eficiente que os métodos QVSM, QVP, QVPC2 e QVPCP2, para quantização das LSF [16][17]. Esse esquema, denominado QVPC4, utiliza uma estrutura com 4 QVs multiestágios com busca em árvore, sendo 3 QVPs e 1 QVSM. Uma variação dessa estrutura é a utilização de 4 QVPs (QVPCP4)[9]. O desempenho dos métodos de QV citados até aqui serão discutidos mais a diante.

Uma questão fundamental em um esquema de quantização vetorial é a medida de distância empregada para identificar o que melhor representa o vetor de entrada dentre os vetores-código do QV. Como essa medida tem que ser repetida muitas vezes para cada vetor de entrada quantizado, o seu cálculo não deve requerer grande esforço computacional. Por outro lado, é importante que a medida escolhida tenha boa correlação com uma medida subjetiva de desempenho [9]. A distorção ou medida escolhida é o erro-quadrático ponderado, onde dados o vetor LSF $\mathbf{f}$ e um vetor candidato a ser uma aproximação $\hat{\mathbf{f}}$, é definida por

$$d(\mathbf{f}, \hat{\mathbf{f}}) = \sum_{i=1}^{p} \alpha_i(\mathbf{f})(f_i - \hat{f}_i)^2 \tag{2-5}$$

onde $\alpha(\mathbf{f}) = (\alpha_1(\mathbf{f})\alpha_2(\mathbf{f}) \ldots \alpha_p(\mathbf{f}))$ é o vetor de pesos ou função de ponderação. A função de ponderação é definida por

$$\alpha_i(\mathbf{f}) = \frac{1}{f_i - f_{i-1}} + \frac{1}{f_{i+1} - f_i} \tag{2-6}$$

onde $i = 1, \ldots, p$ e $f_0 = 0$ e $f_{p+1} = 0,5$. Como os pesos dão maior ênfase às regiões dos formantes do espectro, a medida de distorção em questão é mais correlacionada com a distância espectral do que uma medida não ponderada.

O desempenho dos QVs é avaliado através da distância espectral (DE)

expressa por

$$DE = \left[\sum_{f=0}^{4000} \frac{1}{4000}\Big(10log10 \mid \frac{S(f)}{\hat{S}(f)} \mid\Big)^2\right]^{1/2} \quad (dB) \tag{2-7}$$

onde $S(f)$ e $\hat{S}(f)$ representam a envoltória espectral original e quantizada, respectivamente.

Os métodos de QVPC que inicialmente foram considerados nessa dissertação são apresentados na Tabela 2.1. Para codificar o quadro de LSF, o QVPC realiza uma busca em todos os QV de sua estrutura e então seleciona o quantizador que minimiza um critério de distorção. Sendo i o índice do QV testado, o $i_{\text{ótimo}}$ é dado por:

$$i_{\text{ótimo}} = \arg\min \left\{ d(\mathbf{f}, \hat{\mathbf{f}}^{(\mathbf{i})}) \right\}_{i=1,\ldots,N_{VQ}} \tag{2-8}$$

onde $d(\mathbf{f}, \hat{\mathbf{f}})$ é o erro-quadrático ponderado e N_{VQ} é o número de quantizadores vetoriais usados pelo QVPC.

A Figura 2.3 [9] mostra o desempenho de todos os esquemas de QV, descritos nessa seção, em termos da distorção espectral média (DE). Os desempenhos são mostrados em diferentes taxas de bits (20, 21, 22, 23 e 24 bits por quadro) e em canal livre de ruído. É importante destacar, que todos os quantizadores vetoriais utilizados aqui, usam uma estrutura de multestágios com busca em árvore, onde cada QV possui 4 estágios e o procedimento de busca em árvore usa uma aproximação dos $M = 12$ melhores vetores-código. A alocação de bits por estágio dos quantizadores vetorias em diferentes taxas, pode ser visto na Tabela 2.2.

Tabela 2.1: Esquemas de QVPC.

