

O Codificador GSM a 13kbit/s

Francisco Vaz, António J. S. Teixeira

Resumo – O objectivo deste texto é fazer uma introdução à constituição de um codificador de uso generalizado no sistema europeu de telemóveis GSM. Apresentam-se primeiro, para contextualizar, aspectos gerais sobre a codificação da fala para fins de transmissão da voz humana em redes telefónicas, descrevendo-se as principais ferramentas de processamento de sinal utilizadas, nomeadamente a predição linear. Em seguida descreve-se a estrutura do codificador e decodificador GSM 6.10 e de cada um dos blocos constitutivos.

Abstract – The objective of this paper is to introduce the structure of a coder used widely on the european mobile phone system GSM. First is presented the general principles to code human speech for transmission over telephone networks by introducing the main signal processing tools used, specially addressing linear prediction. After, the structure of the GSM 6.10 coder and decoder, is presented describing each of their partial blocks.

Palavras chave – Codificação, descodificação, codec, modelo de predição linear, LPC, vocoder, GSM

Keywords – Coding, decoding, codec, linear prediction model, LPC, vocoder, GSM

I. CODIFICADORES DE FALA

A. Introdução

Codificador é um sistema que implementa códigos, entendidos como regras para converter a informação numa outra forma ou representação, não necessariamente do mesmo tipo. Formalmente dados dois conjuntos de símbolos A e B , um código será uma aplicação de A em B , associando a qualquer sequência de elementos de A uma sequência de elementos de B .

Em telecomunicações, codificação é o processo através do qual a informação proveniente de uma fonte é convertida em símbolos que poderão ser transmitidos e a descodificação será a operação inversa, convertendo estes símbolos em informação compreensível por um receptor. No caso particular que se irá considerar, a comunicação telefónica, a fonte da informação será a fala representada na forma digital, isto é, uma sequência de amostras feitas à frequência de amostragem f_s , sendo cada uma representada por um dado número de bits. A sua codificação tem como objectivo principal a redução da quantidade de informação que se transmite ou seja, a compressão de dados para optimizar a utilização de um dado canal ([1] e [2]). A compressão de dados faz-se à custa da eliminação de informação, em parte redundante, garantindo no entanto que a fala depois de descodificada mantenha o grau de inteligibilidade suficiente para a aplicação desejada, neste caso, a comunicação telefónica.

Os codificadores de fala existentes costumam agrupar-se em diversas classes: codificadores de forma de onda, codificadores de fonte ou *vocoders* (*voice coders*) e codificadores híbridos. Os codificadores de forma de onda procuram reproduzir com a maior fidelidade possível a forma de onda no tempo, usando técnicas no domínio do tempo (PCM- *Pulse Code Modulation* e ADPCM- *Adaptive Differential Pulse Code Modulation* sendo os mais utilizados os definidos pelas normas G711, G721). Os codificadores de fonte ou *vocoders* baseiam-se num modelo simplificado de produção da voz, procurando a modelação das características importantes para a inteligibilidade. Finalmente, os codificadores híbridos procuram reproduzir a forma de onda mas recorrendo simultaneamente a um modelo de produção da voz. O codificador que se irá descrever, o GSM a 13kbit/s, ou GSM 6.10, como é conhecido, enquadra-se na categoria dos *vocoders* e foi desenvolvido para o sistema europeu de telemóveis GSM-*Global System for Mobile Communications*, ou Sistema Global para Comunicações Móveis, originalmente, *Groupe Special Mobile*.

Existem ainda codificadores de outros tipos como os codificadores fonéticos que apresentam uma grande capacidade de compressão. No entanto, não há ainda uma normalização que permita o seu uso generalizado em comunicações.

B. Modelo da produção de fala

O aparelho produtor de fala humana pode ser reduzido a um tubo acústico que se inicia na glote (espaço entre as cordas vocais) e termina nos lábios, incluindo as cavidades nasais. Este tubo acústico designado por tracto vocal é excitado por uma onda de pressão periódica provocada pela vibração das cordas vocais, produzindo à saída uma onda sonora que é função desta excitação e das suas características anatómicas. Os diferentes sons são produzidos modificando a configuração do tracto vocal e órgãos associados, podendo considerar-se que a fala resulta da aplicação à onda glotal de um filtro variável no tempo. Bem cedo, no fim da década de 50 do século XX, G.Fant propõe esta teoria da fonte-filtro para modelar o tracto vocal [3], [4].

As cordas vocais, pelo seu movimento periódico de abrir e fechar, modulam o fluxo de ar proveniente dos pulmões transformando-o num fluxo pulsado que pode ser modelado por um trem de impulsos de Dirac filtrado por um filtro formador.

No entanto, a fonte não pode ser um simples gerador de impulsos glotais. De facto, um estudo um pouco mais aprofundado mostra que certos sons não resultam directamente de uma excitação periódica, apresentando características de ruído. Isto é explicável pela dinâmica da circulação do ar no

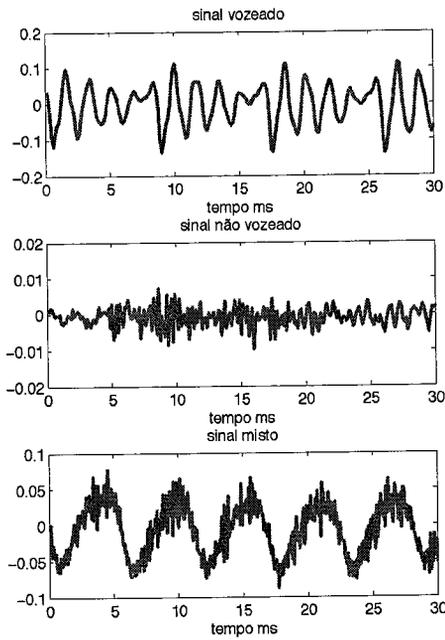


Fig. 1 - Tipos de sons produzidos pelo tracto vocal

tracto que nuns casos se faz de forma lamelar e noutros de forma turbilhonar, permitindo uma primeira classificação dos sons produzidos em sons vozeados (por exemplo a vogal "a") e sons não vozeados (por exemplo a consoante "p"). Existem também sons que são simultaneamente devidos aos dois tipos de escoamento do ar (por exemplo a consoante "v"). Na figura 1 mostram-se as formas de onda temporais de exemplos de diversos tipos de sons.

