

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

ESCOLA POLITÉCNICA

DEPARTAMENTO DE ELETRÔNICA E DE COMPUTAÇÃO

APERFEIÇOAMENTO DO CODIFICADOR DE VOZ CELP

Autor:

Thiago de Moura Prego

Orientador:

Prof. Sérgio Lima Netto, Ph.D.

Examinador:

Prof. Eduardo Antônio Barros da Silva, Ph.D.

Examinador:

Vagner Luís Latsch, M.Sc.

Rio de Janeiro, RJ – Brasil

Agosto de 2007

Agradecimentos

Meus sinceros agradecimentos:

- ao Professor Sérgio Lima Netto, pela orientação dada durante todo o período de desenvolvimento deste projeto e por ter me dado a oportunidade de fazer uma Iniciação Científica, o que certamente me incentivou a gostar mais da Engenharia Eletrônica;
- ao Professor Luiz Wagner Pereira Biscainho, por me acompanhar e incentivar desde o 4º período do curso de Engenharia Eletrônica e de Computação;
- a todas as pessoas que me ajudaram neste projeto por meio de dicas, orientação ou material de estudo;
- aos meus colegas de turma, pela amizade e ajuda durante esses anos da graduação.

Thiago de Moura Prego

Resumo

Este trabalho tem por objetivo otimizar o sistema de codificação de voz CELP existente no Departamento de Engenharia Eletrônica e Computação da Universidade Federal do Rio de Janeiro.

Para situar o leitor no panorama atual da codificação de voz, foi feita uma introdução à codificação de voz, mostrando diversas técnicas diferentes, focando principalmente no sistema CELP, alvo deste trabalho. O codificador inicial, codificador utilizado como base deste trabalho, também é descrito, assim como os bancos de voz utilizados.

Para a otimização do sistema CELP foi incorporado o bloco de detecção de silêncio, uma rotina para diminuir a taxa média de transmissão e a complexidade computacional do algoritmo. Além disso, foram refeitas as quantizações dos coeficientes do filtro de síntese e dos ganhos dos dicionários adaptativo e fixo; alterou-se o processo de interpolação dos coeficientes do filtro de síntese; por fim, analisamos o duplo ciclo da busca das melhores excitações do dicionário adaptativo e fixo.

Índice

CAPÍTULO 1	1
INTRODUÇÃO	1
1.1 Proposta do trabalho	2
1.2 Organização da dissertação	2
CAPÍTULO 2	4
CODIFICADORES DE VOZ	4
2.1 Introdução	4
2.2 Métodos de qualificação de codificadores de voz	4
2.2.1 MOS (Mean Opinion Score)	4
2.2.2 PESQ	5
2.3 Codificadores de voz	6
2.3.1 Codificadores por forma de onda	6
2.3.2 Codificadores paramétricos	6
2.3.3 Codificadores híbridos	9
2.3.4 Comparação entre os codificadores	9
2.4 Codificador CELP	10
2.4.1 Janelamento do sinal de voz	11
2.4.2 Modelo do trato vocal	11
2.4.3 Filtro perceptivo	14
2.4.4 Dicionários	15
2.4.5 Análise por síntese	16
2.4.6 Resumo do sistema CELP	17
2.5 Conclusão	20
CAPÍTULO 3	21
BANCO DE VOZ	21
3.1 Introdução	21
3.2 O que é um banco de voz	21
3.2.1 O recorte	22
3.3 Características do banco de voz	23
3.4 Bancos de treinamento e teste	24
3.5 Conclusão	25
CAPÍTULO 4	26
CODIFICADOR INICIAL	26
4.1 Introdução	26
4.2 Parâmetros do codificador inicial	27
4.2.1 Quantização	27
4.2.2 Duplo ciclo	30
4.2.3 Interpolação dos coeficientes LPC	32
4.2.4 Coeficiente de ponderação γ	33
4.2.5 Detecção de silêncio	33
4.2.6 Resumo	33
4.3 Resultados obtidos	34
4.4 Conclusão	34
CAPÍTULO 5	36
ALTERAÇÕES NO CODIFICADOR	36
5.1 Introdução	36
5.2 Pré-processamento e pós-processamento	37
5.3 Interpolação dos coeficientes LPC	38
5.4 Requantização dos coeficientes DLSF	40
5.4.1 O processo de requantização dos coeficientes DLSF	40
5.4.2 Testes	41
5.5 Reavaliação do coeficiente de ponderação γ	45
5.6 Requantização dos ganhos G_a e G_f	46
5.7 Reavaliação de K_a e K_f	49