QVPC	Estrutura dos Quantizadores
QVPC2	QVPC com 2 classes, onde 1 bit chaveia entre um QVP e um QVSM. Os QVs possuem 4 estágios.
QVPCP2	QVPC com 2 classes, onde 1 bit chaveia entre 2 QVPs. Os QVPs possuem 4 estágios.
QVPC4	QVPC com 4 classes, onde 2 bits chaveam entre 3 QVPs e um QVSM. Os QVs possuem 4 estágios.
QVPCP4	QVPC com 4 classes, onde 2 bits chaveam entre 4 QVPs. Os QVs possuem 4 estágios .

Observa-se na Figura 2.3 que o QVPCP4 obteve melhor desempenho. Entretanto, testes realizados em canais ruidosos [17] mostram que o QVPC4 tem melhor desempenho se comparado ao QVPCP4, pois apresentou o melhor compromisso entre desempenho e robustez contra erros no canal.

Tabela 2.2: Alocação de bits/estágio dos esquemas de QV com 4 estágios.

Taxa de bits/vetor	20	21	22	23	24
QVSM	5 5 5 5	6 5 5 5	6 6 5 5	6 6 6 5	6 6 6 6
QVP	5 5 5 5	6 5 5 5	6 6 5 5	6 6 6 5	6 6 6 6
QVPC2	5 5 5 4	5 5 5 5	6 5 5 5	6 6 5 5	6 6 6 5
QVPCP2	5 5 5 4	5 5 5 5	6 5 5 5	6 6 5 5	6 6 6 5
QVPC4	5 5 4 4	5 5 5 4	5 5 5 5	6 5 5 5	6 6 5 5
QVPCP4	5 5 4 4	5 5 5 4	5 5 5 5	6 5 5 5	6 6 5 5

Já o QVSM apresentou o pior desempenho, mas em canais com altas taxas de erro de bits, o QVSM foi o mais eficiente [17].

Nessa dissertação, utilizaremos 3 tipos de quantizadores vetoriais para avaliarmos a influência da quantização das LSF sobre a qualidade de voz em codecs a baixas taxas operando em redes IP e em ambientes ruidosos. O QVPC4 e o QVSM foram escolhidos, tendo em vista que testes em [17] mostraram que estes são mais robustos em canais com erros de bits. O terceiro quantizador vetorial que será utilizado nas simulações, é o QVPCP2 proposto por McCree e De Martin [15], de modo que se possa avaliar, também, o desempenho de um sistema composto só por QVPs em canais com perda de quadros. As simulações serão feitas utilizando 21 bits para codificar cada quadro de LSFs, mantendo o compromisso entre desempenho e a menor taxa de bits possível.