Tendo em conta o atrás exposto, no modelo fonte-filtro a fonte é modelada por dois geradores comutáveis, um de impulsos periódicos a uma frequência f_0 designada por frequência fundamental e outro de ruído branco. O tracto vocal pode ser modelado, tendo em consideração as suas características anatómicas (não considerando as cavidades nasais) e as leis da mecânica de fluidos, como um sistema linear só com pólos (ver [5]). Um modelo de produção da fala pode ser então o que se indica na figura 2

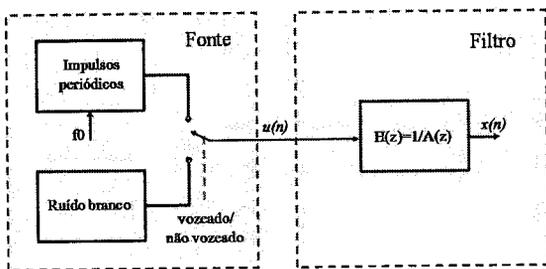


Fig. 2 - Modelo de produção da fala

O sinal de fala pode então considerar-se como a saída $x(n)$ de um sistema linear invariante no tempo com uma função

de transferência

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

só com pólos e excitado por um sinal $u(n)$ constituído por um trem de impulsos periódico (frequência f_0) no caso de ser fala vozeada, ou por ruído branco gaussiano, se a fala tiver características não vozeadas.

C. Modelo LPC

O modelo $H(z) = 1/A(z)$ é um modelo só com pólos e designado por autoregressivo. A determinação dos seus parâmetros a_k é um problema bem estudado por exemplo [6]. A ordem p do modelo que se utiliza para o estudo e caracterização do sinal de fala é em geral baixa dado que o tracto vocal tem características dimensionais que permitem, no caso de um adulto do sexo masculino, um máximo cinco ressonâncias espectrais na banda sonora até 5 kHz banda onde se concentra a informação relevante para a compreensão da generalidade dos sons da fala. Como uma ressonância espectral é modelável por um modelo de 2ª ordem com pólos complexos, basta usar modelos de ordem 10 para as aplicações sobre voz humana (de facto no GSM usa-se $p = 8$ havendo no entanto codificadores similares onde se usa uma ordem um pouco mais alta).

Uma vez que

$$x(n) = u(n) - \sum_{k=1}^p a_k x(n-k)$$

fazendo

$$\tilde{x}(n) = - \sum_{k=1}^p a_k x(n-k)$$

tem-se

$$x(n) = u(n) + \tilde{x}(n)$$

Este resultado pode ser interpretado como este sistema ter, para cada instante, uma saída composta por uma combinação linear das saídas anteriores e da entrada actual. Daqui resulta a designação comum de modelo de predição linear, interpretando $\tilde{x}(n)$ como uma predição do sinal de saída feita a partir do passado recente do sistema e, desta forma, os parâmetros a_k que caracterizam o sistema recebem o nome de coeficientes de predição linear ou LPC (sigla obtida a partir da versão inglesa *Linear Prediction Coefficients*).

A determinação dos a_k faz-se a partir dos $p + 1$ primeiros valores da autocorrelação do sinal, ou seja, utiliza-se a informação estatística inferida de uma pequena janela de p amostras (neste caso com $p = 8$ e $f_s = 8$ kHz e uma duração de 1 ms) pelo que se designa, por vezes, esta predição como uma predição de curta duração.

D. Predição de longa duração

Na figura 3 mostra-se um segmento de fala vozeada e o comportamento da modelação LPC, quando a ela aplicada. Na figura mostra-se o erro de predição ou resíduo que se

