Estimação do Espectro – Base em Codificação de Voz por Transformada Adaptativa

Abraham Alcaim, J. Roberto Boisson de Marca e Carlos F. B. Jaccoud

Em codificação por transformada adaptativa, a estimativa do espectro-base tem um papel fundamental sobre o desempenho do sistema, pois ela é utilizada tanto para alocação de bits como para adaptação dos quantizadores. Neste trabalho são examinadas detalhadamente três técnicas de estimação do espectro-base em codificação de voz por transformada adaptativa a 16 kbit/s. A análise é feita através de resultados de testes de escuta informais. Além disso, são comparadas formas de onda e espectros do sinal original e do sinal reconstruído e fornecidas medidas objetivas de desempenho para três sentenças, tanto utilizando voz masculina como voz feminina.

1. Introdução

Nos últimos anos tem-se notado um grande interesse no desenvolvimento de codificadores digitais de sinais de voz para a taxa de 16 kbit/s. Esse interesse tem sido motivado por uma grande variedade de aplicações, dentre as quais a possibilidade de se utilizar um maior número de usuários em sistemas de comunicações limitados em faixa ou a possibilidade de se armazenar uma maior quantidade de informação de voz para aplicação em diversos tipos de serviços automáticos.

Um dos métodos de codificação digital de voz, que fornece bom desempenho para a taxa de 16 kbit/s, é o método de codificação por transformada adaptativa ATC ("Adaptive Transform Coding") [1] . Esta técnica tem despertado interesse também para outras aplicações como, por exemplo, a armazenagem de sinais de áudio. Em um trabalho recente [2] foi alcançada uma economia da ordem de 75% sobre o PCM ("Pulse Code Modulation") para armazenagem de sinais de voz e música.

A. Alcaim e J. R. B. de Marca são professores do Centro de Estudos em Telecomunicações da PUC/Rio, 22453, Rio de Janeiro, RJ.
C. F. B. Jaccoud é engenheiro do Departamento de Desenvolvimento de Recursos Humanos, EMBRATEL, Rua da Assembléia 10, 20011, Rio de Janeiro, RJ.

O diagrama em blocos do sistema ATC é mostrado na **Fig. 1**. Nesse sistema o sinal de voz é dividido, inicialmente, em blocos de N amostras {s(1), ..., s(N)}, sendo cada um dos quais normalizado através de uma estimativa do valor r.m.s. quantizado ($\hat{o}_B$) das amostras do bloco. Ou seja, cada amostra s(i) dentro de um bloco assume o valor $x_i = s(i)/\hat{o}_B$ após a normalização. Essas amostras formam um vetor $\mathbf{x} = (x_1, \ldots, x_N)T$ que é transformado para $\mathbf{y} = (y_1, \ldots, y_N)T$, onde T indica o operador transposto. Essa transformação de **x** em **y** é usualmente feita através de uma transformada cosseno discreta DCT ("Discrete Cosine Transform") de N pontos [**2**]. Os elementos de **y** representam componentes de freqüência do sinal de voz que são codificadas separadamente através de codificadores PCM adaptativos (APCM). Na reconstrução do sinal, os elementos de **y** são decodificados e transformados por uma DCT inversa em um vetor $\hat{\mathbf{x}}$. Este valor é multiplicado pelo fator de normalização $\hat{o}_B$ de modo a reconstruir as amostras do sinal.

O método ATC tem a vantagem de explorar critérios baseados nos modelos de produção e percepção da voz, sem contudo tornar os algoritmos totalmente dependentes destes modelos, como no caso de codificadores paramétricos [**4**]. Assim, às componentes de freqüência mais baixa, por exemplo, onde a estrutura de periodicidade dos sons sonoros e a primeira formante precisam ser cuidadosamente preservadas, é em geral alocado um maior número de bits por amostra. Nesses codificadores, a alocação de bits é feita de forma dinâmica, de modo a acompanhar apropriadamente as variações espectrais da voz ao longo tempo.

Os algoritmos utilizados para alocação de bits e adaptação dos quantizadores em ATC constituem os elementos fundamentais do sistema, uma vez que deles depende, em grande parte, o desempenho alcançado. Esses algoritmos, por sua vez, dependem essencialmente de um conjunto de N parâmetros $\{\sigma(1), \ldots, \sigma(N)\}$, que correspondem a uma estimativa dos desvios padrão dos coeficientes $\{y_1, \ldots, y_N\}$ no domínio da transformada. Tal conjunto de parâmetros pode ser visto também como um modelo do espectro do sinal, chamado espectro-base.

Uma estimativa do espectro-base deve ser transmitida, como informação paralela, a cada bloco específico de amostras de voz. Note-se, porém, que para manter a eficiência do sistema de codificação, essa informação paralela deve conter o menor número possível de parâmetros, de modo a não prejudicar a quantização da informação principal $(y_1, \ldots, y_N)$. A partir daqueles parâmetros é, então, feita a estimação do espectro-base.

No presente trabalho são examinadas, em detalhe, três técnicas de estimação do espectro-base em sistemas ATC a 16 kbit/s. Nos sistemas aqui analisados, N=256 e a quantização dos coeficientes no domínio da transformada é feita após normalização dos coeficientes pela estimativa do espectro-base. Ou seja, o que é quantizado é a razão $y_i/\sigma(i)$.