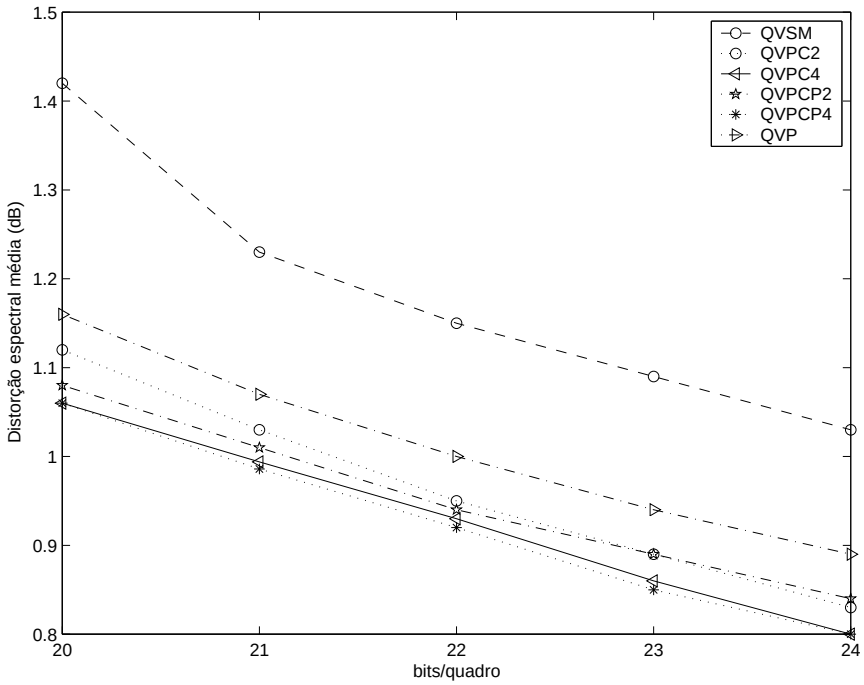

Figura 2.3: Desempenho dos QVs em termos da DE média para diferentes taxas de bits.

2.2 Descrição do Codec

A maioria dos modernos codificadores de voz a baixas taxas, pode entregar uma voz codificada de boa qualidade, trabalhando em uma taxa de 2,4 kb/s. Entretanto, novas técnicas de codificação veêm sendo desenvolvidas com o objetivo de minimizar a taxa de bits, mantendo o compromisso com a qualidade da voz. Nessa dissertação, é utilizado um codificador a baxas taxas, proposto recentemente por de Lamare e Alcaim [7][18], que fornece uma voz de boa qualidade, operando em uma taxa média de 1,2kb/s. As principais características desse codec será descrito de forma sucinta nessa seção.

Uma questão fundamental em codecs a baixas taxas é o algoritmo de deteção do período fundamental, pois estes, além de serem responsáveis pela computação do perído fundamental, são responsáveis pela classificação dos quadros sonoros e surdos. Aqui, o algoritmo de deteção do período fundamental é baseado no método proposto por Unno *et al.* em [19], que utiliza uma janela deslizante, a fim de reduzir os valores incorretos de perído fundamental e de decisões sonoras. O emprego de janelas deslizantes pode reduzir o ruído artificial usualmente encontrado em segmentos não-estacionários que contem vogais e fornecer decisões sonoras e períodos fundamentais mais precisos. O valor do período fundamental T, usando o método da janela deslizante, equivale ao deslocamento para o qual $R_i(T)$ é maximizada e é definido por

$$R(T) = max_{i=-T_s}^{T_s-1} [max_T R_i(T)] \tag{2-9}$$

$$R_i(T) = \frac{C(i, T+i)}{\sqrt{C(i,i)C(T+i, T+i)}} \tag{2-10}$$

onde T_s é o valor máximo do deslizamento e $R_i(T)$ é o valor da função autocorrelação normalizada para o retardo i. A função de autocorrelação $C(k,l)$ é limitada entre 20 e 160 amostras (em uma freqüencia de 8kHz) e é expressa por

$$C(k,l) = \sum_{n=0}^{N-1} s(n+k)s(n+l) \tag{2-11}$$

onde $s(n)$ é o sinal de voz passa-baixa, N é o comprimento do quadro e k e l são os retardos. O valor final do período fundamental é quantizado em uma escala logarítmica por um quantizador escalar uniforme com 63 níveis. Os valores do período fundamental são mapeados em uma palavra-código de 6

bits e interpolados. O nível de quantização associado à palavra-código igual a zero é reservado para uso de uma excitação genérica em quadros surdos.

O ganho é quantizado uniformemente com 5 bits por quadro e a excitação é codificada com 3 bits por quadro. O modelo de excitação utilizado pelo codec explora as vantagens de diferentes técnicas de codificação. Seu algoritmo de classificação de sons foi desenvolvido com o objetivo de separar os quadros em sonoros, fricativos surdos, oclusivos surdos e silêncio. Para a codificação dos sons sonoros é usada a técnica denominada excitação mista em multibandas (EMM) [20], que permite representar de maneira mais adequada os sons sonoros do que o tradicional modelo de excitação sonoro/surdo em banda única. Em aplicações a baixas taxas de bits, é necessário reduzir o número de bandas de freqüências utilizadas no modelo de excitação, e por esse motivo, os quadros de voz classificados como sonoros são divididos em 3 sub-bandas de freqüências, que são implementadas com bancos de filtros fixos. As bandas de freqüencias utilizadas são 0-1kHz, 1-2kHz e 2-4kHz. Em seguida é feita uma análise em cada sub-banda, a fim de determinar se ela é sonora ou surda [9].

