Capítulo 6

Quantificação

Antes de transmitir qualquer amostra de um parâmetro entre o emissor e o receptor, esta deverá ser codificada com o menor número de bits possíveis de modo a reduzir o débito binário total. Se o parâmetro a transmitir for contínuo, é necessário previamente quantificá-lo, ou seja, aproximá-lo por parâmetro com valores discretos, capazes de serem codificados com um número finito de bits. Esta operação produz ruído de quantificação que deverá ser mantido dentro de limites aceitáveis, segundo um determinado critério. Um dos critérios mais utilizados é o da relação sinal-ruído (SNR) de quantificação, que deverá ser a maior possível. Este capítulo discute alguns dos métodos de quantificação mais utilizados, quer em codificadores de fala quer noutro tipo de codificadores. A parte final do capítulo é dedicada à quantificação de dois importantes conjuntos de parâmetros, utilizados na maioria dos codificadores de fala: os coeficientes LPC e o valor do período de vibração das cordas vocais.

6.1 Quantificação uniforme

Um dos métodos mais comuns de quantificação consiste na divisão da gama de variação [-*V*:*V*] do parâmetro a quantificar em *m* intervalos iguais, sendo o *nível de quantificação* o valor central do respectivo *intervalo de quantificação*. Este procedimento denomina-se por *quantificação uniforme*. Para simplificar a codificação binária, utiliza-se um número de níveis coincidente com uma potência de 2, de modo a optimizar o número *n* de bits de codificação por amostra. Nestas condições, o intervalo de quantificação é dado por:

$$\Delta = \frac{2V}{m} = \frac{2V}{2^n}. \tag{6.1}$$

Utilizando um número suficiente de níveis de quantificação (por *ex.*, $m \geq 16$), o ruído de quantificação pode ser considerado como sendo uma variável aleatória com distribuição uniforme no intervalo de quantificação, centrado na origem. A sua função densidade de probabilidade é, nestas condições, aproximadamente uniforme em torno de cada nível de quantificação, sendo o valor máximo de $\Delta/2$. A potência do ruído de quantificação N_q é dada pela variância σ_n^2,

$$N_q = \sigma_n^2 = \frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} e^2 de = \frac{\Delta^2}{12} = \frac{V^2}{3 \times 2^{2n}}, \tag{6.2}$$

pelo que a potência do ruído de quantificação cresce com o aumento do intervalo de quantificação, ou seja, com a amplitude *V* ou a diminuição de n ou m. A relação entre a potência P_x do sinal e a potência N_q do ruído virá então:

$$SNR = \frac{P_x}{N_q} = 3 \times 2^{2n} \times \frac{P_x}{V^2}, \tag{6.3}$$

ou em decibéis:

$$(SNR)_{dB} = 6{,}02n + 10\log_{10}\left(\frac{3P_x}{V^2}\right). \tag{6.4}$$

A relação sinal ruído é tanto melhor quanto maior for o número n de bits de codificação por amostra, aumentando 6,02 dB por cada bit, sendo também função da potência do sinal de entrada. Por este motivo, deve-se utilizar toda a gama dinâmica possível. Por exemplo, se for utilizada apenas metade da gama dinâmica, que corresponde a uma diminuição em potência de 0,25, a SNR diminui de 6,02 dB, podendo este resultado ser interpretado como a utilização de menos 1 bit de codificação por amostra. Para 0 bits de codificação a SNR resultante é de 1 ou 0 dB, o que equivale a dizer que a potência do ruído é igual à potência do sinal ($x_q[n]=0$; $n[n]=x[n]$).

6.2 Quantificação não uniforme

Uma das desvantagens da utilização da quantificação uniforme é a grande dependência da relação sinal-ruído com a potência do sinal de entrada (ver equação 6.4). A utilização de intervalos com dimensões diferentes, *i.e.*, *quantificação não uniforme*, pode resolver este problema e adequar-se melhor à distribuição do parâmetro a quantificar. Para calcular a relação sinal-ruído nestas circunstâncias, repare-se que este procedimento é equivalente a aplicar uma não linearidade ao parâmetro de entrada seguido de um quantificador uniforme, como ilustra a figura 6.1. No descodificador aplica-se a característica inversa de modo a não distorcer o parâmetro quantificado.