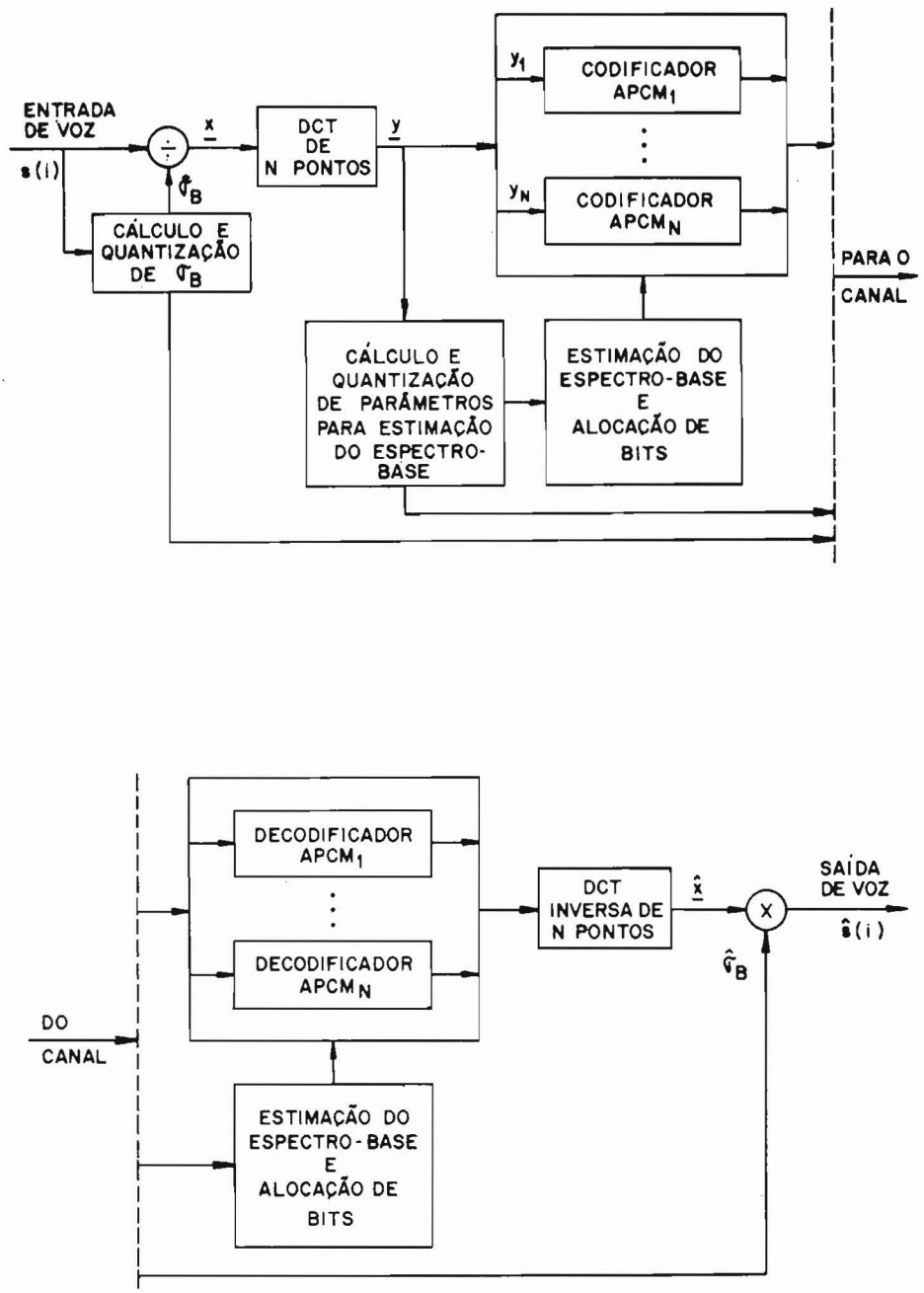

Figura 1. Sistema ATC.

O número de bits n(i) alocado para a quantização de cada coeficiente é determinado a partir da expressão que minimiza o erro médio quadrático de quantização global [**1**] , [**5**] , dada por

$$n(i) = \bar{n} + \log_2 \sigma(i) - \frac{1}{256} \sum_{i=1}^{256} \log_2 \sigma(i), \quad i = 1 \ldots, 256 \qquad (1)$$

onde $\bar{n}$ é o número médio de bits por amostra dedicados para transmissão da informação principal (vetor **y**). Note-se que para se obter uma taxa de bits de 16 kbit/s com valores inteiros para n(i) deve-se estabelecer algum procedimento adicional para a distribuição dos bits. Aqui foi utilizada uma regra em que $0 \leqslant n(i) \leqslant 5$ e quando bits precisam ser retirados, isso é feito dos coeficientes com menor número de bits, enquanto, quando bits precisam ser adicionados isso é feito aos coeficientes com maior número de bits [**5**]. Uma regra oposta a essa [**6**] (ou seja, coeficientes com menor número de bits são privilegiados e com maior número de bits são prejudicados) foi também considerada, porém forneceu um desempenho um pouco inferior.

As seções 2 e 3 deste artigo apresentam, respectivamente, as técnicas de estimação do espectro-base e a avaliação de seu desempenho. Na Seção 4 são fornecidas as principais conclusões do trabalho.

2. Estimação do Espectro-Base

Nesta seção serão descritas as três técnicas de estimação do espectro-base consideradas neste trabalho. Serão fornecidos ainda detalhes relativos à discretização dos parâmetros a partir dos quais o espectro-base é obtido, assim como as características dos quantizadores usados na codificação da informação principal ($y_1, \ldots, y_{256}$). Os parâmetros associados ao espectro-base e o valor r.m.s. do bloco de amostras (σ_B) formam a informação paralela.

Técnica de Zelinski e Noll: ATC-ZN

A técnica de Zelinski e Noll [**1**] supõe um espectro-base suave, explorando a similaridade de componentes de freqüência (no domínio da DCT) adjacentes através de uma média dos valores quadráticos de componentes vizinhas. Mais especificamente, essa técnica consiste em inicialmente dividir o bloco de 256 coeficientes em 16 sub-blocos de 16 coeficientes. Os parâmetros associados ao espectro-base são exatamente os valores quadráticos $\beta_1^2, \beta_2^2, \ldots, \beta_{16}^2$ desses 16 sub-blocos, os quais são quantizados através de quantizadores logaritmicos (lei μ=100) de 3 bits por amostra. O parâmetro σ_B é também quantizado, com o mesmo tipo de quantizador, resultando em $\hat{\sigma}_B$.

A estimativa do espectro-base é obtida através de uma interpolação linear do logaritmo (base 2) dos parâmetros quantizados dos 16 sub-blocos. Tomando o antilog, obtém-se finalmente os coeficientes $\{\sigma(1),...,\sigma(256)\}$ que formam o espectro-base. Note-se, então, que nesse sistema o espectro-base, composto de 256 coeficientes, é obtido a partir de apenas 16 parâmetros [**5**]. Assim, para cada bloco de amostras, (16+1)x3=51 bits são gastos para transmitir a informação paralela e 512-51=461 bits são usados para a informação principal. Essa última é codificada utilizando-se quantizadores ótimos projetados para entrada Gaussiana.

Técnica Invariante em Altas Freqüências: ATC-IAF

Essa técnica corresponde a uma simplificação da anterior, com o objetivo de fornecer mais bits à informação principal. Como é sabido e facilmente observável, as componentes de freqüência mais baixa, que são em geral as que contêm maior energia, recebem a maioria dos bits disponíveis. Além disso, as componentes de freqüência mais alta podem, em geral, ser codificadas da forma mais grosseira do que as componentes de freqüência mais baixa. Por esses motivos, achou-se conveniente examinar a possibilidade de codificar apenas a metade inferior dos 16 parâmetros asssociados ao espectro-base, deixando-se a metade superior invariante e igual a seus valores médios dados por

$\beta_9=7,4$ $\beta_{10}=7,2$ $\beta_{11}=7,0$ $\beta_{12}=5,5$

$\beta_{13}=3,4$ $\beta_{14}=2,6$ $\beta_{15}=1,9$ $\beta_{16}=1,1$

Com isso, além de se evitar o cálculo de metade dos parâmetros associados ao espectro-base, consegue-se fornecer mais bits à informação principal. Nesse caso, (8+1)x3=27 bits por bloco são gastos na informação paralela e 512-27=485 bits por bloco na principal.