Na detecção de sons oclusivos é usado o valor de pico do sinal residual LPC $r(n)$ [9] e uma janela deslizante é empregada a fim de localizar a posição do quadro que maximiza o valor de pico [19]. O valor de pico com a janela deslizante é dado por

$$P = max_{i=-T_s}^{i=T_s} P_i \tag{2-12}$$

$$P_i = \frac{\frac{1}{N}\sum_{n=0}^{N-1} r(n+i)^2}{\sqrt{\frac{1}{N}\sum_{n=0}^{N-1} |r(n+i)|}} \tag{2-13}$$

onde N é o comprimento do quadro de voz e T_s é o valor máximo do deslizamento. Além da medida de pico, a energia do sinal passa-baixa é calculada e usada para distinguir os rápidos ataques de vogais dos sons oclusivos. A detecção de sons fricativos é baseada no uso de limiares apropriados para o número de cruzamentos por zero e a energia de cada quadro. Em geral, estes sinais de baixa energia apresentam entre 60 e 140 cruzamentos por zero, enquanto os quadros sonoros típicos não cruzam o eixo mais de 60 vezes por quadro [21]. Um limiar de energia também é empregado para distinguir os sons fricativos dos quadros em silêncio. Note que apenas os sinais fricativos e oclusivos surdos efetivamente necessitam deste modelo. O algoritmo de detecção do período fundamental separa os fricativos e oclusivos surdos dos sonoros.

Para codificar os sons em sonoros, fricativos surdos, oclusivos surdos e

silêncio, é empregado um dicionário com 3 bits ou 8 índices, dentre os quais 4 índices são designados para quadros sonoros, 2 para quadros oclusivos surdos, 1 para quadros fricativos surdos e 1 para quadros em silêncio.

O codificador realiza uma análise de predição linear a cada quadro de 20 ms. Os parâmetros LPC são transformados em parâmetros LSF e estes são codificados com 21 bits por quadro. Nos experimentos dessa dissertação, utilizaremos 3 esquemas de quantização vetorial das LSF, o QVPC4, o QVPCP2 e o QVSM [16][17]. O objetivo é avaliar o desempenho desses esquemas com o codec operando em redes IP e em presença de ruído ambiente. Esses esquemas foram descritos na Seção 2.1.2.

No decodificador, os sinais de excitação para os quadros sonoros são filtrados por um par de bancos de filtros. Para a reprodução destes quadros de voz, a excitação mista é gerada como a soma das excitações periódica e ruidosa filtradas. Para os quadros surdos, a excitação é declarada totalmente surda e o sinal de excitação não é aplicado ao banco de filtros. Em seguida, aplica-se o sinal de excitação ao filtro de síntese LPC com os coeficientes correspondentes às LSF interpoladas e o ganho decodificado ao sinal de voz sintetizado. Com o objetivo de reduzir o ruído de codificação e melhorar a qualidade da voz decodificada, é empregado um pós-filtro restaurador da *eebsc* e redutor de ruído de codificação (PFRR), que combina duas técnicas para melhorar a qualidade da voz decodificada. O PFRR utiliza um pós-filtro adaptativo de realce espectral (ASE - *Adaptive Spectral Enhancement*)[22] e um pós-filtro adptativo restaurador da envoltória espectral (SER - *Spectral Envelope Restoration*). Isso permite obter um desempenho superior aos métodos ASE e SER [18]. O PFRR tem a seguinte função de transferência:

$$H_{PFRR} = \frac{\tilde{A}(z/\zeta)}{A(z/\eta)}(1 - \nu z^{-1}) \tag{2-14}$$

onde $A(z)$ e $\tilde{A}(z)$ são os modelos da *eebsc* original e reconstruida, respectivamente. Os valores apropriados para ζ, η e ν a baixas taxas de bits são 0.82, 0.9 e $0.3k_1$, respectivamente, onde k_1 é o primeiro coeficiente de reflexão do modelo de predição linear [22]. Esse filtro é seguido de um filtro de supressão de ruído que adota uma versão simplificada da Subtração Espectral Suavizada (SES) [23] [24], de maneira semelhante ao MELP, em razão da sua baixa complexidade quando comparada a SES [9].