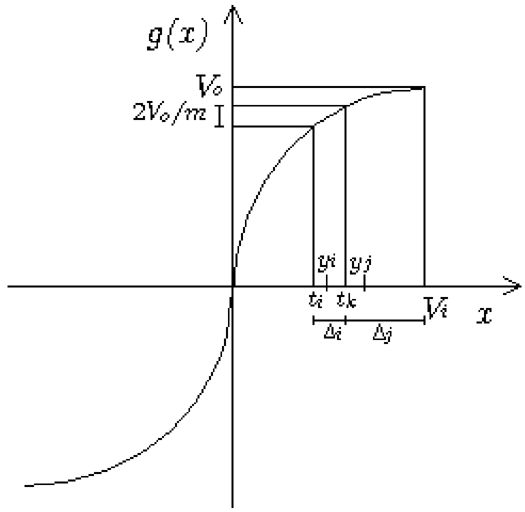

Figura 6.1
Não linearidade seguida de quantificação uniforme, equivalente à quantificação não uniforme.

Repare-se que de modo a ser escolhido o nível de quantificação mais próximo do valor do sinal de entrada, os níveis de decisão t_i que definem o respectivo intervalo deverão estar equidistantes dos níveis de quantificação y_k,

$$t_k = \frac{y_k + y_{k-1}}{2} . \tag{6.5}$$

A função $g(x)$ é quantificada uniformemente com intervalos de comprimento $2V_o/m$. Assim, o comprimento do k-ésimo intervalo de quantificação é:

$$\Delta_k \approx \frac{2V_0}{mg'(y_k)} , \tag{6.6}$$

em que $g'(y_k)$ é o valor da derivada da não linearidade em torno de y_k. Se for conhecida a função densidade de probabilidade do sinal de entrada $f(x)$, a potência do ruído de quantificação para o k-ésimo intervalo é dada pela variância centrada no nível de quantificação y_k:

$$N_k = \int_{t_k}^{t_{k+1}} (x - y_k)^2 f(x) dx \,. \tag{6.7}$$

Se assumirmos um número elevado de níveis, $f(x)$ é aproximadamente constante em cada intervalo, ou seja todos os valores de x têm aproximadamente a mesma probabilidade que $f(y_k)$ e,

$$N_k = f(y_k) \int_{t_k}^{t_{k+1}} (x - y_k)^2 dx = f(y_k) \frac{\Delta_k^3}{12} \,, \tag{6.8}$$

que conjuntamente com a equação 6.6 pode rescrever-se como:

$$N_k = \frac{V_o^2}{3m^2} \frac{f(y_k)}{(g'(y_k))^2} \Delta_k \,. \tag{6.9}$$

A potência total do ruído é a soma da potência do ruído em cada k-ésimo intervalo de quantificação. Esta soma pode ser aproximada por um integral desde que haja um número elevado de níveis:

$$N_q = 2\sum_{k=1}^{m/2} N_k = \frac{2V_o^2}{3m^2} \sum_{k=1}^{m/2} \frac{f(y_k)}{(g'(y_k))^2} \Delta_k \approx \frac{2V_o^2}{3m^2} \int_0^V \frac{f(x)}{(g'(x))^2} dx \,, \tag{6.10}$$

sendo preciso conhecer a derivada da não linearidade para se calcular a SNR final:

$$SNR = \frac{P_x}{N_q} = \frac{3m^2 P_x}{2V_o^2 \int_0^V \frac{f(x)}{(g'(x))^2} dx} \,. \tag{6.11}$$

No caso particular da distribuição uniforme, a derivada de $f(x)$ é 1 e a equação 6.11 reduz-se à equação 6.3. Esta conclusão é válida mesmo quando $g(x)=Cx$, pois se a derivada valer $C=V_o/V_i$ teremos,

$$2V_o^2\int_0^{V_i}\frac{f(x)x}{(g'(x))^2}dx = 2V_o^2\frac{V_i^2}{V_o^2}\int_0^{V_i}f(x)dx = V_i^2 . \qquad (6.12)$$