Técnica Cepestral: ATC-CEP

A terceira técnica a ser considerada neste trabalho é uma técnica específica para sinais de voz [**6**] , que utiliza um modelo de produção da voz que separa a envoltória suave do espectro e a excitação [**4**] . A primeira característica é representada por um pequeno número de parâmetros e a segunda é representada por parâmetros que levam em conta a estrutura quase-periódica de excitação dos sons sonoros. A idéia de utilização dessa técnica em codificadores ATC foi proposta em 1978 por Tribolet e Crochiere [**7**] , com base em um modelo LPC ("Linear Predictive Coding"). No presente trabalho será usada uma idéia semelhante. Porém, a extração dos parâmetros

que representam a envoltória espectral e a excitação será feita através de processamento homomórfico ou "cepestral". [6]

O "cepestro" é definido como a transformada de Fourier[1] inversa do logaritmo do módulo da transformada de Fourier [8] . A propriedade de interesse do cepestro é que apenas sua porção baixa (primeiros coeficientes) é suficiente para representar a envoltória espctral da voz. Além disso, a quase-periodicidade associada à excitação de sons sonoros aparece no cepestro como picos acentuados em sua porção alta. São extamente esses parâmetros (porção baixa e porção alta do cepestro) que são utilizados nesta técnica para estimação do espectro-base em ATC. Um exemplo de cepestro correspondente a um segmento (bloco) de amostras de um sinal de voz é mostrado na **Fig. 2(a)**. O método será aqui denotado por ATC-CEP, e sua descrição será feita a seguir.

A DCT do bloco de amostras é aqui representada por $\{ y_i ; i=1, \ldots . 256\}$. Dessa forma, o cepestro é definido por

$$\{c(i)\} = DCT^{-1} \{ \log_2 |y_1| \} \quad (2)$$

onde DCT^{-1} representa DCT inversa. Os primeiros N_e (12 a 14) coeficientes do cepestro são utilizados para caracterizar a envoltória suave do espectro da voz. Com relação aos parâmetros associados à excitação, é utilizado um procedimento que consiste essencialmente da busca de um pico principal na porção alta de $|c(i)|$ e de dois outros picos: um contíguo ao pico principal e outro numa vizinhança do dobro ou da metade de sua posição [6] . Esse procedimento pode ser descrito pelos seguintes passos:

(1.°) Determina-se posição e amplitude do pico principal:

$$\ell_1 = \max_i \{ |c(i)| ; i=N_e+1, N_e+2, \ldots , 256 \} \quad (3)$$

$$\alpha_1 = c(\ell_1) \quad (4)$$

(2.°) Determina-se posição e amplitude do pico vizinho:

$$\ell_2 = \max_i \{ |c(i)| ; i=\ell_1-1 , \ell_1+1\} \quad (5)$$

$$\alpha_2 = c(\ell_2) \quad (6)$$

(3.°) Determina-se a posição ℓ_3 e amplitude α_3 do terceiro pico através das seguintes relações:

1. Aqui será utilizada transformada cosseno.

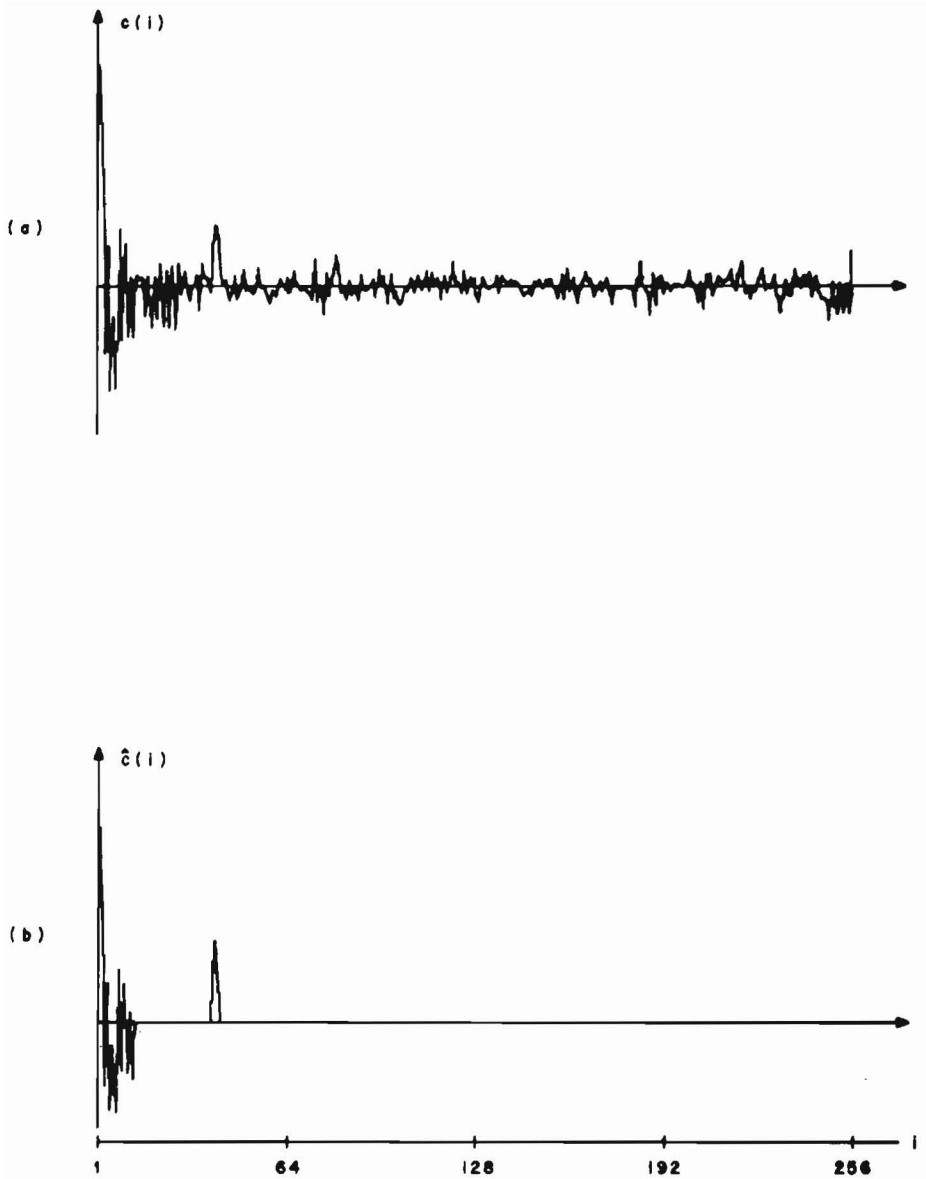

Figura 2. (a) Cepestro para um segmento de voz sonora falado por um locutor masculino; (b) Conjunto de coeficientes c(i) utilizado na construção do espectro-base em um codificador ATC-CEP.

$$k_m = \max_i \{ |c(i)| \; ; \; i = \llcorner \ell_1/2 \lrcorner - 8, \ldots, \llcorner \ell_1/2 \lrcorner + 7 \} \quad (7)$$

onde $\llcorner x \lrcorner$ é o maior inteiro menor ou igual a x

$$k_d = \max_i \{ |c(i)| \; ; \; i = (2\ell_1) - 8, \ldots, (2\ell_1) + 7 \} \quad (8)$$

$$\ell_3 = \max_i \{ |c(i)| \; ; \; i = k_m, k_d \} \quad (9)$$

$$\alpha_3 = c(\ell_3) \quad (10)$$

A quantização dos parâmetros que definem os primeiros N_e coeficientes do cepestro, assim como α_1, α_2 e α_3 dadas por (4), (6) e (10) é feita com quantizadores logaritmicos lei μ=50, de 3 bits/amostra e com nível máximo igual a 2,8. Os parâmetros do cepestro associados às posições dos picos que definem a excitaçao do modelo, de acordo com (3), (5) e (9), são quantizados diretamente com 8 bits para ℓ_1, 1 bit para ℓ_2 e 5 bits para ℓ_3. Para o valor rms do bloco de amostras, σ_B, é utilizado um quantizador logaritmico, lei μ=100, de 5 bits/amostra e com nível máximo igual a 600.