5.8 Codificador final sem detecção de silêncio.....	54
5.9 Detecção de silêncio.....	55
5.10 Comparação entre os codificadores.....	59
5.11 Conclusão.....	60
CAPÍTULO 6	62
CONCLUSÃO	62
6.1 Contribuições do trabalho	62
6.2 Propostas para trabalhos futuros.....	63
REFERÊNCIAS BIBLIOGRÁFICAS.....	64
APÊNDICE A	65

Capítulo 1

Introdução

A cada dia que passa, as pessoas sentem mais necessidade de falar umas com as outras, independente da distância física. Um dos objetivos da área de Telecomunicações é suprir esta necessidade a partir do desenvolvimento de mecanismos e dispositivos para tal necessidade.

A evolução destes dispositivos é cada vez mais rápida, o que aumenta ainda mais o interesse das pessoas nesta área, o que estimula mais a evolução e assim por diante. Um dos grandes motivos desta rápida evolução é a revolução digital, em que sinais de diversas naturezas podem ser tratados como seqüências de bits, o que torna os computadores pessoais uma ferramenta efetiva de comunicação.

São feitas três etapas para representar um sinal de maneira digital: amostragem, quantização e codificação. A amostragem é o processo de transformar um sinal analógico, isto é, contínuo na amplitude e contínuo no tempo, em um sinal discreto no tempo e contínuo na amplitude. Para transformar este sinal amostrado em digital, é feita a quantização, processo que mapeia as infinitas possibilidades de amplitude de um sinal contínuo em um conjunto finito de valores pré-determinados. Quanto maior a quantidade de possibilidades nas quais as amplitudes serão mapeadas, melhor será a volta para o domínio contínuo, isto é, o sinal digital representará mais fielmente o sinal analógico e essa capacidade de representação é chamada de qualidade de codificação (ou quantização). Este número de amplitudes geralmente é do tipo 2^n , onde n é o número de bits.

A codificação é a forma pela qual os bits representarão o sinal digital. Para cada tipo de codificação existe certa quantidade de bits necessária para representar o sinal num determinado período de tempo, sendo esta quantidade chamada de taxa de codificação. Sendo assim, podemos comparar codificadores como mais ou menos eficientes, sendo que aquele que tiver a menor taxa de codificação para uma mesma qualidade de codificação será mais eficiente.

Entre os codificadores de sinais de voz que possuem baixa taxa de transmissão (codificação), os que mais se destacam atualmente são os codificadores baseados na técnica CELP (*Code Excited Linear Prediction*). Estes codificadores fazem uso de regressões lineares e dicionários de excitações (conceitos estes apresentados mais adiante) que serão utilizadas para a reprodução de voz. O fato de apresentarem um bom compromisso entre taxa de transmissão e qualidade de codificação, faz com que estes codificadores sejam amplamente utilizados na área de Telecomunicações, apesar da complexidade computacional que possuem.

1.1 Proposta do trabalho

Este projeto final tem por finalidade melhorar o sistema desenvolvido em [4] reavaliando processos existentes e da incluindo dois novos processos. Serão reavaliadas as quantizações dos coeficientes do filtro de síntese e dos ganhos dos dicionários adaptativo e fixo contendo as excitações a serem aplicadas ao filtro de síntese, assim como os pesos dados na interpolação dos coeficientes do filtro de síntese e do número de excitações armazenadas pelo processo de busca pelas melhores excitações de cada dicionário.

Foram adicionados os pré e pós-processamentos, com o objetivo de melhorar a qualidade de codificação, além do bloco de detecção de silêncio que visa diminuir a taxa de transmissão, com o mínimo possível de diminuição da qualidade de codificação.

1.2 Organização da dissertação

O capítulo 2 fornece uma idéia geral do campo da codificação de voz, com ênfase na técnica de codificação CELP, objeto de estudo deste trabalho. O funcionamento de um codificador baseado nesta técnica é descrito de maneira detalhada.

O capítulo 3 explica o que é um banco de voz e qual a sua importância para um sistema de codificação de voz. A maneira de obtenção dos dois bancos de voz utilizados neste trabalho (Banco de Treinamento e Banco de Teste) é descrita com detalhes, assim como a função de cada um desses bancos para o sistema CELP deste trabalho.