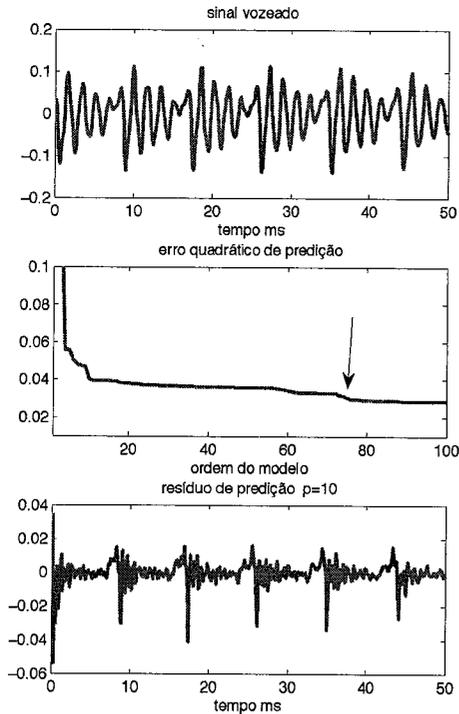


Fig. 3 - Erro de predição e resíduo para um sinal vozeado

obtem aplicando ao sinal o filtro inverso $A(z)$ e, portanto, $u(n) = x(n) - \sum_{k=1}^p a_k x(n-k)$. Este resíduo mostra a diferença entre o sinal predito e o sinal original. Mostra-se também a curva do valor médio quadrático do erro de predição em função da ordem do modelo, indicada como erro quadrático de predição e que mede o grau de afastamento do resíduo ao sinal original. Pode observar-se que, para ordens do modelo próximas de 10, o erro quadrático de predição já é bastante pequeno, justificando a utilização de uma ordem baixa, como já referido. Este erro quadrático de predição mantém-se quase constante a partir daí, verificando-se que há um novo decrescimento sensível para uma ordem entre 70 e 80, conforme se assinala na figura 3. Isto deve-se ao facto de para modelos desta ordem se estarem a considerar as correlações para atrasos de cerca de 70 amostras (cerca de 9 ms ou $f_0 = 110$ Hz) o que corresponde à periodicidade observada no resíduo. Este facto indica que se pode melhorar a predição para tempos bastante superiores ao 1 ms referido. No entanto o aumento da eficácia de predição é feito à custa de aumentar a complexidade do algoritmo que exigiria o cálculo de cerca de 70 coeficientes de correlação e resolver um conjunto de equações mais complexo. Por essa razão se divide o problema: mantém-se a caracterização feita pelo preditor de ordem baixa ou de curta duração e introduz-se uma nova etapa de predição sobre o resíduo apenas considerando a correlação do sinal para atrasos da ordem do período fundamental. Esta predição é designada predição de longa duração e para a sua determinação começa-se por determi-

nar o modelo de curta duração usando o filtro inverso $A(z)$. Ao resíduo assim obtido aplica-se um novo preditor conforme se indica na figura 4. Este preditor é um preditor muito simples pois apenas considera na sua operação uma amostra com um atraso de T , isto é, o valor predito no instante n é proporcional ao valor da amostra de ordem $n-T$, ou seja se $\tilde{e}(n) = bu(n-T)$ for uma predição o sinal à saída será $e(n) = u(n) - bu(n-T)$.

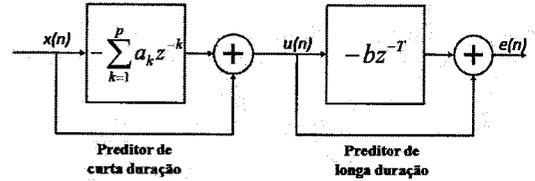


Fig. 4 - Modelo da predição de curta e longa duração

O preditor de longa duração fica então caracterizado pelos dois parâmetros: b o ganho de longa duração e T o atraso. O algoritmo de cálculo consiste em fazer variar T entre valores considerados limites para a voz humana e determinar qual o valor que conduz a maior correlação. Usando este valor e o resíduo, facilmente se determina o ganho b minimizando o valor quadrático médio do erro de predição $E[(e(n) - \tilde{e}(n))^2]$. A sua determinação detalhada pode encontrar-se em [1].

E. Quantificação dos parâmetros do modelo

Em muitas aplicações da predição linear, particularmente na presente, é vantajoso usar um outro conjunto de parâmetros em vez dos referidos a_k . De facto como estes parâmetros vão sofrer uma operação de quantificação, a distorção de quantificação introduzida pode conduzir ao aparecimento de modelos instáveis (com pólos no exterior do círculo unitário) após a descodificação. O algoritmo que permite determinar os a_k (conhecido como algoritmo de Durbin [6]) permite calcular em simultâneo um outro conjunto de parâmetros equivalentes, designados por coeficientes de reflexão $k(i)$, $i = 1, \dots, p$ que gozando da propriedade de $|k(i)| < 1$ em modelos estáveis [6], permitem facilmente controlar a estabilidade do modelo após a descodificação. Como, por sua vez, também estes parâmetros apresentem grande sensibilidade à quantificação podendo ser causa de distorção, opta-se neste *vocoder* por caracterizar o modelo pelos parâmetros *LAR* assim designados por serem o acrónimo da sua designação em inglês *log-area ratios* e derivados dos anteriores através de $LAR(i) = \log_{10} \frac{1+k(i)}{1-k(i)}$, $i = 1, \dots, p$.

F. Determinação do sinal de excitação

De facto no GSM 6.10 não se vai usar uma fonte de excitação tal como foi indicada na descrição do modelo fonte-filtro. É evidente que se se usasse o resíduo a reconstrução seria quase perfeita pois apenas estariam presentes erros de quantificação, mas tal processo não interessa pois não haveria redução significativa dos dados a transmitir, mesmo recorrendo à utilização de códigos otimiza-



Fig. 5 - Os grandes blocos do algoritmo do codificador GSM 6.19

dos. Os codificadores de fonte distinguem-se exactamente aqui, na escolha de qual informação sobre o resíduo se deve transmitir para se obter um sinal reconstruído de boa qualidade, cumprindo-se no entanto as necessárias restrições ao ritmo de transmissão. No GSM 6.10, seguindo a proposta de Kronn em [7] o resíduo é decimado por um factor de 4, sendo portanto possíveis 4 diferentes conjuntos de impulsos. A escolha recai no conjunto de impulsos de maior energia, procedimento justificado por Atal que em [8] mostra que a não inclusão de impulsos de baixa amplitude do resíduo não afecta muito a qualidade da fala sintetizada. Como os impulsos resultam de uma decimação, estão regularmente afastados no tempo (de facto a uma frequência de $8/4 = 2$ kHz) o que justifica o nome atribuído a este *vocoder* : GSM 6.10 RPE-LTP, onde RPE indica que se trata de um sistema que inclui *Regular Pulse Excitation*. A designação LTP justifica-se pela utilização de predição de longa duração, *Long Term Prediction* na língua inglesa.