Fora os N_e coeficientes do cepestro $\{c(i)\}$ que representam a envoltória espectral e os 3 coeficientes que modelam a excitação, os outros valores de c(i) são feitos iguais a zero. Além disso, os 3 coeficentes que modelam a excitação têm sua amplitude modificada por um fator de ganho fixo G, com o objetivo de casar melhor as características espectrais do modelo com aquelas da voz original. O efeito da variação desse fator sobre o desempenho será também avaliado neste trabalho.

Resultam, então, após a quantização dos coeficientes do cepestro e as operações que se acabou de mencionar, um novo conjunto de coeficientes que representa o cepestro da voz para o bloco de amostras em consideração. Esse conjunto que será denotado por $\{\hat{c}(i)\}$ é exemplificado na **Fig. 2(b)** para o mesmo bloco de amostras para o qual foi calculado o cepestro da **Fig. 2(a)**.

Tanto no codificador, como no decodificador, é feita a estimação do espectro-base $\{\sigma(i)\}$ a partir $\{\hat{c}(i)\}$. Para isto, são realizadas operações inversas às expressas em (2), ou seja,

$$\sigma(i) = 2^{DCT\{\hat{c}(i)\}} \quad (11)$$

A quantização das componentes de frequência y_i é feita, nesse caso, normalizando-as primeiramente por $2\sigma(i)$. O fator 2 se tornou necessário pelo fato

de $y_i/\sigma(i)$ se encontrar dentro de uma faixa de amplitudes entre aproximadamente −2 e +2 conforme mostra a **Fig. 3.** Finalmente, após a normalização, a quantização é feita utilizando-se quantizadores ótimos de Max, projetados para distribuição Gaussiana de média nula e variância unitária.

3. Avaliação de Desempenho

O desempenho dos codificadores por transformada adaptativa descritos na seção anterior foi avaliado através de simulação em computador digital. Foram utilizadas como material de entrada, para as simulações, as sentenças:

1. O bispo lançou a cabeça para trás e sorriu.

2. Como o ato de fé, o ato de amor é uma rendição.

3. Para o biógrafo, a personalidade de um homem é importante.

Cada uma das sentenças foi falada por um homem e uma mulher, ambos adultos.

Antes de ser processado pelos codificadores em estudo, o sinal elétrico correspondente às sentenças mencionadas sofreu uma filtragem passa-baixas em conformidade com o padrão CCITT (Comitê Consultivo Internacional de Telefonia e Telegrafia), foi amostrado a uma taxa de 8.000 amostras por

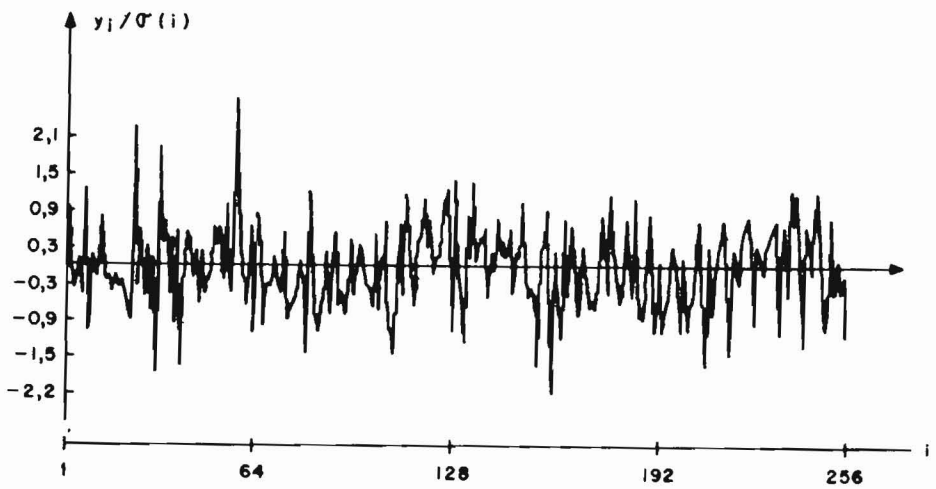

Figura 3. Variação dos coeficientes da transformada y_i já normalizados pelo espectro-base $\sigma(i)$ para um segmento de voz.

segundo e, em seguida, foi digitalizado empregando-se um quantizador uniforme com resolução de 12 bits/amostra.

As medidas de desempenho adotadas neste trabalho foram a razão sinal-ruído segmentada (RSS) [**9**] e um teste subjetivo informal, o qual resultou em um número indicador, para cada codificador, da média de opiniões dos ouvintes. Antes de apresentar valores numéricos para essas medidas, é bastante útil, para compreensão do comportameto dos diversos codificadores, fazer uma análise comparativa das características nos domínios do tempo e da freqüência em diferentes pontos dos codificadores.

Inicialmente a **Fig. 4(a)** apresenta, em base logaritmica, a magnitude dos coeficientes da transformada DCT de um bloco de 256 amostras do sinal de entrada. Como foi visto, o chamado espectro-base é justamente uma estimativa para o valor rms dos coeficientes da DCT. Essa estimativa serve, então, para determinar os parâmetros dos quantizadores que serão empregados na discretização de cada coeficiente. As **figuras 4(b)** até **4(d)** mostram os espectros-base, também em base logaritmica, produzidos pelas três técnicas descritas na Seção 2. Como pode ser visto, o método cepestral fornece uma aproximação para a transformada do sinal que pode ser considerada bastante boa, captando de forma adequada a estrutura fina (associada ao período fundamental) do espectro. Já o método de Zelinski e Noll produz um espectro-base que acompanha de forma razoável a envoltória da DCT mas perde completamente a estrutura periódica. Por último, a técnica com invariância nas altas freqüências consegue um espectro-base bastante semelhante ao ATC-ZN até aproximadamente metade da faixa. A partir daí no entanto, o espectro-base não apresenta semelhança, mesmo em termos de envoltória, com c(i). Isto de certa forma podia ser esperado, já que na região de freqüências mais altas, esta técnica constrói o espectro-base a partir de valores médios da DCT e portanto, tem uma dependência pequena com o valor real da transformada do sinal para um determinado bloco.