O capítulo 4 fornece uma descrição resumida do codificador inicial, codificador este que foi desenvolvido em [4] e que foi utilizado como base para este trabalho.

O capítulo 5 descreve detalhadamente as modificações e inclusões feitas por este trabalho no codificador inicial, com um resumo comparativo das duas versões do codificador final (implementadas neste trabalho) e o codificador inicial.

O capítulo 6 resume todo o projeto com comentários a respeito dos resultados obtidos e uma lista contendo propostas de novos trabalhos.

Capítulo 2

Codificadores de voz

2.1 Introdução

Este capítulo explica de forma sucinta os tipos de codificadores de voz, mostrando suas vantagens e desvantagens, dando maior ênfase no codificador CELP, pois é o codificador utilizado neste trabalho.

A seção 2.2 explica o método de qualificação de codificadores de voz utilizado neste trabalho: o MOS (*Mean Opinion Score*), que neste trabalho é estimado pelo PESQ (*Perceptual Evaluation of Speech Quality*).

A seção 2.3 descreve os diferentes tipos de codificadores de voz, sendo estes os codificadores paramétricos, por forma de onda e híbridos. Exemplos desses tipos serão mostrados, assim como as vantagens e desvantagens de cada um, além de uma comparação entre os seus desempenhos.

O codificador CELP é descrito com maior detalhamento na seção 2.4, incluindo um resumo de seu funcionamento.

2.2 Métodos de qualificação de codificadores de voz

Antes de falar sobre a maneira como codificar um sinal de voz, é importante informar como é feita a avaliação de um codificador de voz.

2.2.1 MOS (*Mean Opinion Score*)

O MOS é a recomendação P.800.1 [6] da ITU (*International Telecommunications Union*) que descreve métodos e procedimentos que permitem a avaliação subjetiva da qualidade de sinais de voz. São atribuídas notas de 1 a 5, segundo a tabela 2.1.

Tabela 2. 1 - Escala MOS.

MOS	Qualidade do sinal de voz
5	Excelente
4	Bom
3	Regular
2	Ruim
1	Pobre

O *MOS* é calculado de forma subjetiva reunindo-se um grupo de pessoas treinadas para fazer testes subjetivos de voz e são apresentados os sinais originais e codificados, sendo então as notas dadas de forma comparativa. Podemos perceber que este processo é bastante demorado e custoso, algo que fomentou a utilização de uma forma objetiva de avaliar as frases codificadas.

2.2.2 PESQ

A recomendação P.861.1 [7], também da ITU, foi utilizada para estimar a nota *MOS*. O cálculo da qualidade de voz é baseado em uma medida denominada *PESQ* (Perceptual Evaluation of Speech Quality), em que este apresenta precisão aceitável em seus resultados em relação à clareza da voz quando afetada pelos seguintes processos ou parâmetros:

- Codificadores de forma de onda;
- Codificadores paramétricos e híbridos a partir de 4kbps;
- Erros no canal de transmissão;
- Perdas de pacotes.

Para calcular o valor *PESQ*, compara-se o sinal de entrada com o sinal de saída do codificador, ambos em formato .wav, que são passados ao programa em sua chamada na linha de comando, junto com a taxa de amostragem dos sinais.

O resultado obtido pelo cálculo $PESQ$, sendo este um resultado objetivo, pode ser mapeado na avaliação subjetiva MOS através da seguinte equação [2]:

$$MOS = 0,999 + \frac{4}{1 + e^{(-1,4945 * PESQ + 4,6607)}}$$

2.3 Codificadores de voz

Os codificadores de voz podem ser classificados em três tipos: por forma de onda, paramétricos e híbridos. A grande diferença entre os tipos de codificadores consiste na informação que é enviada: o sinal de voz ou seus variantes no caso dos codificadores por forma de onda, parâmetros extraídos da manipulação do sinal de voz no caso dos paramétricos e uma mistura dos dois no caso dos híbridos.

2.3.1 Codificadores por forma de onda

Estes codificadores fazem uso das propriedades temporais e espectrais do sinal de voz a ser codificado. Seu objetivo é reconstruir o sinal original a partir destas propriedades, com baixo custo computacional, uma vez que não há qualquer tipo de análise do sinal.

A vantagem destes codificadores é a qualidade bastante elevada (próximo ao 4,5 na escala MOS). Porém, para tal, necessitam de uma grande taxa de bits, sendo esta a maior desvantagem.