II. O CODIFICADOR

Em seguida faz-se a descrição do codificador seguindo de perto a norma ETS 300 580-2, 3ª edição de Dezembro de 2000 [9].

Um diagrama de blocos do codificador GSM 6.10 é apresentado na figura 6. O sinal de fala é, como já referido, amostrado a 8 kHz e 13bit/amostra num dispositivo de conversão analógico digital. A norma prevê ainda a possibilidade de ligação a sistemas já digitais que usem um PCM não uniforme de 8bit/amostra definido na norma G711/714. Neste caso deve ser incluída a necessária conversão do PCM não uniforme para um PCM uniforme de 13bit/amostra.

Embora não sejam indicados detalhes sobre os algoritmos utilizados, deve ser referido que todos eles são implementados numa aritmética de números inteiros com o fim de diminuir os tempos de cálculo. De facto, dado que a codificação se faz por quadros de 20 ms, existe um atraso inerente deste valor entre o sinal de entrada e o sinal de saída. Se não houver muito cuidado na implementação das operações que têm de ser efectuadas, este atraso poderá aumentar de modo significativo. Por isso a norma impõe um tempo de atraso máximo de 30 ms.

Conforme se pode ver do diagrama de blocos da figura 6, o sinal começa por ser pré-processado para otimizar as operações que se seguem e é segmentado em quadros de 160 amostras. Os blocos seguintes efectuam a modelização LPC, ficando o modelo caracterizado pelo conjunto dos parâmetros LAR. O resíduo de curta duração é determinado através do filtro inverso $A(z)$ que é caracterizado por parâmetros iguais aos que serão usados no receptor. Para tal procede-se, para além da necessária codificação dos LAR, à sua descodificação e são estes valores resultantes da aplicação das duas operações os usados para a definição do filtro inverso. Em seguida, determina-se o

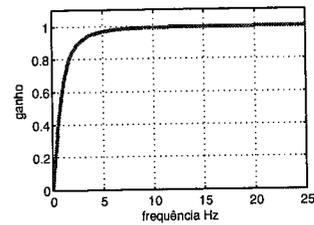


Fig. 7 - Resposta na frequência do filtro de eliminação da componente contínua do GSM 6.10

resíduo de longa duração e para tal subdivide-se o quadro em 4 subquadros de 40 amostras e para cada um são determinados os parâmetros de longa duração. É de salientar que, mais uma vez, esta determinação é feita nas condições da recepção após a codificação e descodificação dos respectivos parâmetros. O resíduo de curta duração é em seguida decimado por factor de 4 e das 4 sequências possíveis de 13 impulsos escolhe-se a de maior energia. Esta sequência e sua posição são em seguida codificadas utilizando um APCM- *Adaptive Pulse Code Modulation*, ou seja, uma codificação por modulação adaptativa de impulsos. Um APCM é um codificador em que os valores a codificar estão permanentemente a ser normalizados por forma a otimizar a gama dinâmica do quantificador que é habitualmente uniforme

Em seguida faz-se uma apresentação mais detalhada de cada uma das quatro partes constituintes, decompostas nos blocos indicados na figura 6.

A. Pré-processamento

A.1 Filtro de eliminação da componente contínua

Trata-se de um filtro *notch* com um pólo e um zero reais, para $z = 0.999$ e $z = 1$, Com uma função de transferência $H(z) = (1 - z^{-1}) / (1 - \alpha_1 z^{-1})$, $\alpha_1 = 32735.2^{-15} = 0.999$. A figura 4 mostra a resposta na frequência deste bloco

A.2 Filtro de pre-ênfase

É conhecido [6] que os modelos autoregressivos não modelam bem os picos espectrais de baixa energia. Como na voz o espectro decai fortemente com a frequência, os picos de alta frequência são em geral de baixa energia e é prática corrente no processamento deste sinal aplicar uma filtragem prévia que acentue as altas frequências. Trata-se de um simples filtro com um zero real para $z = 0.86$ com uma função de transferência $H(z) = 1 - \alpha_2 z^{-1}$, $\alpha_2 = 28180 * 2^{-15} = 0.86$ e a figura 5 mostra a sua resposta na frequência

B. Predição de curta duração

B.1 Segmentação e cálculo do modelo

Para cada segmento de 160 amostras é calculado um modelo de ordem 8. A partir das amostras, calculam-se os 9 valores da autocorrelação usando uma janela rectangular que são depois convertidos nos 8 coeficientes de reflexão $k(i)$ utilizando o algoritmo de Schur [10]. Este algoritmo equivalente ao anteriormente citado algoritmo de Durbin é

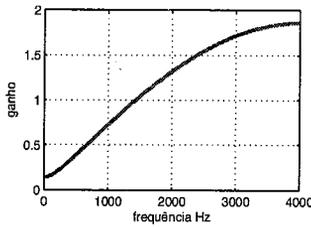


Fig. 8 - Resposta na frequência do filtro de pré-ênfase do GSM 6.10

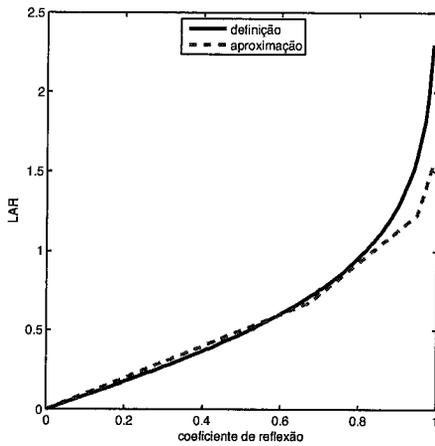


Fig. 9 - LAR em função do coeficiente de reflexão usando a definição e a aproximação

usado por ser facilmente implementável em aritmética inteira.