Como será visto mais adiante, esta má representação do espectro-base por parte do ATC-IAF é fatal para o seu desempenho, que é bastante inferior ao dos demais ATC. A característica de preservação da estrutura fina espectral apresentada pelo ATC-CEP se reflete também na distribuição de bits entre os coeficientes, conforme mostra **Fig. 5(a)**. Os outros dois métodos não apresentam esta propriedade e a distribuição de bits acompanha de certa forma a envoltória da DCT (ver **figuras 5(b)** e **5(c)**) até aproximadamente a metade da faixa. Note que a motivação do método ATC-IAF de garantir um número maior de bits para a faixa mais alta, sem a necessidade de transmissão dos parâmetros dessa faixa, é até certo ponto realizada. No entanto, não existe a garantia de que esses bits estejam na posição correta.

O grau de fidelidade com que um determinado sistema de codificação consegue reproduzir o sinal original pode ser avaliado tanto no domínio do tempo

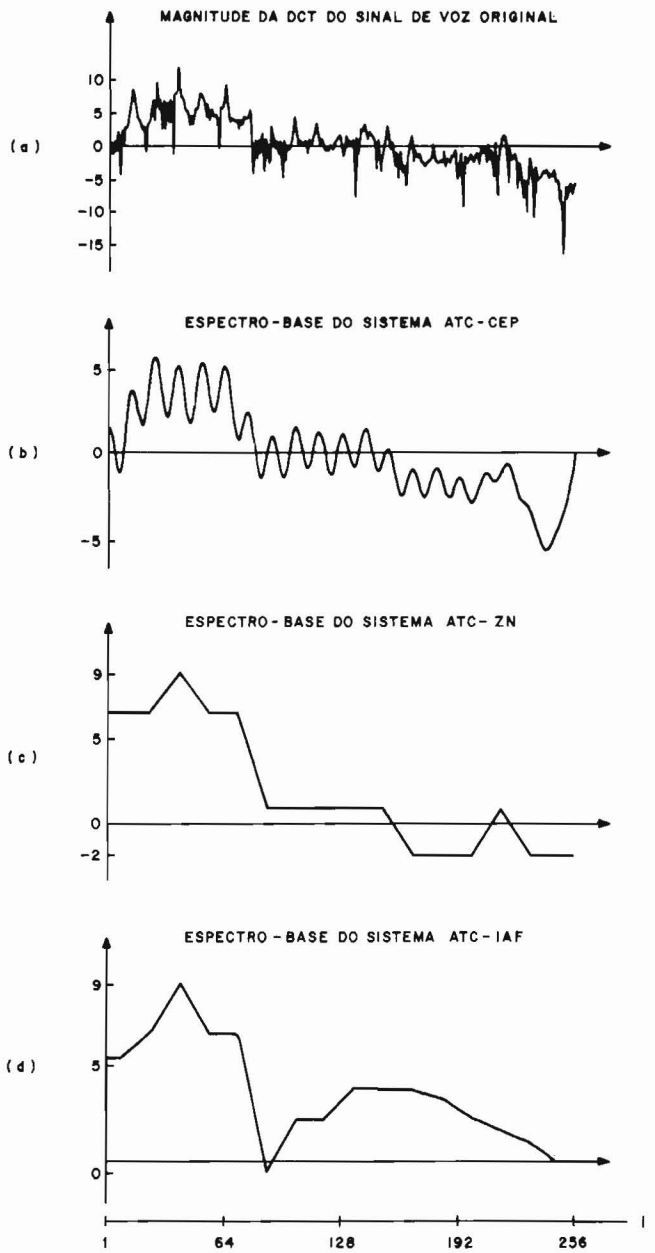

Figura 4. Magnitude da DCT do sinal original (a) e espectros-base gerados pelos codificadores ATC-CEP (b), ATC-ZN (c) e ATC-IAF (d), todos representados no domínio logaritmico.

quanto no da freqüência. O espectro de potência do bloco de amostras do sinal original que foi objeto das **figuras 4** e **5** é mostrado na **Fig.6(a)**. As **figuras 6(b)** até **6(d)** contêm os espectros dos sinais reconstruídos por um sistema ATC-CEP (N_e = 12), e pelos codificadores ATC-ZN e ATC-IAF. Como pode ser visto, todas as três técnicas atingem uma boa representação espectral para freqüências até 2 kHz. Na região entre 2 e 2,5 kHz apenas o sinal produzido pelo ATC-CEP consegue acompanhar de forma razoável o espectro do sinal original. A partir de 2,5 kHz as características espectrais produzidas pelos três codificadores divergem significativamente do original, sendo que a pior situação é novamente a do ATC-IAF.

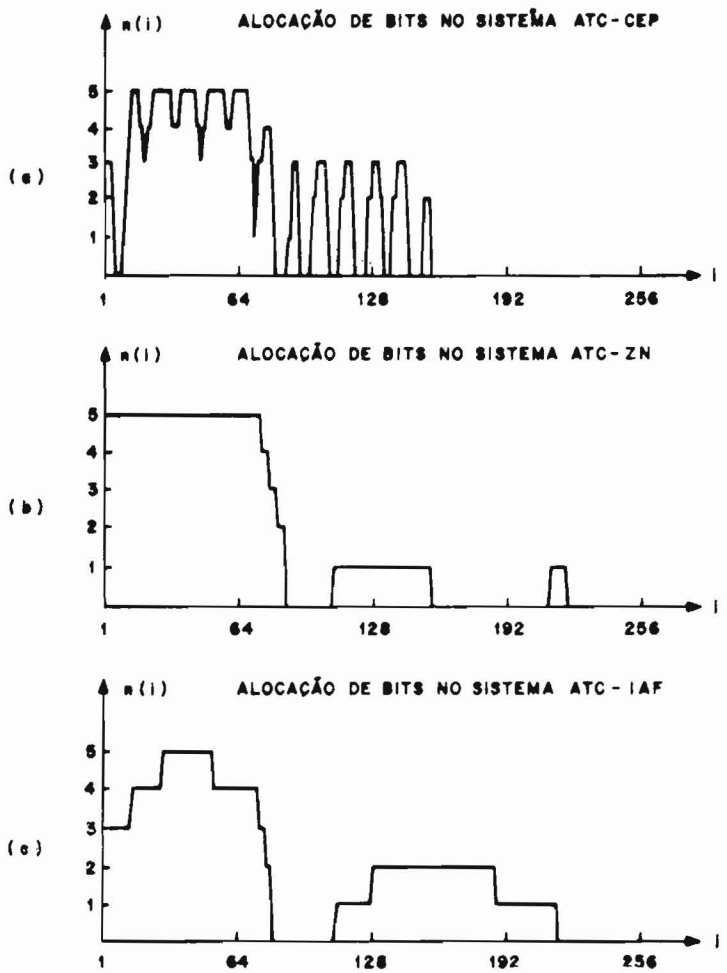

Figura 5. Alocação de bits produzida pelos codificadores ATC-CEP (a), ATC-ZN (b) e ATC-IAF (c).