Os mais conhecidos são os codificadores G.711 (PCM) e G.721 (ADPCM), geralmente utilizados com as taxas de 64 kbps e 32 kbps, respectivamente.

2.3.2 Codificadores paramétricos

Os codificadores paramétricos utilizam características da fonte geradora do sinal a ser codificado a partir da modelagem do trato vocal, fonte geradora do sinal de voz.



Figura 2. 1 - Processo de geração da voz humana no codificador LPC.

Possuem, num geral, taxas de bits baixas, em torno de 2 a 4 kbits/s. O preço para conseguir tais taxas é a baixa qualidade da voz restaurada (algo em torno de 2 na escala *MOS*), sendo esta a grande desvantagem deste grupo de codificadores.

Um exemplo de codificador paramétrico é o codificador LPC (*Linear Predictive Coding*), principal codificador paramétrico para voz, que leva em consideração vários princípios em relação ao trato vocal.

A voz é produzida pela interação dos seguintes órgãos: pulmões, traquéia, laringe, faringe (garganta), cavidade bucal e cavidade nasal. São elementos (articuladores) também responsáveis pela produção da fala as cordas vocais (na laringe), o palato mole (velum, que contém a úvula, vulga “campainha”), palato duro (vulgo “céu-da-boca”), a língua, os dentes, os lábios e a mandíbula. A voz pode ser vista, então, como a saída de um filtro digital que modela estes órgãos, onde a entrada é equivalente ao ar que sai dos pulmões e a vibração das cordas vocais. Este filtro é nomeado filtro de síntese. Podemos, então, classificar um som vocal em [1]:

- Sonoro: Basicamente um trem de impulsos (ou pulsos glotais);
- Surdo: Pode ser considerado ruído branco;
- Misto: Contém componentes sonoros e surdos;
- Silêncio: É, na verdade, a ausência de excitação;
- Plosivo: Silêncio por um instante de tempo, seguido de excitação sonora ou surda (fecha-se o trato vocal, aumentando a pressão do ar e soltando-o em seguida de uma só vez);
- Sussurro: Componentes essencialmente ruidosos até mesmo para os sons originalmente sonoros;
- Ejetivo: Apenas sai ar pela cavidade oral;
- Cliques e Implosivos: Nestes casos o ar é trazido para dentro do trato vocal.

Os dois tipos de sons mais importantes são os sonoros e surdos.

Os sons sonoros são aqueles em que o ar que sai dos pulmões vibra as cordas vocais. O trem de pulsos que representam este tipo de som é pseudo-periódico, em que o período relativo à frequência fundamental é chamado de *pitch*, sendo um importante parâmetro para estes codificadores. Formalmente, *pitch* é definido como frequência fundamental percebida. Em processamento de voz é comum ignorar o aspecto de perceptividade, determinando o *pitch* pela frequência aproximada do sinal. Como exemplos de sons sonoros, podemos citar os sons das vogais e dos encontros vocálicos.

Já os sons surdos são produzidos sem que haja a vibração das cordas vocais pelo ar que sai dos pulmões. Este tipo de som é representado por ruído branco, e como exemplos podemos citar os sons de fonemas consonantais como “ch” em chuchu, “s” em sorvete ou “f” em fósforo.

A voz é um sinal não estacionário e não periódico. Mas se levarmos em consideração pequenos segmentos de voz com duração entre 10 e 30 ms [1], podemos considerar a voz estacionária por partes. Sendo assim, para esses segmentos podemos modelar o processo de geração da voz humana como descrito na figura 2.1.

Este tipo de codificação faz uso da chamada análise LPC, em que, no modelo visto acima, seria um filtro *all-pole*, isto é, um filtro em que todos os zeros estão na origem. Esse filtro é determinado pela técnica de regressão linear e é aplicado a todos os segmentos de voz.

A vantagem deste codificador é o fato de apenas serem transmitidos os 10 coeficientes gerados pela análise LPC, um *flag* sinalizando se o som é surdo ou sonoro chamado de *flag U/UV* (caso o som seja sonoro, o valor do *pitch* também é transmitido) e o ganho do modelo, ao invés de todas as amostras quantizadas. Com isso, a quantidade de dados transmitidos é bem inferior à transmitida pelos codificadores por forma de onda.

Para exemplificar esta diferença na taxa de transmissão, se consideramos intervalos de voz de 20 ms a uma taxa de amostragem de 8 kHz, os codificadores LPC transmitem 13 valores, enquanto os codificadores PCM transmitem 160 valores. O esquema dos codificadores LPC é mostrado na figura 2.2. Ele tem como grande desvantagem o fato de classificar um bloco (ou sub-bloco) inteiro como sendo um som puramente sonoro ou puramente surdo.