B.2 Quantificação LPC e interpolação

A definição dos LAR atrás indicada envolve o cálculo de logaritmos. Para evitar qualquer complexidade de cálculo usa-se a aproximação linear segmental que a seguir se indica

$$LAR(i) = \begin{cases} k(i), & |k(i)| < 0.675 \\ \text{sign}(k(i)) \cdot (2|k(i)| - 0.675), & |k(i)| \in [0.675, 0.95] \\ \text{sign}(k(i)) \cdot (8|k(i)| - 6.375), & |k(i)| \in [0.95, 1] \end{cases}$$

A qualidade da aproximação pode ser avaliada observando a figura 9

Os LAR são, em seguida, quantificados e codificados segundo a fórmula

$$\hat{LAR}(i) = Nint((A(i)LAR(i) + B(i)))$$

sendo $Nint(z) = int(z + 0.5sign(z))$ o inteiro mais próximo de z . Os valores de $A(i)$ e $B(i)$ têm em conta a variabilidade estatística dos parâmetros $LAR(i)$ e estão indicados na seguinte tabela. O valor de $LAR'(i)$ assim determinado fica ainda limitado aos valores máximo e mínimo também definidos na tabela

	A(i)	B(i)	mínimo	máximo	Nº de bits
LAR(1)	20.000	0.000	-32	31	6
LAR(2)	20.000	0.000	-32	31	6
LAR(3)	20.000	4.000	-16	15	5
LAR(4)	20.000	-5.000	-16	15	5
LAR(5)	13.637	0.184	-8	7	4
LAR(6)	15.000	-3.500	-8	7	4
LAR(7)	8.334	-0.666	-4	3	3
LAR(8)	8.824	-2.235	-4	3	3

Os valores codificados, para além de serem enviados para o multiplexador de saída, são descodificados usando a transformação inversa

$$LAR'(i) = (\hat{LAR}(i) - B(i)) / A(i)$$

Para evitar variações bruscas no filtro predictor, os LAR codificados são interpolados com os do segmento anterior em 4 diferentes grupos que serão usados para cada um dos 4 subsegmentos

$$LAR'_j = p_j LAR'_j + p_{j-1} LAR'_{j-1}$$

Pêsos para a interpolação dos parâmetros LAR

intervalo	quadro anterior p_{j-1}	quadro actual p_j
[0, 12]	0.75	0.25
[13, 26]	0.50	0.50
[27, 39]	0.25	0.75
[40, 159]	0.00	1.00

B.3 Filtro predictor de erro

Os LAR interpolados são transformados em coeficientes de reflexão usando as fórmulas

$$k'(i) = \begin{cases} LAR'(i), & |LAR'(i)| < 0.675 \\ \text{sign}(LAR'(i)) \cdot (0.5|LAR'(i)| + 0.337500), & LAR'(i) \in [0.675, 1.225] \\ \text{sign}(LAR'(i)) \cdot (0.125|LAR'(i)| + 0.796875), & |LAR'(i)| \in [1.225, 1.625] \end{cases}$$

que vão ser usados no filtro de análise, implementado numa estrutura em *lattice*. À sua saída ter-se-á o erro de curta duração $d(n)$

C. Predição de longa duração

C.1 Subsegmentação

Cada quadro de 160 amostras do resíduo é dividido em 4, correspondendo a cada um 40 amostras ou 5ms.

Cada segmento é agora constituído pelas amostras $d(k_j + k)$, $j = 0, 1, 2, 3$, $k_j = 1 + 40j$, $k = 0, \dots, 39$.

C.2 Predição de longa duração

Para cada subquadro é calculado um modelo de predição de longa duração caracterizado por um atraso T_j e por um ganho b_j . A sua determinação é feita a partir do cálculo da correlação cruzada entre o resíduo actual e o resíduo reconstruído do quadro anterior $d'(k_j + i)$, $i = -120, \dots, -1$,

$$R_j(T) = \sum_{i=0}^{39} d(k_j + i)d'(k_j + i - T),$$

$$k_j = 1 + 40j, T = 40, \dots, 120$$

O valor óptimo do atraso ou período T será, para cada subquadro, o valor que maximiza a correlação cruzada, partindo do resíduo anterior reconstruído

$$T_j = \arg \max_T R_j(T)$$

O ganho de longa duração será para cada subquadro

$$b_j = \frac{R_j(T)}{\sum_{i=0}^{39} d'^2(k_j + i - T)}$$

Como $T_j \in \{41, \dots, 120\}$ são usados directamente e representados em 7 bits, não havendo propriamente uma codificação. Os ganhos b são codificados como \hat{b} e decodificados como b' com dois bits de acordo com a tabela seguinte:

b	\hat{b}	b'
$(0.8, \infty)$	3	1.00
$(0.5, 0.8]$	2	0.65
$(0.2, 0.5]$	1	0.35
$(-\infty, 0.2]$	0	0.10

C.3 Filtro de análise de longa duração

Partindo do resíduo anterior reconstruído $d'(k_j + i)$ e dos parâmetros de longa duração determina-se $d''(k_j + i)$ e, em seguida, o resíduo de longa duração $e(k_j + i)$:

$$\begin{aligned} d''(k_j + i) &= b_j d'(k_j + i - T_j), \\ e(k_j + i) &= d(k_j + i) - d''(k_j + i), \\ j &= 0, 1, 2, 3, k_j = 1 + 40j, i = 0, \dots, 39 \end{aligned}$$