Uma não linearidade importante é a logarítmica, *i.e.*, $y=\ln(|x|)$, cuja derivada é $1/|x|$. Sabendo que o valor do integral de $2x^2f(x)dx$ calculado no intervalo entre 0 e V é igual à potência do sinal de entrada P_x, o valor da SNR virá igual a $3m^2$, deixando de ser dependente da potência do sinal de entrada para ser dependente apenas do número de níveis de quantificação, ou seja, do número de bits de codificação. A função logarítmica não pode no entanto ser utilizada, pois converte o intervalo 0 a 1 no intervalo $-\infty$ a 0, sendo necessário um número infinito de níveis de quantificação.

Para codificação de um sinal de fala amostra-a-amostra (PCM), encontram-se normalizadas 2 funções pseudo-logarítmicas, que convertem o intervalo entre 0 e 1 no mesmo intervalo, através de uma compressão, tendo a função uma característica impar: a Lei-*A* utilizada na Europa; e a Lei-μ utilizada nos EUA e Japão, ambas descritas na recomendação ITU-T G.711. Para recuperar o sinal sem distorção deverá ser aplicada a não linearidade inversa de expansão, pelo que este procedimento toma o nome de *companding* (*comp*ressor - ex*panding*). Os valores de entrada e saída estão normalizados em relação à amplitude, pelo que $V_o=V_i=1$. Ambas as leis utilizam 8 bits de codificação por amostra e uma frequência de amostragem de 8 kHz, resultando um débito binário de 64 kbit/s. A lei-*A* é descrita por:

$$\begin{aligned} g(x) &= \pm\frac{1+\ln(A|x|)}{1+\ln(A)} \quad \frac{1}{A} \le |x| \le 1 \\ g(x) &= \frac{Ax}{1+\ln(A)} \qquad 0 \le |x| \le \frac{1}{A} \end{aligned}, \tag{6.13}$$

em que o parâmetro A governa o grau de compressão. Para valores baixos a Lei-A tem um comportamento linear, enquanto que para valores médios e altos tem um comportamento logarítmico.

Para a Lei-A, a função derivada vale:

$$\begin{aligned} g'(x) &= \frac{1}{|x|(1+\ln(A))} \qquad \frac{1}{A} \le |x| \le 1 \\ g'(x) &= \frac{A}{1+\ln(A)} \qquad 0 \le |x| \le \frac{1}{A} \end{aligned}. \tag{6.14}$$

Para potências médias e altas, ou seja, quando a potência normalizada do sinal de entrada P_x, relativamente ao valor máximo de quantificação, é razoavelmente superior a $1/A^2$, a primeira sub-equação da equação 6.13 é o termo dominante e a SNR virá, usando o valor normalizado para A de 87,56:

$$SNR \approx \frac{3m^2 P_x}{(1+\ln(A))^2 \int_{-1}^{1} x^2 f(x)dx} = \frac{3m^2}{(1+\ln(A))^2} = 0{,}1m^2, \tag{6.15}$$

ou, em decibéis:

$$(SNR)_{dB} = 6{,}02n - 10. \tag{6.16}$$

Repare-se que a SNR só depende do número de bits de codificação por amostra, deixando de depender da potência do sinal de entrada. Para baixas potências, ou seja quando a potência normalizada do sinal de entrada é razoavelmente inferior a $1/A^2$, o termo dominante é o inferior da equação 6.13, com um comportamento linear, pelo que a SNR é dada pelas equações 6.3 e 6.4.

Para o codificador normalizado de 8 bits (64 kbit/s)[1], este valor máximo da SNR dado pela equação 6.16 é de 38,16 dB e mantêm-se praticamente constante para uma variação apreciável de potência do sinal de entrada. É esta característica quase constante do *companding* que o faz ter um desempenho médio superior ao PCM com quantificação uniforme com o mesmo número de bits.