A alocação de bits do codificador, assim como a taxa média de bits para codificação de quadros sonoros e surdos são mostradas na Tabela 2.3. E a Figura 2.4 mostra o diagrama em blocos do codec descrito nessa seção.

Tabela 2.3: Alocação de Bits

Parâmetros	Quadro sonoro	Quadro surdo
LSFs	21	0
Ganho	5	5
Excitação	3	3
Fundamental	6	0
Total bits/20 ms	35	8
Bit rate	1,75 kb/s	0,4 kb/s

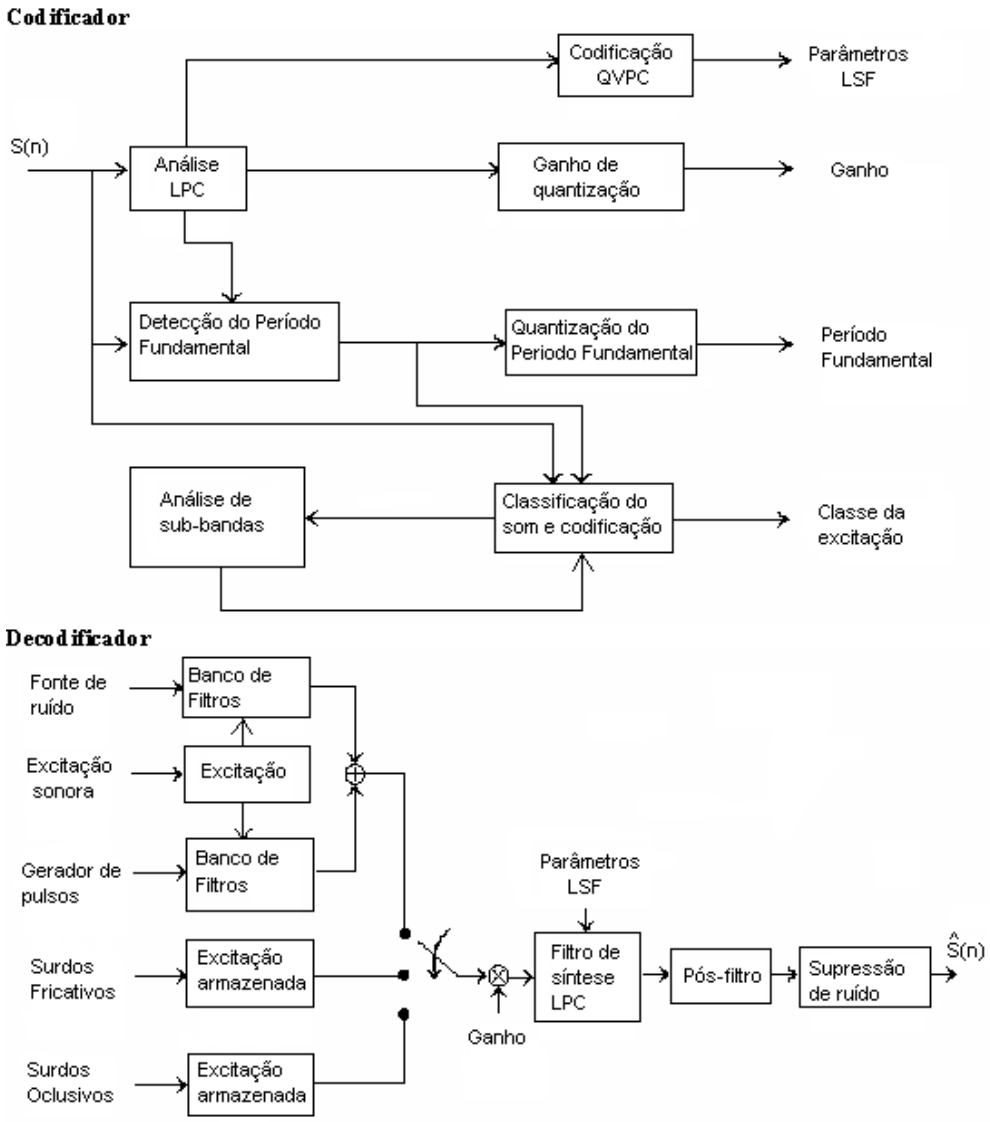

Figura 2.4: Diagrama em blocos do codec.

2.3 Transmissão e Recuperação dos Parâmetros LSF em Redes com Perdas de Pacotes

A primeira idéia para transmitir voz através de uma rede de pacotes é digitalizar o sinal, produzindo um arquivo de dados e enviá-lo através de um protocolo de transferência de arquivos (Figura 2.5). Ao chegar no destino, utilizamos um decodificador para reconstituir o sinal de voz. Entretanto a transferência em tempo real é diferente de uma simples transferência de um arquivo de dados. Em uma transferência de dados, o usuário somente acessará o arquivo após o final da transferência, interessando somente o tempo total de transferência e a integridade do arquivo. No caso de transmissão de voz a taxa de transmissão deve ser relativamente constante, pois o usuário estará ouvindo em tempo real e, em alguns casos, a integridade dos dados pode não ser o elemento mais importante.