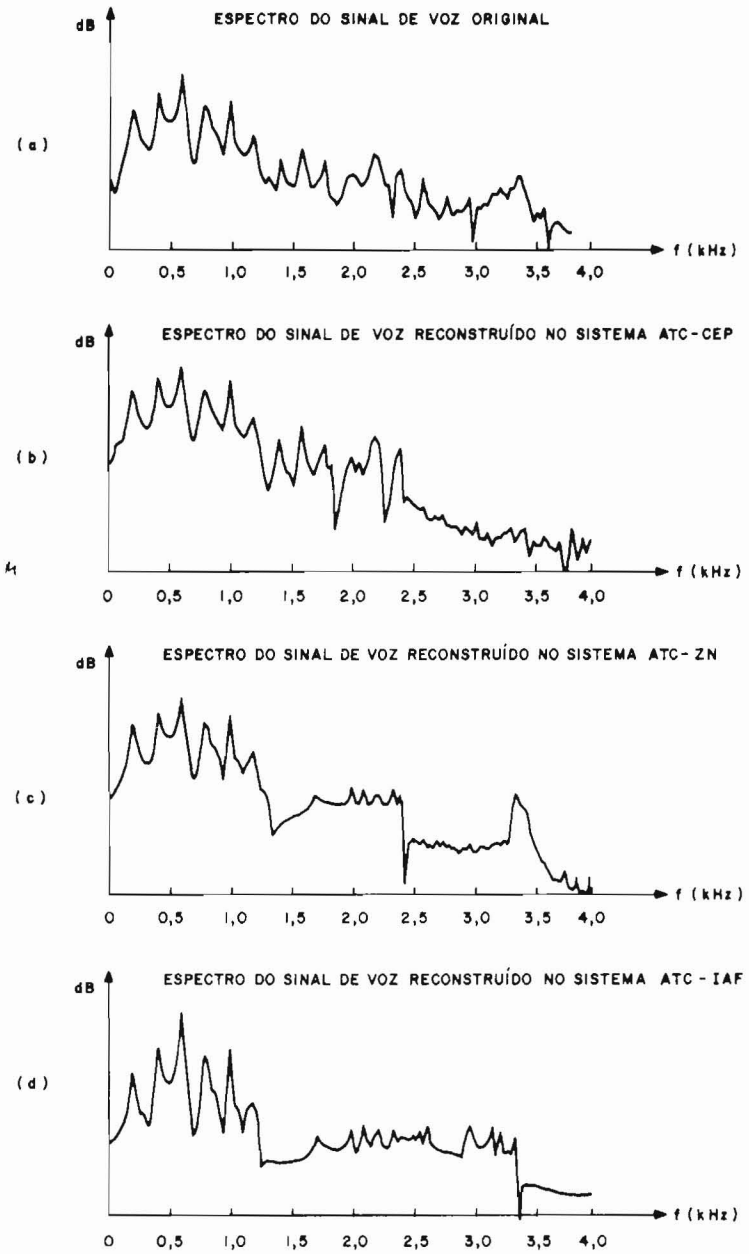

Figura 6. Espectros de freqüência do sinal original para voz masculina (a) e dos sinais reconstruídos pelos codificadores ATC-CEP (b), ATC-ZN (c) e ATC-IAF (d).

O bom desempenho dos três métodos para segmentos de voz com alto conteúdo de baixas freqüências (típico de sons sonoros) também pode ser observado a partir do domínio do tempo. As formas de onda correspondentes ao sinal original e aos sinais de saída dos três sistemas são mostradas na **Fig. 7**. Essas formas de onda dizem respeito ao mesmo segmento de voz sonora para o qual os espectros de freqüência apresentados na **Fig. 6** foram calculados.

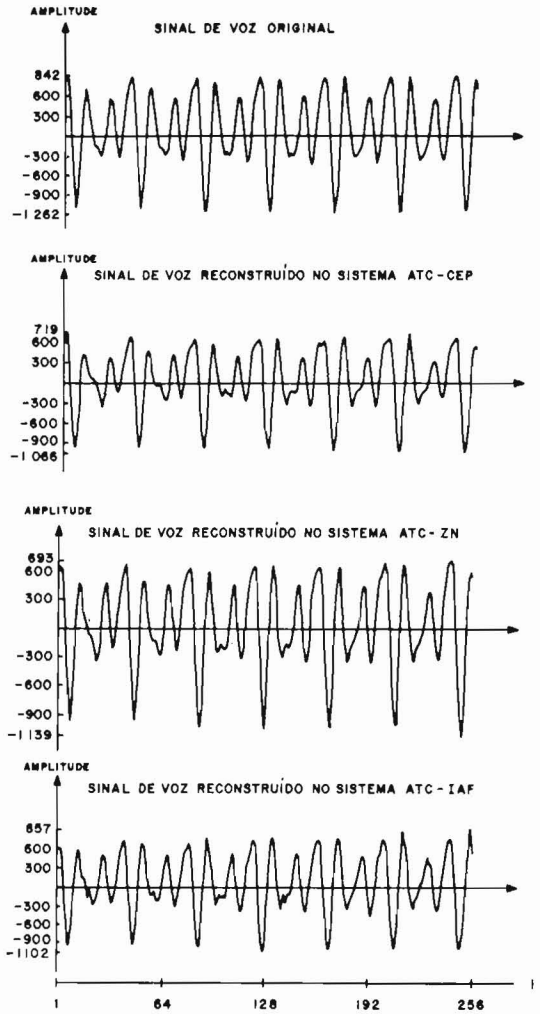

Figura 7. Formas de onda do sinal original para voz masculina (a) e dos sinais reconstruídos pelos codificadores ATC-CEP (b), ATC-ZN (c) e ATC-IAF (d).

Os resultados numéricos obtidos para as medidas de desempenho mencionadas anteriormente (RSS e indicador de opinião) são mostrados nas **tabelas 1 e 2.** Algumas observações de interesse podem ser feitas com relação à **Tabela 1** que ilustra a variação do desempenho, em termos de RSS, de todos os codificadores ATC em estudo e também de dois log-PCM, em função de cada frase e de cada locutor. As três primeiras colunas apresentam os resultados quando as 3 frases mencionadas no início desta seção são faladas por um locutor masculino (H). As três colunas seguintes contém os resultados correspondentes à situação em que o locutor é do sexo feminino (M). A última coluna indica o valor de RSS para cada codificador, quando são considerados os seis estímulos (H1, H2, H3, M1, M2 e M3)

CODIFICADOR	FRASE						
	H1	H2	H3	M1	M2	M3	MÉDIA
log-PCM, 64 kbit/s	37,0	36,7	37,3	36,9	39,2	38,8	37,7
log-PCM, 48 kbit/s	25,5	26,2	25,6	25,8	27,2	26,6	26,2
ATC-IAF	13,6	15,3	14,7	12,3	13,5	14,0	13,9
ATC-ZN	13,9	15,0	15,3	13,5	13,8	13,9	14,2
ATC-CEP N_e=12, G=0,8	11,7	13,5	14,4	12,6	13,4	16,3	13,7
ATC-CEP N_e=13, G=0,8	13,2	13,6	15,3	13,0	13,6	16,3	14,2
ATC-CEP N_e=14, G=0,8	14,0	13,8	16,0	13,1	13,7	16,4	14,5
ATC-CEP N_e=12, G=1,0	12,4	13,9	14,7	13,4	14,2	17,4	14,3
ATC-CEP N_e=12, G=1,2	12,9	14,1	14,6	14,0	14,0	18,2	14,8

Tabela 1. Valores de razão sinal-ruído segmentada RSS (em dB) para 6 sentenças: três faladas por locutor masculino (H1, H2, H3) e as mesmas três faladas por uma mulher (M1, M2, M3).