Para a Lei-μ:

$$g(x) = \pm \frac{\ln(1+\mu|x|)}{\ln(1+\mu)} \qquad 0 \le |x| \le 1 \quad , \qquad (6.17)$$

em que μ governa o grau de compressão, sendo o valor normalizado de 255. Para valores baixos possui também um comportamento linear, uma vez que $\ln(1+\mu|x|) \approx \mu|x|$, sendo que para valores elevados tem um comportamento logarítmico, uma vez que para $\mu|x| >> 1$ resulta $\ln(1+\mu|x|) \approx \ln(\mu|x|)$. A *SNR* virá, assumindo uma distribuição de Laplace[2] para o sinal de fala:

$$\begin{aligned} SNR_{dB} = {} & 6{,}02n + 4{,}77 - 20\log_{10}\left(\ln(1+\mu)\right) \\ & - 10\log_{10}\left(\frac{1+\mu^2 P_x + \mu\sqrt{2P_x}}{\mu^2 P_x}\right). \end{aligned} \qquad (6.18)$$

Para potências normalizadas elevadas (acima dos -15 dB, gama impossível de atingir pelos sinais de fala), a quantificação uniforme tem um desempenho melhor que o *companding*. Nas centrais telefónicas em que está implementada a recomendação G.711, a compressão é

[1] Demonstração na página da disciplina (demos/cod_forma/cod_forma.html)
[2] As distribuições de Laplace e Gamma têm sido usadas como boa aproximação do modelo probabilístico dos sinais de fala.

normalmente desligada quando são detectados na linha sinais sinusoidais típicos da modulação digital (*modems*), cuja potência normalizada é de -3 dB. Nesta situação é utilizada quantificação uniforme.

Embora a principal motivação da recomendação G.711 seja tornar a SNR independente da potência do sinal de entrada, é um exemplo do mascaramento auditivo, pois produz maior potência de ruído de quantificação para valores elevados em que o ruído é mascarado, sendo este minimizado para valores menores em que o ouvido humano é mais sensível.

6.3 Quantificação óptima

Quando existe maior probabilidade de ocorrência de alguns dos níveis de quantificação do que outros, deve-se diminuir o comprimento dos intervalos de quantificação mais prováveis, aumentando o comprimento dos intervalos menos prováveis. Por exemplo, os sinais de fala têm uma distribuição com maiores ocorrências para valores mais baixos, pelo que o *companding* é, como já visto, uma melhor alternativa em relação a um quantificador uniforme. No entanto a principal vantagem do *companding* é a de tornar a SNR praticamente independente da potência do sinal de entrada. Para sinais de grande potência a utilização de *companding* resulta mesmo numa diminuição da SNR em relação à quantificação uniforme, pelo que outro tipo de não linearidade deve ser utilizado.

Foram propostos por Max [Max(60)] e Linde *et al*. [Linde(80)] algoritmos para a determinação dos intervalos de quantificação óptimos e respectivos níveis de saída, baseados num modelo probabilístico da entrada. Dada a função densidade de probabilidade do sinal de entrada

$f(x)$, a distorção total, definida como a potência do ruído de quantificação, será a soma dos erros quadráticos médios (variâncias N_i) do erro de quantificação em todos os intervalos:

$$N_i = \int_{t_i}^{t_{i+1}} (x - y_i)^2 f(x)dx, \tag{6.19}$$

sendo t_i a t_{i+1} os extremos do intervalo de quantificação no nível de quantificação y_i. Os valores óptimos são obtidos minimizando a distorção, ou seja, tomando as derivadas parciais de N_i em ordem a t_i e a y_i. A resolução do conjunto de equações resultante não é fácil, sendo no entanto possível tirar as seguintes conclusões: os níveis de quantificação óptimos são dados por:

$$y_i = \frac{\int_{t_i}^{t_{i+1}} x f(x)dx}{\int_{t_i}^{t_{i+1}} f(x)dx}, \tag{6.20}$$

ou seja, y_i é a média de x no respectivo intervalo de quantificação; os níveis de decisão t_i óptimos são dados pela equação 6.5.

Descreve-se a seguir um algoritmo iterativo para a estimação dos níveis óptimos de quantificação, conhecido por Lloyd-Max [Max(60)], que tem como entrada a função densidade de probabilidade do parâmetro a quantificar, estimada como o respectivo histograma. Na figura 6.2 é mostrado o histograma de um sinal de fala e os níveis de quantificação iniciais (uniforme) e finais, obtidos por este algoritmo. O ruído de quantificação baixou 2,8 vezes, ou seja, foi produzido um aumento da SNR de 4,5 dB.