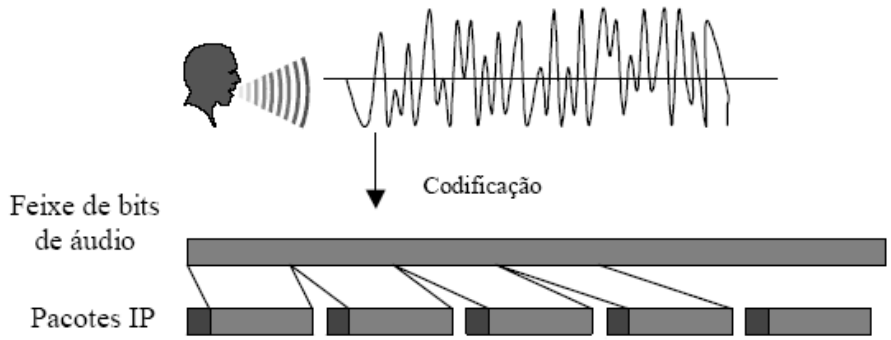

Figura 2.5: Sinal de voz digitalizado.

A voz sobre IP consiste no uso das redes de dados que utilizam o conjunto de protocolos TCP/UDP/IP para a transmissão de sinais de voz em tempo real na forma de pacotes. A voz é digitalizada e transmitida usando uma infra-estrutura LAN ou WAN. Neste caso, não há garantia de serviço, isto é, dependendo do tráfego podem ocorrer retardos na transmissão ou descartes de pacotes. Quando chegam ao seu destino, os dados são convertidos novamente em sinais analógicos. A vantagem é que, usando a Internet, por exemplo, as chamadas telefônicas de voz trafegam juntamente com outros tipos de informação, evitando os custos que essas mesmas chamadas teriam se fossem enviadas isoladamente através da rede de telefonia pública comutada. O impacto mais importante está na separação efetiva entre o controle das chamadas e o transporte. A infra-estrutura necessária para Voz sobre IP necessita de um cabeamento preparado para o transporte de grandes volumes de dados, com priorização de tráfego. Os equipamentos

de rede, principalmente os *switches*, devem possuir uma boa capacidade de tráfego e recursos de qualidade de serviço (QoS).