O resíduo de longa duração reconstruído $e'(k)$ é processado em subsegmentos de 40 amostras $e'(k_j + i)$ a que se adiciona a estimativa do resíduo de curta duração $d''(k_j + i)$ para se obter o resíduo de curta duração reconstruído $d'(k_j + i)$

$$\begin{aligned} d'(k_j + i) &= e'(k_j + i) + d''(k_j + i), \\ j &= 0, 1, 2, 3, k_j = i + 40j, i = 0, \dots, 39 \end{aligned}$$

D. Codificação RPE

D.1 Filtro ponderador

É aplicado a cada subquadro um filtro FIR de fase linear que permite suavizar as transições entre subquadros. A a sua resposta impulsional é:

$$h(k) = [-0.0164, -0.0457, 0, 0.2507, 0.7008, 1, 0.7008, 0.2507, 0, -0.0457, -0.0164]$$

e a sua resposta impulsional e na frequência estão na figura 10:

Note que a convolução entre o filtro e o subsegmento produzem em princípio $40 + 11 - 1 = 50$ amostras. Neste caso

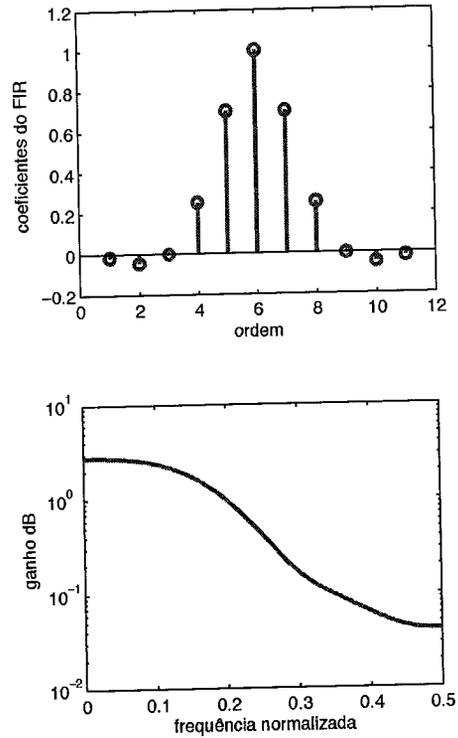


Fig. 10 - Resposta impulsional e na frequência do filtro ponderador

apenas se consideram as 40 amostras centrais da operação convencional:

$$x(k) = \sum_{i=0}^{10} h(i)e(k + 5 - i), k = 0, \dots, 39$$

com $e(k + 5 - i) = 0, k + 5 - i < 0$ ou > 39

D.2 Selecção de posição

O erro de predição depois de aplicado o filtro ponderador. $x(n)$, é decimado por um factor de 3, resultando em 4 seqüências com 13 impulsos regularmente espaçados

$$x_m(n) = x(m + 3n), m = 0, 1, 2, 3; n = 0, 1, \dots, 12$$

O critério de selecção é o de escolher a seqüência com maior energia de entre as 4 disponíveis:

$$m = \arg \max_m \left(E_m = \sum_{n=0}^{12} x_m^2(n) \right)$$

A posição resultante m é então codificada com 2 bits como \hat{M}

D.3 Codificação APCM

A seqüência de impulsos seleccionada é codificada usando um codificador de ganho adaptativo para a frente. Para cada seqüência de 13 amostras determina-se a amplitude do pico $x_{pico} = \max(x_m(n))$ valor que é quantificado usando um

quantificador logarítmico de 6 bits e que é utilizado para normalizar as amostras $x'_m(n) = x_m(n)/\hat{x}_{pico}$. Os valores normalizados $x'(n)$ são em seguida quantificados usando um quantificador uniforme de 3 bits

E. Resumo da alocação de bits

Para terminar, indica-se na tabela I a alocação dos bits neste codificador. Para cada 20 ms a informação codificada ocupa 260 bits, ou seja, o sistema GSM 6.10 tem um ritmo de transmissão de 13kb/s. Faz-se notar que o sistema GSM prevê em normas complementares que se transmita uma quantidade significativa de bits, destinados a controlo e correcção de erros.

III. DESCODIFICADOR

O descodificador GSM 6.10 está representado pelo seu diagrama de blocos na figura 11

A descodificação começa por reconstruir o resíduo de longa duração e' , a partir dos valores codificados transmitidos da posição da grelha (\hat{M}) e da aplicação de um APCM inverso aos valores característicos dos impulsos (\hat{x}_m e \hat{x}_{pico}). A utilização do preditor de longa duração caracterizado por T' e b' permite determinar a contribuição da predição de longa duração d'' que, adicionada ao resíduo de longa duração e' define o resíduo de curta duração reconstruído d' que será aplicado ao filtro de síntese. Este é implementado numa estrutura em *lattice* e usa os valores dos coeficientes de reflexão $k'(i)$ determinados a partir dos LAR transmitidos e após descodificação interpolação nas condições definidas no codificador. O sinal de saída sintetizado obtém-se depois da aplicação do filtro com uma função de transferência $H(z) = 1/(1 - \alpha_2 z^{-1})$, $\alpha_2 = 28180 * 2^{-15} = 0.86$ que faz a compensação da operação de pré-ênfase efectuada no codificador.

Todas as operações indicadas na descodificação são em tudo semelhantes às efectuadas na codificação, sendo portanto definidas em aritmética inteira. Finalmente refira-se que o sinal sintetizado é definido com 13 bits e será encaminhado para o dispositivo de conversão analógica ou para o PCM não uniforme a 64kbit/s.