Entrada:

Número m de níveis a quantificar;

Entradas do histograma normalizado dos valores de entrada
x_i: i=1:k; k>>m;

Histograma normalizado dos valores de entrada
$f(x_i)$: i=1:k;

Índice de distorção ε;

Inicialização:

Calcula-se arbitrariamente m níveis de quantificação
y_j; j=1:m; (*e.g.*: quantificação uniforme);

Inicializa-se a ordem de iteração n=0;

Inicializa-se a distorção $N_{-1} = \infty$;

Iteração:

1) Calcula-se os espaços S_j, j=1:m com $x_i \in S_j$ se

$(x_i - y_j)^2 \leq (x_i - y_l)^2$, com $l \neq j$, l=1:m;j=1:m;

2) Calcula-se o ruído de quantificação total N_n

$$N_n = \sum_{j=1}^{m} \sum_i \left(x_i - y_j\right)^2 f(x_i) \quad \text{em que } x_i \in S_j;$$

3) Se $\varepsilon \geq \dfrac{N_{n-1} - N_n}{N_n}$ termina-se, sendo os níveis óptimos dados por y_j ; j=1:m, e os níveis de decisão calculados segundo a equação 6.5;

4) Calculam-se novos níveis de quantificação como a média dos espaços S_j:

$$y_j = \frac{\sum_i x_i f(x_i)}{\sum_i f(x_i)} \qquad j=1 \ldots m, \text{ em que } x_i \in S_j;$$

5) Incrementa-se n e retorna-se ao ponto 1);

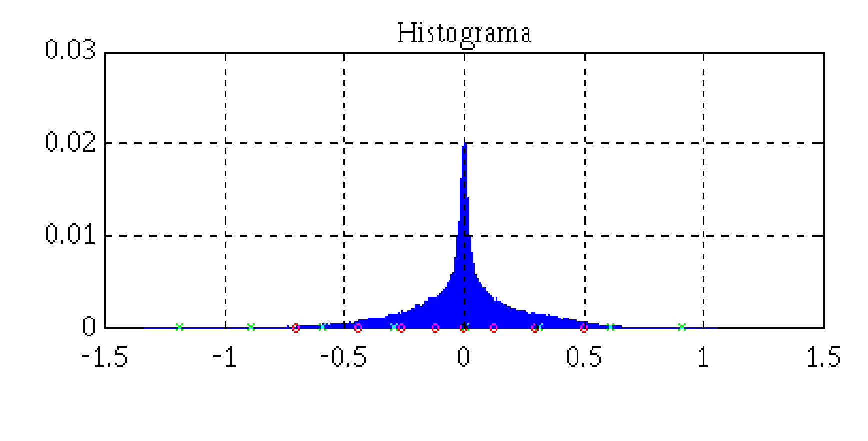

Figura 6.2
Histograma de um sinal de fala.
x - Níveis para quantificação uniforme (distorção de 170.000)
o - Níveis óptimos (distorção de 60.000).

Na tabela 6.1 são mostrados os níveis de quantificação y_i para sinais de entrada com funções densidade de probabilidade Gaussiana e Gamma, para 2, 3 e 4 bits de codificação [Max (60)].

Distribuição Gaussiana			Distribuição Gamma		
N=2	N=3	N=4	N=2	N=3	N=4
0,4528	0,2451	0,1284	0,313	0,155	0,073
1,510	0,7560	0,3831	2,223	0,899	0,387
	1,344	0,6568		2,057	0,795
	2,152	0,9424		4,121	1,307
		1,256			1,959
		1,618			2,822
		2,069			4,061
		2,733			6,195

Tabela 6.1
Valores dos níveis de quantificação y_i para sinais com funções densidade de probabilidade Gaussiana e Gamma, para 2, 3 e 4 bits de codificação.