Nessa dissertação o sistema considerado é uma rede IP onde no terminal de transmissão, um conjunto de parâmetros LSF é quantizado e codificado a cada quadro de voz em uma palavra-código de 21 bits (Figura 2.6). Cada seqüência de conjuntos de parâmetros LSF codificados está associada a uma seqüência de quadros, que por sua vez é caracterizada por uma seqüência de bits. Esses quadros são encapsulados em pacotes, de acordo com o mecanismo de transmissão da rede, e enviados pelo canal. No destinatário, o enquadramento é desfeito pelo decodificador e os parâmetros LSF quantizados são recebidos.

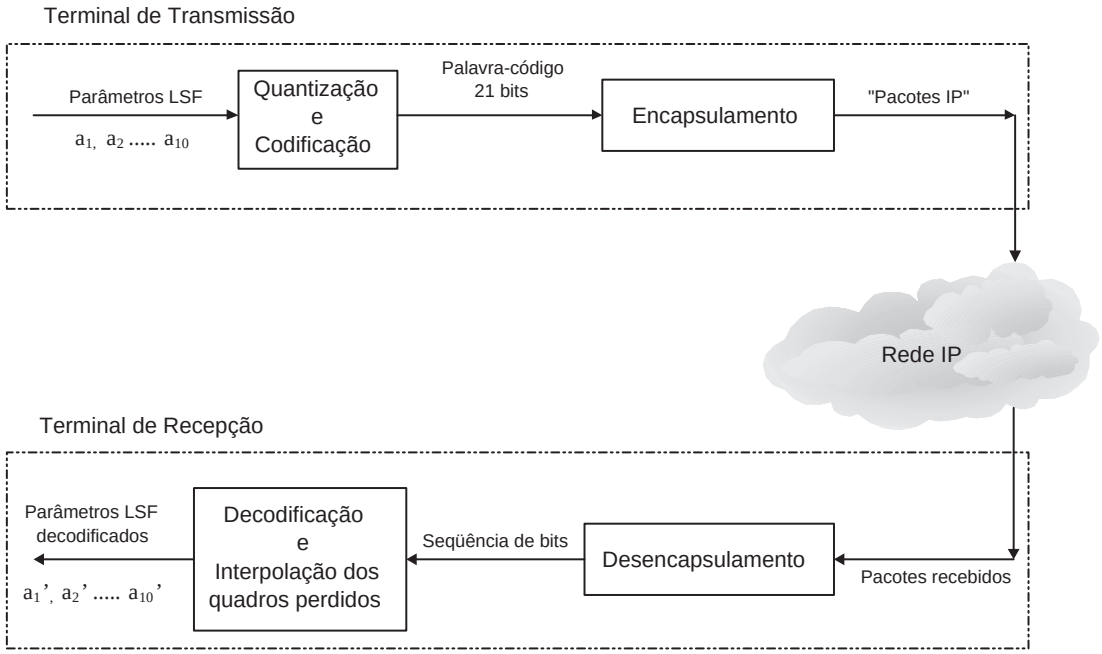

Figura 2.6: Diagrama em blocos do sistema de transmissão e recepção de pacotes IP.

Em uma rede IP, havendo congestionamento, poderá ocorrer a situação de *buffer overflow* nos *switches* ou roteadores levando ao descarte ou perda de pacotes, caracterizando um canal com perda ou apagamento de quadro (*Frame Erasure*–FE). As perdas de pacotes em redes IP normalmente ocorrem em rajadas. Supõe-se, sem perda de generalidade, que um quadro do quantizador é encapsulado em um pacote. Para avaliar o desempenho dos quantizadores em redes IP, usualmente adota-se um modelo Markoviano de dois estados, também conhecido como Modelo de Gilbert. Bolot[8, 25] estudou a distribuição de perdas de pacotes na Internet e concluiu que ela poderia ser modelada dessa forma. Os dois estados se referem aos eventos "pacote recebido" e "pacote perdido", respectivamente. Como mostrado na

Figura 2.7, p é a probabilidade de transição do estado "pacote recebido" para o estado "pacote perdido", e q a probabilidade de transição do estado "pacote perdido" para "pacote recebido". A taxa de perda de quadro (TPQ), também conhecida como probabilidade de perda incondicional é dada por : $TPQ = p/(p+q)$. O comprimento médio da rajada (B) é dado por $B = 1/(1-ppc)$, onde ppc é a probabilidade de perda condicional, que é a probabilidade de transição do estado "pacote perdido" para "pacote perdido" ($ppc = 1$ - q).