CODIFICADOR	LOCUTOR								
	MASCULINO			FEMININO			MASCULINO/ FEMININO		
	M. O.	σ	I. C.	M. O.	σ	I. C.	M. O.	σ	I. C.
log PCM, 64kbit/s	4,53	0,55	0,19	4,13	0,78	0,26	4,33	0,70	0,16
log PCM, 48kbit/s	4,04	0,58	0,23	3,63	0.79	0,27	3,83	0,76	0,18
log PCM, 32kbit/s	2,00	0,46	0,16	1,51	0,54	0,18	1,76	0,56	0,13
SBC, 16kbit/s	2,50	0,76	0,26	2,63	1,05	0,36	2,56	0,91	0,21
ATC-IAF	2,18	0,56	0,19	1,76	0,63	0,21	1,97	0,63	0,15
ATC-ZN	3,78	0,68	0,23	3,10	0,95	0,32	3,44	0,89	0,21
ATC-CEP N_e=12, G=1,2	3,69	0,88	0,29	3,35	0,88	0,30	3,52	0,88	0,21
ATC-CEP N_e=12, G=1,2	3,69	0,88	0,29	3,35	0,88	0,30	3,52	0,88	0,21
ATC-CEP N_e13, G=0,8	3,58	0,80	0,30	3,46	0,81	0,28	3,52	0,85	0,20
ATC-CEP N_e14, G=0,8	3,71	0,85	0,27	3,07	0,71	0,24	3,39	0,81	0,19

Tabela 2. Resultados de teste subjetivo: valores médios de opinião (M. O.), desvio padrão (σ) e intervalo de confiança (I. C.) para nível de confiança de 95%.

Como era de se esperar a partir da análise feita anteriormente, o desempenho dos sistemas ATC-IAF e ATC-ZN para voz masculina é superior aos desempenhos respectivos para voz feminina. Este fato se deve a que a voz feminina contém uma maior quantidade de energia em altas freqüências, que não são bem reproduzidas por aqueles codificadores. De uma forma geral, os codificadores ATC encontraram maior dificuldade com a sentença número 1, possivelmente devido ao fato dela apresentar um maior número de fonemas fricativos e oclusivos.

A partir da **Tabela 1** vemos que dentre os codificadores ATC, o ATC-CEP é o que atinge melhor desempenho, enquanto o ATC-IAF é o que se mostra pior, resultado coerente com observações anteriores. No entanto, é necessário ter cuidado ao se utilizar a RSS como medida para avaliar o desempenho relativo de codificadores. Como será visto, através dos resultados do teste subjetivo, a qualidade do sinal reconstruído pelo sistema ATC-CEP, N_e = 12, G = 1,2, é muito mais próxima daquela do log-PCM com 6 bits por amostra (48 kbit/s) do que indicam os quase 12 dB de diferença em RSS. Da mesma forma, o desempenho do ATC-IAF é bastante inferior àquele do ATC-ZN, fato que não pode ser apreciado dos valores de RSS. Por outro lado, a RSS é útil na otimização dos parâmetros de um determinado codificador. Por exemplo, a indicação de que para o ATC-CEP, N_e = 12, o melhor valor para o ganho G é 1,2 foi confirmada pelos testes de audição. Obviamente a obtenção de valores de RSS é muito mais simples do que preparar e executar um teste subjetivo.

A **Tabela 2** apresenta os resultados de um teste subjetivo informal. Para este teste foi preparada uma fita na qual 13 codificadores foram utilizados para processar as três sentenças já apresentadas, faladas tanto pelo locutor masculino quanto pelo feminino. Os 78 estímulos foram então submetidos a 12 ouvintes que, para cada estímulo, atribuíam um grau entre 1 (inaceitável) e 5 (excelente). Na **Tabela 2**, as médias de opinião são apresentadas separadamente para as frases faladas pelo locutor masculino e para aquelas faladas por uma mulher e também a média incluindo todas as frases. A **Tabela 2** também fornece, para cada situação, o desvio padrão verificado no teste subjetivo e o tamanho do intervalo de confiança (I.C.) para a estimativa da média de opinião correspondente a um nível de confiança de 95%.[2] Como pode ser visto daquela tabela, os melhores sistemas ATC alcançaram médias de opinião em torno de 3,5. Este nível é próximo daquele fornecido pelo sistema log-PCM com taxa três vezes maior (48 kbit/s). Outro fato digno de nota é que os sistemas ATC apresentaram um desempenho bastante superior ao de um codificador em sub-bandas SBC ("Sub-Band Coder") com alocação de bits adaptativa e mesma taxa de bits (16 kbit/s) [**10**] . Focali-

2. Por exemplo, para o sistema log-PCM, taxa de 64 kbit/s e locutor masculino, a média de opinião está no intervalo 4,53 $\pm$ 0,19 com probabilidade 0,95.

zando agora apenas nos codificadores ATC pode-se observar que o sistema proposto por Zelinski e Noll embora mais simples e não dependente do modelo de produção da voz alcançou um desempenho praticamente equivalente àquele do ATC-CEP (é importante ressaltar que não foi feita uma busca ampla de valores para os parâmetros N_e e G usados no método cepestral, sendo portanto concebível a obtenção de um desempenho ainda superior para o ATC-CEP). Por outro lado, o teste subjetivo indicou que a qualidade do sinal reconstruído pelo ATC-IAF é bastante inferior àquela fornecida pelas duas outras técnicas ATC, em razão do seu espectro-base não acompanhar adequadamente o sinal transformado na região de altas freqüências. A baixa qualidade do ATC-IAF é percebida como um abafamento da voz. Esse efeito foi considerado bastante desagradável pelos ouvintes, quando comparado aos outros métodos.

Deve-se mencionar ainda que para todos os métodos ATC é possível obter uma melhoria adicional de desempenho se for utilizada uma transformada que reduza o efeito de descontinuidade entre blocos [**11**] .

4. Comentários Finais

Este trabalho apresentou um estudo comparativo do desempenho de três técnicas de codificação por transformada adaptativa, que diferem fundamentalmente no método de obtenção do espectro-base. A análise do comportamento dos sistemas foi feita através de comparação de formas de onda e espectros do sinal original e do sinal reconstruído, razão sinal-ruído segmentada e resultados de teste de escuta.

Os valores de opinião média obtidos do teste de escuta mostram que a qualidade do sinal reconstruído utilizando tanto o ATC-CEP quanto o ATC-ZN, a uma taxa de 16 kbit/s, é de boa para muito boa, estando situada entre aquelas produzidas pelos sistemas log-PCM de 5 e 6 bits por amostra.