6.4 Quantificação vectorial

A quantificação individual dos parâmetros a transmitir pode ser pouco eficaz em termos do número total de bits transmitidos. A quantificação vectorial combina num único vector $\vec{x}_i$ de dimensão M um conjunto de parâmetros correlacionados, quantificando-os simultaneamente, de modo a minimizar uma distância d entre estes e um vector de quantificação $\vec{y}_j$. A distância normalmente utilizada é a distância euclidiana pesada,

$$d(\vec{x},\vec{y}) = \sqrt{\sum_{k=1}^{M} w_k (x_k - y_k)^2}\ , \tag{6.21}$$

em que w_k é o peso do elemento k, normalmente escolhido conforme a importância relativa deste elemento em relação aos outros. Caso não se queira diferenciar entre cada um dos elementos, considera-se w_k=1.

Seguidamente apresenta-se um algoritmo iteractivo [Linde (80)] para escolha dos vectores óptimos de quantificação, dado um conjunto de vectores de entrada utilizado para o treino do quantificador. Este algoritmo é conhecido por LBG, as iniciais dos três investigadores que o propuseram em 1980 para codificação de sinais de fala (Yoseph Linde, Andrés Buzo, Robert M. Gray). Repare-se que se o vector de entrada possuir apenas um elemento (escalar), então este algoritmo é equivalente ao apresentado para quantificação óptima escalar, embora menos eficiente ao especificar as amostras de entrada individualmente e não através do histograma.

Para se quantificar um vector $\vec{x}_i$, compara-se este com todos os vectores de quantificação $\vec{y}_k$ resultante do treino, sendo escolhido aquele que apresentar a menor distância d_{ij}. A quantificação vectorial é também utilizada no contexto do reconhecimento com treino não supervisionado (sem conhecer a classe de cada elemento de treino mas apenas o número de classes pretendidas), sendo os vectores de treino agrupados em classes representadas pelo respectivo *centróide* (nível de quantificação) e a classe reconhecida para um dado vector de entrada a correspondente ao *centróide* que apresentar a menor distância.

Entrada:

Número m de vectores de quantificação;
Vectores de entrada $\vec{x}_i$;
Índice de distorção ε;

Inicialização:

Calcula-se arbitrariamente m vectores de quantificação $\vec{y}_j$, j=1:m;
Inicializa-se a ordem de iteração n=0;
Inicializa-se a distorção $D_{-1} = \infty$;

Iteração:

1) Calcula-se os espaços S_j, j=1:m com $\vec{x}_i \in S_j$ se:

$$d_{ij} \leq d_{il} \text{ , com } l \neq j,\ l=1:m; j=1:m;$$

sendo d_{ij} a distância entre o vector de entrada $\vec{x}_i$ e o vector de quantificação $\vec{y}_j$, tomando cada espaço S_j a dimensão N_j;

2) Calcula-se a distorção total D_n

$$D_n = \sum_{j=1}^{m} \sum_{i=1}^{N_j} d_{ij} \text{ ,}$$

sendo d_{ij} a distância entre o vector de entrada $\vec{x}_i$ e o vector de quantificação $\vec{y}_j$, em que $\vec{x}_i \in S_j$;

3) Se $\varepsilon \geq \dfrac{D_{n-1} - D_n}{D_n}$ termina-se, sendo os vectores óptimos dados por $\vec{y}_j$; j=1:m;

4) Calculam-se novos vectores de quantificação como a média (*centróide*) dos espaços S_j.

$$\vec{y}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \vec{x}_i \qquad j=1\ldots m, \text{ em que } \vec{x}_i \in S_j,$$

individualmente para cada valor do vector $\vec{y}_j$;

5) Incrementa-se n e retorna-se ao ponto 1).