IV. NOTAS FINAIS

A figura 12 mostra o resultado da aplicação do codificador GSM 6.10 a um curto registo de voz com a duração de 0.5 s. Pode verificar-se que a forma de onda do sinal descodificado segue com razoável aproximação o sinal original. Nos espectrogramas já se nota uma clara perda das altas frequências acima dos 3kHz, o que não altera a qualidade da voz segundo os critérios da comunicação telefónica. Apresenta-se também o sinal de erro, definido pela diferença $x(n) - x'(n)$ entre o sinal original $x(n)$ e o sinal descodificado $x'(n)$. Se se definir distorção da codificação como $D = \log_{10}(\sum_n x(n)^2 / \sum_n (x(n) - x'(n))^2)$, terá um valor para este caso de 15.5 dB, valor este normal para segmentos desta duração quando codificados por este sistema. Contudo este ruído introduzido não altera a inteligibilidade pelo que o ouvinte classificará, em geral, o som ouvido depois de descodificado como de boa qualidade. Nos testes de qualidade efectuados do tipo MOS (*Mean Opinion*

Score) a fala, depois de processada por este codificador, recebe uma pontuação superior a 4 numa escala que vai desde o 1 para ininteligível até 5 para qualidade natural

A codificação/descodificação foram obtidas utilizando o programa SoX que é um programa *freeware* suportado em diversas plataformas (Windows, Linux, MacOS X, etc.) que permite a conversão de ficheiros audio entre vários formatos e ainda a aplicação de efeitos sobre os sinais sonoros gravados. O leitor interessado poderá encontrá-lo em sox.sourceforge.net.

REFERÊNCIAS

- [1] W.C. Chu, *Speech Coding Algorithms Foundation and Evaluation of Standardized Coders*, John Wiley and Sons, New Jersey, 2003.
- [2] A.M. Kondoz, *Digital Speech Coding for Low Bit Rate Communications Systems*, John Wiley and Sons, England, 1994.
- [3] G. Fant, *Acoustic Theory of Speech Production*, Mouton and Co., sGravenhage, 1959.
- [4] G. Fant, "The acoustics of speech", em *Proceedings of third International Congress on Acoustics*. 1960, Elsevier.
- [5] J. R. Deller, J. H. L. Hansen, e J. Proakis, *Discrete Time Processing of Speech Signals*, IEEE Press, New York, 2000.
- [6] John Makhoul, "Linear prediction: A tutorial review", *Proceedings of IEEE*, vol. 63, pp. 561-580, 1975.
- [7] P. Kronn e R.J. Sluyter, "Regular-pulse excitation. a novel approach to effective and efficient of multipulse coding of speech", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 5, pp. 1054-1063, 1986.
- [8] B.S. Atal e J.R. Rench, "A new method of lpc excitation for producing natural sounding speech at low bit rates", em *Proceedings of ICASP*. ICASP, 1982.
- [9] European Telecommunications Standards Institute, *Digital cellular telecommunications system (Phase 2); Full rate speech; Part 2: Transcoding; (GSM 06.10 version 4.2.1)*, ETSI, France, 2000.
- [10] P. Strobach, "New forms of Levinson and Schur algorithms", *IEEE Signal Processing Magazine*, vol. 8, 1991.

TABELA I
ATRIBUIÇÃO DE BITS NO CODIFICADOR GSM 6.10 RPE-LTP

parâmetro	nº/quadro	resolução(bit)	nº bits/quadro
LPC(LAR)	8	6, 6, 5, 5, 4, 4, 3, 3	36
Período (\hat{T})	4	7	28
Ganho de longa duração(\hat{b})	4	2	8
Posição (\hat{M})	4	2	8
Amplitude do pico (\hat{x}_{pico})	4	6	24
Amplitude das amostras (\hat{x}_m)	4 * 13	3	156
		Total	260