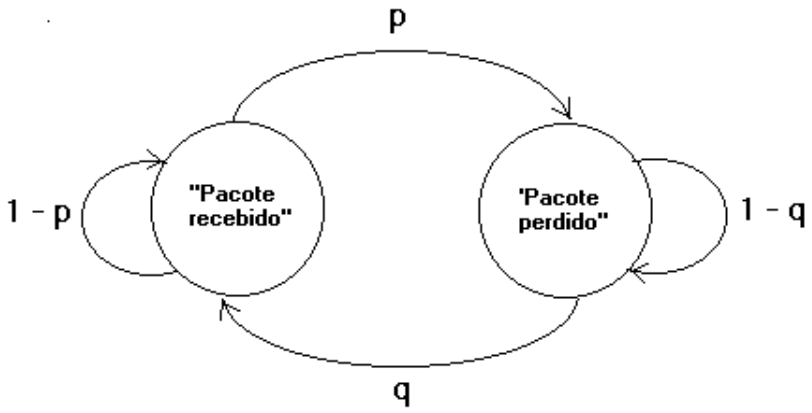

Figura 2.7: Modelo de Gilbert.

Para combater os efeitos de perda de quadros existem algumas contramedidas. A primeira e mais simples delas é silenciar (*muting*) a saída de voz enquanto não se estiver recebendo as LSF perdidas por causa do descarte de quadros. Entretanto, pode-se obter um melhor desempenho se ao invés do *muting* for realizada a repetição (*repeating*) do último quadro de parâmetros LSF recebido. Uma solução um pouco mais atraente e adotada nessa dissertação é a interpolação dos parâmetros LSF. Devido às suas propriedades, as LSF podem ser interpoladas linearmente para formar um conjunto válido de parâmetros LSF. Seja $\widehat{\mathbf{f}}_i$ o conjunto de parâmetros LSF quantizados recebidos no quadro i. Considere o caso em que as L LSFs seguintes são perdidas e $\widehat{\mathbf{f}}_{i+L+1}$ é recebida. Nesse caso, as LSF interpoladas $\widehat{\mathbf{f}}_{i+1} \ldots \widehat{\mathbf{f}}_{i+L}$ são dadas por:

$$\widehat{\mathbf{f}}_{i+x} = \frac{L+1-x}{L+1} \times \widehat{\mathbf{f}}_i + \frac{x}{L+1} \times \widehat{\mathbf{f}}_{i+L+1} \qquad (2\text{-}15)$$

para $x = 1 \ldots L$.

Considere-se, por exemplo, que um conjunto de parâmetros LSF#1 é recebido e, por causa das imperfeições do canal, deixa-se de receber as LSF#2 e recebe-se as LSF#3. A interpolação permite que se obtenha

uma aproximação das LSF#2 às custas de um pequeno retardo adicional. Nota-se que é possível realizar a interpolação de mais de um conjunto de parâmetros LSF às custas de um retardo cada vez maior [26, 27]. De acordo com Wang[27], se n quadros consecutivos, de duração t cada um, forem perdidos, então o atraso D_i devido à interpolação será dado por

$$D_i = n \times t + RTT/2 \tag{2-16}$$

onde RTT (*Round Trip Time*) é o tempo que um pacote leva para ir e voltar de um destinatário (valores típicos para RTT são de 10-700ms). Ainda segundo [27], atrasos aceitáveis para aplicações de voz sobre IP (VoIP) devem ser menores que 800ms.