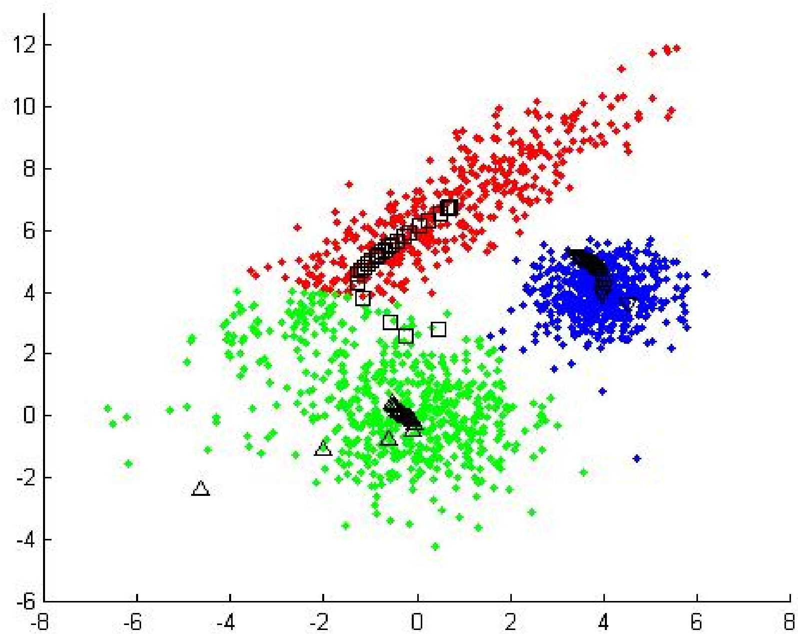

Figura 6.3
Exemplo de quantificação vectorial com três classes.

Uma descrição pormenorizada da quantificação vectorial e sua utilização na transmissão pode ser encontrada nos capítulos 9 e 10 [Hedelin (95)] do livro de [Kleijn (95)].

6.5 Quantificação dos coeficientes LPC

Os coeficientes LPC devem ser recalculados para tramas de duração entre os 10 e os 30 ms, de modo a acompanhar a evolução temporal do espectro dos sinais de fala. Os coeficientes obtidos têm que ser quantificados de um modo eficiente para serem transmitidos para o receptor. A quantificação directa dos coeficientes LPC não é viável devido à sua grande gama dinâmica e à possibilidade do filtro se tornar instável. É então necessário operar transformações dos coeficientes LPC para minimizar estas desvantagens. Os parâmetros transformados podem ser quantificados quer escalarmente quer vectorialmente. São a seguir apresentadas três transformações usadas na quantificação dos coeficientes LPC, podendo ser encontrada uma descrição pormenorizada no capítulos 12 [Paliwal (95)] do livro de [Kleijn (95)].

6.5.1 Coeficientes de Reflexão

Os coeficientes de reflexão ou de correlação parcial (PARCOR) têm a mesma informação que os coeficientes LPC, mas têm a vantagem de garantir a estabilidade do filtro mesmo depois da quantificação, já que esta é assegurada pela condição descrita na equação 3.33 (módulo inferior a 1). Na quantificação deve ser levada em conta a sensibilidade em forma de U, mais sensível perto de ± 1 [Viswanathan (75)], sendo para tal necessário utilizar quantificação não uniforme.

6.5.2 Coeficientes LAR

Outra não linearidade possível de ser utilizada para quantificar os coeficientes de reflexão é a do *arco-seno*, pois uma sinusóide tem uma função de probabilidade com uma característica próxima da destes. Mais utilizada ainda é a transformação para coeficientes LAR (*Logarithmic Area Ratio*) ou logaritmo da relação de áreas. Fisicamente os coeficientes LAR podem ser interpretados do modo seguinte: considerando um tubo acústico com p secções, cada uma com área A_i, a relação de áreas[3] é dada por:

$$\frac{A_{i+1}}{A_i} = \frac{1-k_i}{1+k_i} \quad , \tag{6.22}$$

sendo k_i os coeficientes de reflexão entre as secções. Os coeficientes LAR são então definidos como:

[3] Numa linha de transmissão, a relação de áreas é o inverso do factor de onda estacionária, dado pela relação de impedâncias)

$$g_i = log\left(\frac{1-k_i}{1+k_i}\right), \tag{6.23}$$

e os coeficientes de reflexão são recalculados dos LAR através do procedimento inverso:

$$k_i = \frac{1-10^{g}i}{1+10^{g}i}. \tag{6.24}$$

Através da equação 6.23 facilmente se verifica que esta transformação mapeia o intervalo $-1 \leq k_i \leq 1$ no intervalo $-\infty \leq g_i \leq \infty$. Os parâmetros LAR têm uma distribuição que se aproxima da uniforme, com uma baixa interdependência entre parâmetros, o que é importante na presença de erros no canal de transmissão.