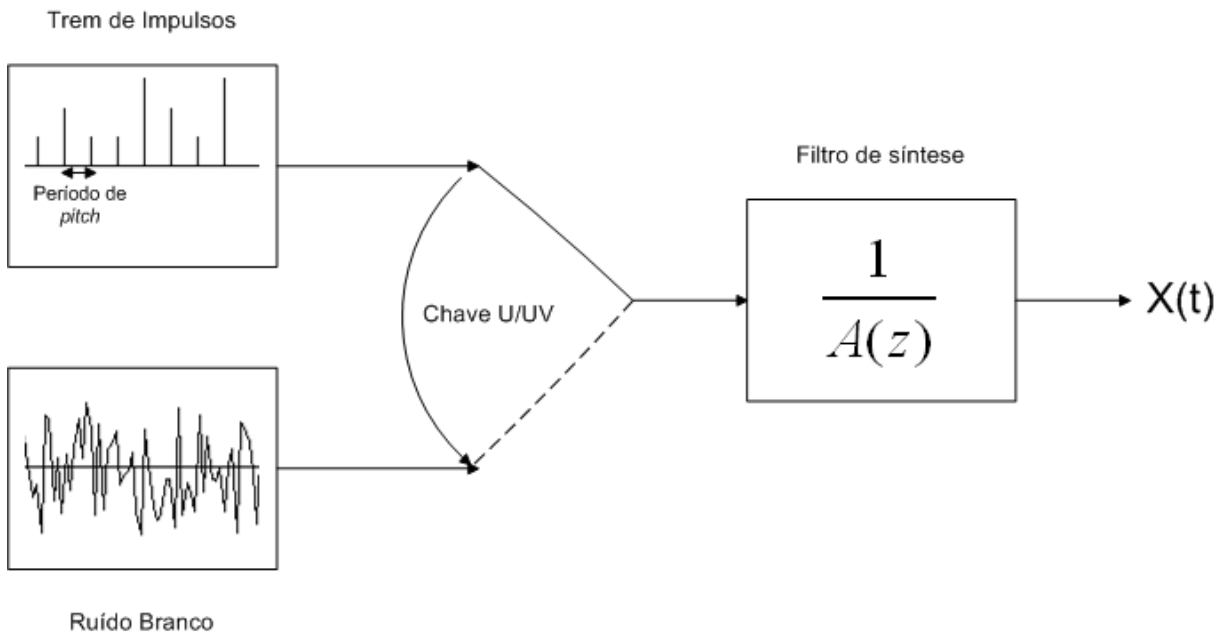


Figura 2. 2 - Esquema da codificação LPC.

2.3.3 Codificadores híbridos

Os codificadores híbridos extraem parâmetros dos sinais de voz assim como os codificadores paramétricos, e ao mesmo tempo utilizam características temporais e espectrais dos sinais como os codificadores de forma de onda. Desta forma, conseguem obter uma boa qualidade do sinal reconstituído ao final do sistema com taxas relativamente baixas, entre 2 e 16 kbps.

Uma forma de fazer isto é denominada técnica CELP, objeto de estudo deste trabalho. Um estudo mais detalhado desta técnica é realizado na seção 2.4.

2.3.4 Comparação entre os codificadores

A eficiência de um codificador de voz pode ser medida a partir da sua taxa de transmissão (taxa de bits) e pela qualidade da voz reconstruída.

A figura 2.3 faz a comparação entre os três tipos de codificadores de voz a partir destes dois parâmetros.