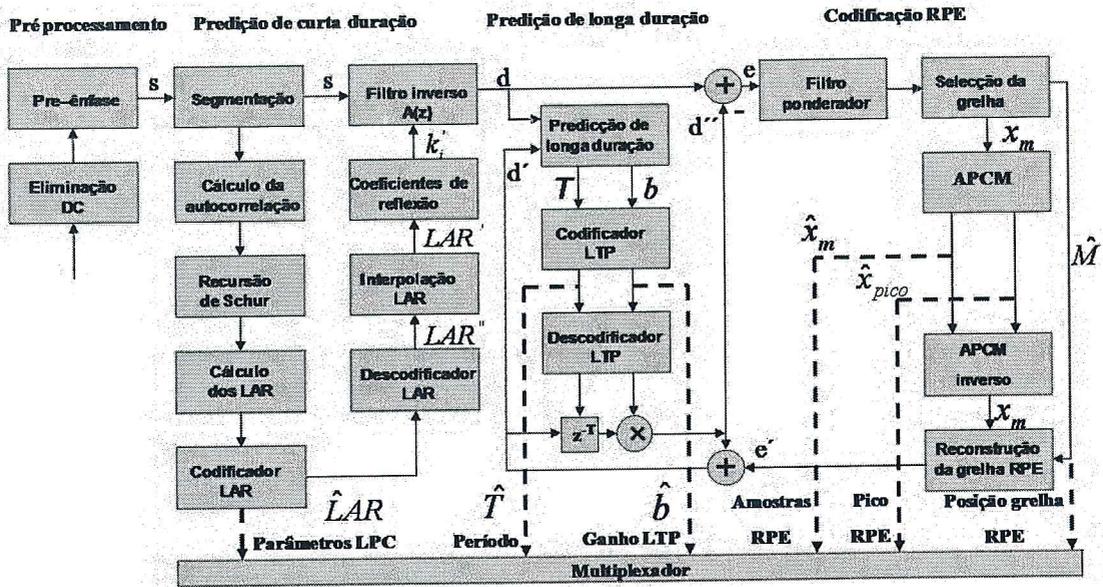


Fig. 6 - Diagrama de blocos do codificador GSM 6,10

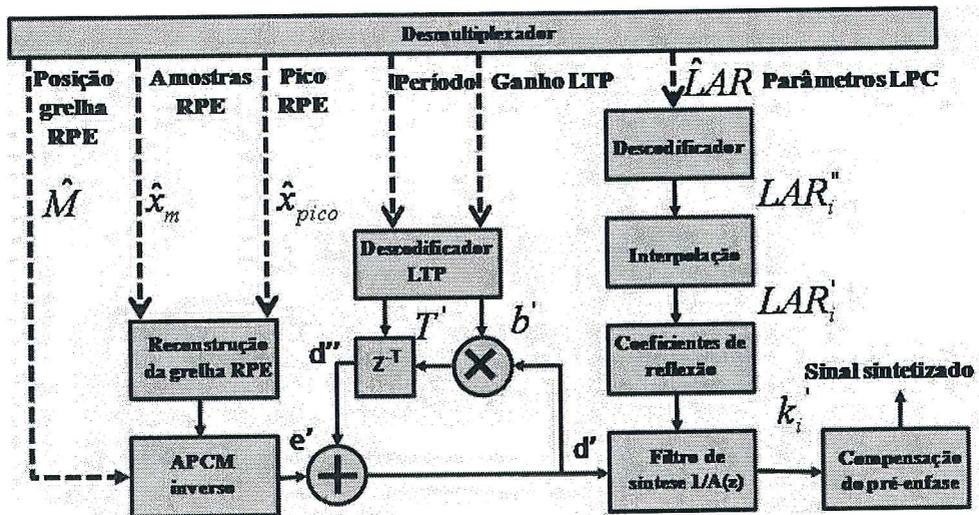


Fig. 11 - Decodificador GSM 6.10

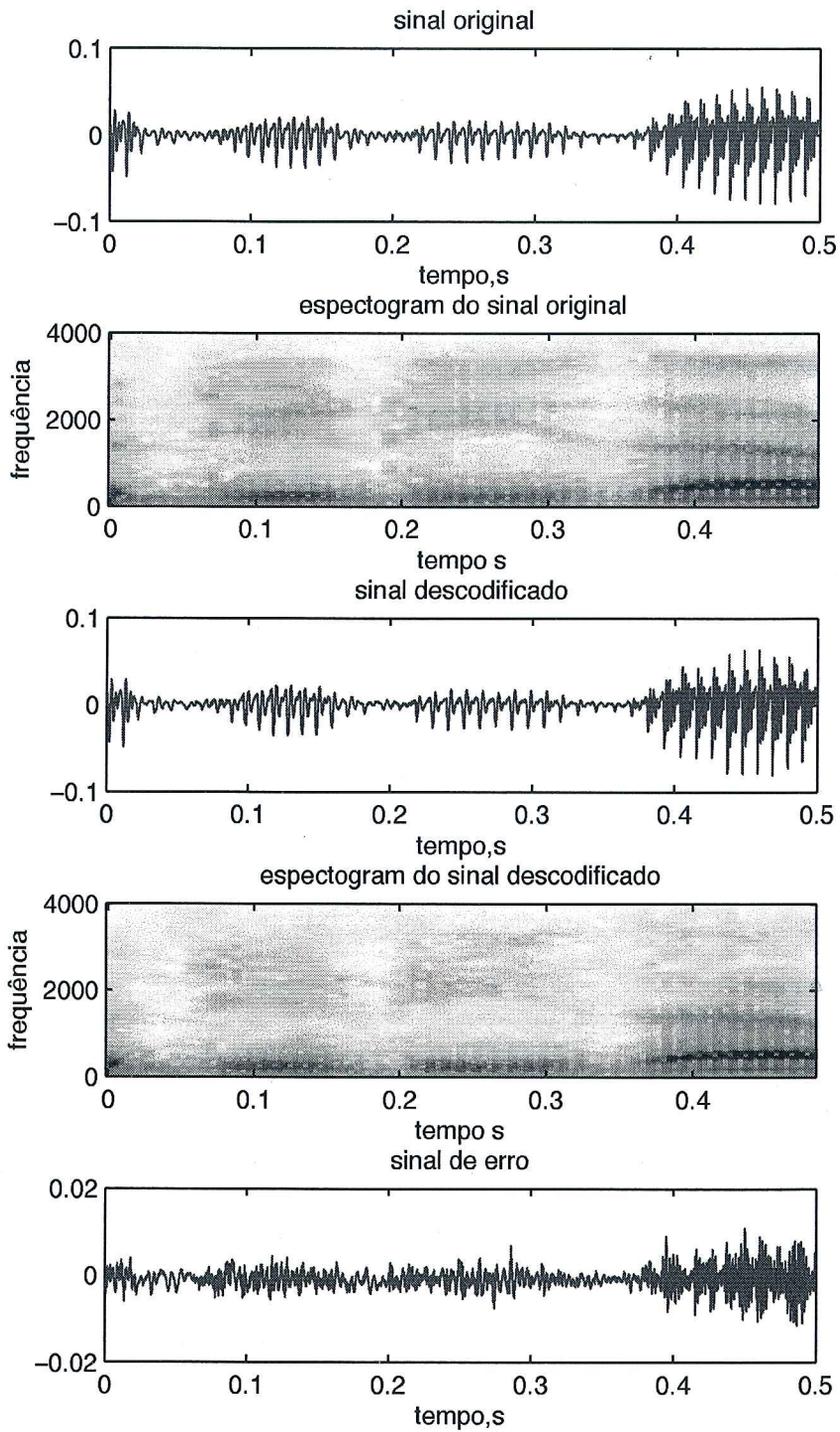


Fig. 12 - Exemplo de codificação-descodificação GSM6.10