6.5.3 Coeficientes LSF

Os coeficientes LSF (*Line Spectrum Frequency*), apresentados no capítulo 3, são hoje os mais utilizados na quantificação da informação dos coeficientes LPC. Recorde-se que, para manter a estabilidade do filtro, os coeficientes LSF devem apresentar-se sob uma ordem crescente no intervalo $[0:fs/2]$ (ou $[0:\pi]$), encontrando-se os coeficientes dos polinómios $P(z)$ e $Q(z)$ entrelaçados.

A quantificação individual de cada frequência é pouco eficaz do ponto de vista do débito binário. Embora menos imune a erros do canal, como as frequências têm valor crescente, pode-se quantificar com muito maior eficiência a diferença de frequência em relação à frequência anterior (LSFD - *Line Spectrum Frequency Difference*) [Soong (88)].

Uma técnica híbrida entre quantificação vectorial e escalar para os coeficientes LSF é proposta por Mano [Mano (89)]. Nesta, para um vector representando os p coeficientes LSF a quantificar, são escolhidos os 4 vectores que produzem menor distorção, de entre um conjunto de 256, sendo a procura mantida com uma carga computacional aceitável. Seguidamente, a diferença de cada coeficiente LSF em relação a estes vectores é também quantificada, agora individualmente (escalarmente), sendo escolhido o par vector-escalar que origina a menor distorção.

Na quantificação vectorial dos coeficientes LSF, a distância espectral utilizada é por exemplo a distância euclidiana pesada dada pela equação 6.21. Os pesos w_k devem reflectir a importância perceptual de cada um dos coeficientes. Por exemplo, Laroia *et al.* [Laroia (91)] propõem que:

$$w_k = \frac{1}{LSF_k - LSF_{k-1}} + \frac{1}{LSF_{k+1} - LSF_k}, \qquad (6.25)$$

baseados na propriedade de que, quanto mais próximos os valores de dois coeficientes LSF consecutivos estiverem, menor é a largura de banda do pólo do tracto vocal correspondente, marcando muito possivelmente um formante. Estas regiões são perceptualmente mais sensíveis às variações dos coeficientes LSF [Crosmer (85)], sendo-lhes atribuído maior peso no cálculo da distância espectral.

Kemp *et al.* [Kemp (91)] propõem que a quantificação vectorial seja produzida por partes, ou seja, quantificando separadamente conjuntos (3, 3, 4) de coeficientes LSF adjacentes (p=10). A motivação para esta separação é a correspondência entre os conjuntos considerados e os formantes. Normalmente, considera-se que os 3 primeiros coeficientes LSF estão relacionados com o primeiro formante, o segundo conjunto de 3 coeficientes LSF com o segundo formante e os 4 últimos coeficientes com os formantes de ordem superior. O quantificador utiliza um total de 18 bits de quantificação por trama.

6.6 Quantificação de *F*0

Um parâmetro crucial que é transmitido na maioria dos codificadores é o valor da frequência fundamental *F*0 nas zonas vozeadas. Os métodos de estimação da frequência fundamental, nomeadamente os que utilizam a função de autocorrelação (ou suas variantes), estimam o período fundamental em múltiplos do período de amostragem. Assumindo que as frequências possíveis de vibração das cordas vocais situam-se entre os 55 Hz e os 400 Hz, o que para a frequência de amostragem de 8 kHz corresponde a um período de 146 a 20 intervalos de amostragem, estes são capazes de ser codificados com um código de 7 bits. Como este valor só é válido nas zonas vozeadas, é necessário transmitir a decisão de vozeamento, o que pode ser efectuado com mais um bit de codificação. Produz-se assim um código com apenas 1 bit nas zonas não vozeadas e 8 bits nas zonas vozeadas. De modo a evitar um código de dimensão variável e perda de robustez na presença de um erro na transmissão do bit com informação de vozeamento, muitos codificadores utilizam um código fixo de 7 bits, representando uma das hipóteses uma trama não vozeadas e as outras 127 hipóteses uma trama vozeada com informação no código do respectivo período em número de amostras.