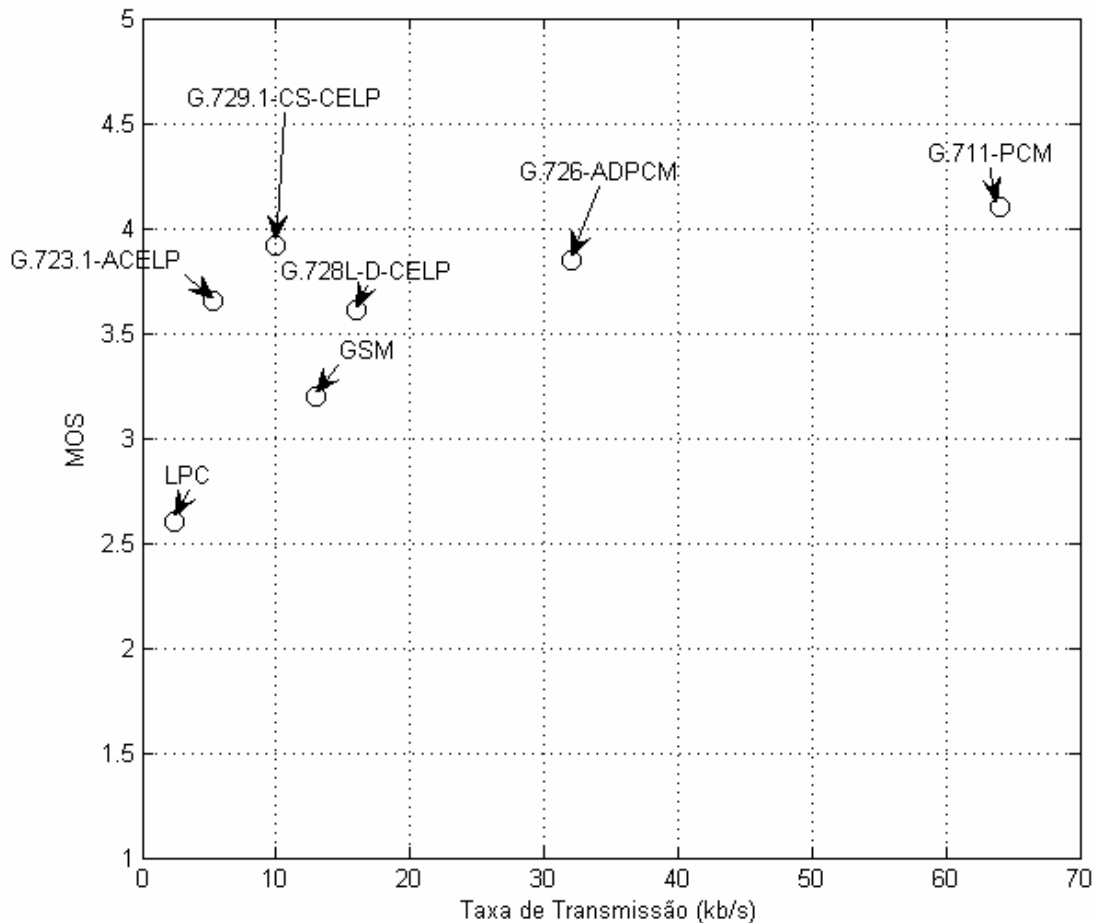


Figura 2.3 - Relação entre qualidade de codificação e taxa de transmissão de alguns codificadores.

De modo geral, por esta figura observamos que os codificadores híbridos (G.723.1-ACELP, GSM, G.729.1-CS-CELP e G.728L-D-CELP) representam um excelente compromisso entre taxa de transmissão e qualidade de codificação quando comparados com os codificadores por forma de onda (G.726-ADPCM e G.711-PCM) e o paramétrico (LPC).

2.4 Codificador CELP

O codificador CELP (*Code Excited Linear Prediction*) é um codificador de voz do tipo híbrido que utiliza o procedimento de Análise por Síntese para gerar o sinal codificado de voz. O codificador paramétrico LPC é utilizado como base para o codificador CELP, este fazendo uso de todas as suas características visando reduzir o número de parâmetros a serem transmitidos,

somando-se a exploração da questão da excitação utilizada como entrada do sistema, a grande causa da maior desvantagem do codificador LPC, que é a baixa qualidade da voz reconstruída.

Ele faz uso de um banco de dicionários para determinar as excitações, o que aumenta o número de excitações a serem utilizadas, melhorando, assim, a qualidade da voz reconstruída em relação ao codificador LPC.

2.4.1 Janelamento do sinal de voz

Como já mencionado anteriormente, a voz pode ser considerada estacionária em intervalos de 10 ms a 30 ms e essa característica é utilizada pelo codificador CELP. Para separar o sinal nestes intervalos é necessário fazer o janelamento do sinal de voz em blocos. Historicamente os codificadores CELP utilizam a janela Hamming, descrita por [3]:

$$w_h(n) = \begin{cases} \alpha + (1-\alpha) \cos\left(\frac{2\pi n}{M}\right), & |n| \leq \frac{M}{2} \\ 0, & |n| > \frac{M}{2} \end{cases},$$

onde $\alpha = 0,54$ [3] e M é a ordem do filtro.