Um outro ponto digno de nota é que embora o espectro-base do ATC-CEP aproxime bem melhor o sinal transformado (em especial no que tange à estrutura fina) do que o método de Zelinski e Noll, o desempenho do ATC-ZN se mostrou bastante próximo em termos de percepção auditiva daquele obtido pela técnica cepestral. Este bom desempenho permite que o ATC-ZN possa ser considerado um bom candidato para aplicações em que o sinal na entrada do codificador não seja unicamente voz. Com efeito, a forma como é calculado o espectro-base neste método não pressupõe um sinal de entrada específico, podendo, em princípio, ser usado tanto para voz como para sinais de dados, por exemplo. Já o método cepestral se utiliza da estimativa do período fundamental que é uma característica específica dos sinais de voz.

Finalmente, a tentativa de evitar a transmissão de informação paralela asso-

ciada às altas freqüências não foi bem sucedida. O desempenho do ATC-IAF foi nitidamente inferior aos dos demais codificadores investigados.

Agradecimentos

Os autores gostariam de agradecer a colaboração da EMBRATEL, através do seu Departamento de Desenvolvimento de Recursos Humanos, que permitiu a utilização do seu laboratório para preparação da fita contendo o material para o teste subjetivo.

Referências

[1] R. Zelinski e P. Noll, "Adaptive Transform Coding of Speech Signals", IE E Transactions on Acoustics. Speech and Signal Processing, vol. ASSP-25, n.° 4, Agosto 1977, pp. 299-309.

[2] R. W. Tansony e P. Kabal, "A Variable Rate Adaptive Transform Coder for Digital Storage of Audio Signals", IEEE International Conference on Communications, Philadelphia, Estados Unidos, Junho 1988, pp. 1.374-1.379.

[3] N. Ahmed, T. Natarajan e K. Rao, "Discrete Cosine Transform", IEEE Transactions on Computers, vol. C-23, n.° 1, Janeiro 1974, pp. 90-93.

[4] J. L. Flanagan et al, "Speech Coding", IEEE Transactions on Communications, vol. COM-27, n.° 4, Abril 1979, pp. 710-737.

[5] A. Alcaim, J. R. B. de Marca e C. F. B. Jaccoud, "Distribuição dos Recursos Binários em Codificadores de Voz no Domínio da Freqüência", Anais do 5.° Simpósio Brasileiro de Telecomunicações, Campinas, Setembro 1987, pp. 199-203.

[6] R. V. Cox e R. E. Crochiere, "Real-Time Simulation of Adaptive Transform Coding", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-29, n.° 2, Abril 1981, pp. 147-154.

[7] J. M.Tribolet e R. E. Crochiere,"A Vocoder—Driven Adaptation Strategy for Low—Bit Rate Adaptive Transform Coding of Speech", Proceedings of the International Conference on Digital Signal Processing, Florença, Itália, Setembro 1978, pp. 638-642.

[8] A. V. Oppenheim, "Speech Analysis-Synthesis System Based on Homomorphic Filtering", Journal of the Acoustics Society of America, vol. 45, Fevereiro 1969, pp. 459-462.

[9] J. R. B. de Marca, "Representação Digital de Sinais de Voz", Anais

do 5.º Simpósio Brasileiro de Telecomunicações, Campinas, Setembro 1987, pp. 5-19.

[10] C. F. B. Jaccoud e A. Alcaim, "Improvement of the Coding Structure in the Sub-Band Encoding of Speech Signals", Proceedings of the International Conference on Digital Signal Processing, Florença, Itália, Setembro 1987, pp. 315-318.

[11] H. S. Malvar e R. Duarte, "Codificação de Voz por Transformada Utilizando a LOT", Anais do 7.º Simpósio Brasileiro de Telecomunicações, Florianópolis, Setembro 1989, pp. 116-121.

ABRAHAM ALCAIM formou-se em Engenharia Elétrica pela PUC/Rio em 1975, obteve o título de Mestre em Ciências em Engenharia Elétrica pela mesma Universidade em 1977, e os títulos de D.I.C. e PhD em Engenharia Elétrica pelo Imperial College of Science and Tecnology, University of London, em 1981. Tem mais de 10 anos de experiência na área de codificação digital e transmissão de formas de onda e processamento digital de sinais de voz. Em 1984, desempenhou atividades na área de processamento de voz, como Pesquisador Visitante no Centre National d'Etudes de Télécommunications (CNET), em Lannion, França. Atualmente é Professor Associado do Centro de Estudos em Telecomunicações da PUC/Rio, onde atua em ensino e pesquisa no grupo de Sistemas de Telecomunicações, desenvolvendo trabalhos na área de codificação digital de voz a baixa taxa de bits. O Dr. Abraham Alcaim é o presidente da Comissão de Programa do SBT/IEEE International Telecommunications Symposium (ITS'90) a ser realizado no Rio de Janeiro em setembro de 1990.

JOSÉ ROBERTO BOISSON DE MARCA formou-se Engenheiro Eletricista (Telecomunicações) pela PUC/Rio em 1972, tendo posteriormente recebido os graus de M.Sc. (1975) e PhD (1977) em Engenharia Elétrica, ambos pela University of Southern California, Los Angeles. Foi engenheiro de telecomunicações da EMBRATEL, professor da UNICAMP e desde 1978 é professor do Centro de Estudos de Telecomunicações da PUC/Rio. Desempehou também as atividades de Consultor Científico do AT&T Bell Laboratories, Murray Hill (1986), Pesquisador Visitante da Universidade de Toronto (1981) e por duas vezes a função de Professor Visitante no Politécnico de Turim (1984 e 1989). O Prof. Boisson atuou ainda como presidente da Sociedade Brasileira de Telecomunicações (1984-1987), coordenador do Comitê Assessor em Engenharia Elétrica, Biomédica e Microeletrônica do CNPq (1986-1989),

Presidente da Comissão de Coordenadores de Comitês Assessores (CCCA) do CNPq (1987-1989) e Coordenador Central de Projetos Patrocinados da PUC/Rio (1984-1986). É membro senior do IEEE e presidente do Comitê para a América Latina da IEEE Communications Society. Atualmente exerce o cargo de diretor de Desenvolvimento Científico e Tecnológico do CNPq.

CARLOS FELIPE D. JACCOUD nasceu no Rio de Janeiro, em 9 de março de 1954. Formou-se em Engenharia Elétrica pelo Instituto Militar de Engenharia em 1977 e obteve o grau de Mestre em Ciências em Engenharia Elétrica na PUC/Rio, em 1986. Trabalhou na Cia. Telettra do Brasil entre 1977 e 1979 e a partir de novembro de 1979 ingressou no Departamento de Desenvolvimento de Recursos Humanos da EMBRATEL. Atualmente chefia a Seção de Desenvolvimento e Pesquisa deste departamento, onde vem desenvolvendo trabalhos de pesquisa e implementação de codificadores digitais de sinal de voz a 16 kbit/s, que constituem sua maior área de interesse.