Cada bloco de voz é ainda dividido em quatro sub-blocos de tamanhos idênticos, o que será explicado mais adiante.

2.4.2 Modelo do trato vocal

Como visto anteriormente na sessão 2.3.2, o trato vocal é modelado como um sistema linear (filtro digital) cuja entrada é uma excitação que modela o ar que sai dos pulmões e excita (no caso dos sons sonoros) ou não (no caso dos sons surdos) as cordas vocais. Este filtro digital $H(z)$ é chamado de Filtro de Síntese e é descrito pela função de transferência [2]:

$$H(z) = \frac{1}{A(z)},$$

$$\text{onde } A(z) = 1 - \sum_{j=1}^p a_j z^{-j},$$

e $\{a_1, a_2, a_3, \dots, a_N\}$ são os coeficientes de predição linear ou coeficientes LPC e são obtidos através da análise LPC. A constante p representa a ordem do modelo LPC e denota a precisão com a qual o filtro $H(z)$ irá modelar o trato vocal. No caso do codificador CELP, geralmente a ordem é $p = 10$, o que mantém uma boa relação entre qualidade e taxa de transmissão.

Existem algumas formas de se calcular os coeficientes LPC. A maneira utilizada neste trabalho é conhecida por método da auto-correlação [2], feito da seguinte maneira:

1. Calcula-se a matriz R_n de auto-correlação de cada sub-bloco do sinal de voz:

$$R_n(j) = \frac{1}{N} \sum_n^{N-1} x(n)x(n-j),$$

2. Resolve-se a equação matricial pelo método de Levinson-Durbin:

$$\begin{bmatrix} R_n(0) & R_n(1) & \dots & R_n(p-1) \\ R_n(1) & R_n(0) & \dots & R_n(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_n(p-1) & R_n(p-2) & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix},$$

onde $\{a_1, a_2, a_3, \dots, a_p\}$ são os coeficientes do filtro $H(z)$.

Todavia, os coeficientes LPC são muito sensíveis a erros de quantização, e por isso, ao invés deles, são utilizados coeficientes LSF (Line Spectral Frequency). Tais coeficientes são calculados da seguinte maneira:

1. Calculam-se os polinômios $P(z)$, simétrico, e $Q(z)$, anti-simétrico, a partir de $A(z)$:

$$\begin{aligned} P(z) &= A(z) + z^{-p-1} A(z^{-1}) \\ Q(z) &= A(z) - z^{-p-1} A(z^{-1}), \end{aligned}$$

onde p é o número de coeficientes preditores, isto é, a ordem do sistema de predição.

2. Definimos então o polinômio $P_1(z)$ como o polinômio $P(z)$ sem a raiz -1 , e o polinômio $Q_1(z)$ como o polinômio $Q(z)$ sem a raiz $+1$, ou seja:

$$P_1(z) = \frac{P(z)}{1+z^{-1}} \text{ e } Q_1(z) = \frac{Q(z)}{1-z^{-1}}, \text{ para } p \text{ par}$$

$$P_1(z) = P(z) \text{ e } Q_1(z) = \frac{Q(z)}{1-z^{-2}}, \text{ para } p \text{ ímpar}$$

3. Os polinômios $P_1(z)$ e $Q_1(z)$ são simétricos de ordem par, e como as suas raízes são pares de números complexos conjugados, apenas metade delas precisa ser determinada. Ou seja, para p pares, $p/2$ raízes de $P_1(z)$ e $p/2$ raízes de $Q_1(z)$, totalizando p raízes, podem representar os polinômios $P(z)$ e $Q(z)$, e conseqüentemente, o filtro de síntese $H(z)$. Como as p raízes estão sobre o círculo unitário, necessita-se apenas que se determine os seus argumentos complexos para representar $H(z)$. Estes valores são chamados de parâmetros LSF.

Sabe-se que o trato vocal se move continuamente, podendo assumir infinitas posições. Para suavizar esta movimentação, estimam-se os coeficientes do filtro de síntese para cada sub-bloco de voz, ao invés de fazê-lo para cada bloco. Isso é feito interpolando os coeficientes do bloco de voz atual com os coeficientes do bloco anterior, da seguinte maneira:

$$w_i^n = (1 - q_n)w_i^a + q_n w_i^c,$$

onde w_i^n são os coeficientes do n -ésimo sub-bloco, w_i^a são os coeficientes do sub-bloco anterior, w_i^c são os coeficientes do sub-bloco corrente e $q_n = \{0,25; 0,50; 0,75; 1,00\}$ [4].

Para diminuir ainda mais a taxa de transmissão, enviam-se as diferenças entre os coeficientes LSF consecutivos, ao invés dos próprios coeficientes LSF, pois estas possuem dispersão menor que os coeficientes originais. Estas diferenças são denominadas DLSF.

2.4.3 Filtro perceptivo

Sabe-se que, no domínio da frequência, quanto menor é a amplitude de uma componente do sinal de voz, mais facilmente o ouvido percebe erro/ruído neste componente. Então, o filtro perceptivo tem o intuito de minimizar o erro/ruído dando maior ênfase nas componentes de menor amplitude. O filtro perceptivo é denotado por $W(z)$ e possui a seguinte equação:

$$W(z) = \frac{A(z)}{A\left(\frac{z}{\gamma}\right)},$$

onde $\gamma \in (0,1)$ é o coeficiente de ponderação e é quem indica o grau de mudança no espectro. O valor de γ utilizado na maioria dos codificadores CELP está entre 0,8 e 0,9 [2].

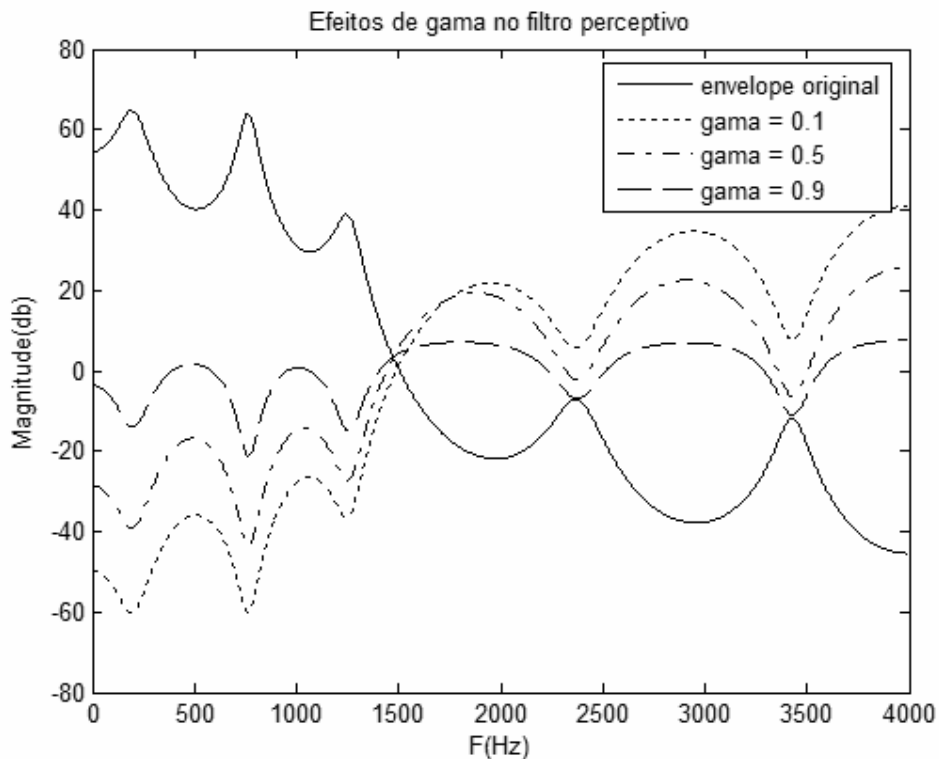


Figura 2. 4 - Efeitos de γ na resposta em frequência do filtro perceptivo.

A Figura 2.4 mostra os efeitos de alguns valores do coeficiente de ponderação γ sob o filtro perceptivo aplicado ao filtro de síntese para um trecho do sinal de voz da vogal “a”. Note que os

formantes do sinal, isto é, os picos encontram-se nas mesmas frequências tanto no envelope do sinal original, quanto nos envelopes modificados por γ .

2.4.4 Dicionários

Na técnica CELP, um dicionário é um conjunto de excitações para o filtro de síntese, da seguinte forma:

$$C = \{[x_0(n)], [x_1(n)], \dots, [x_{K-1}(n)]\}$$

Essa forma de representação indica que o dicionário armazena K seqüências x_n , onde n indica o índice da seqüência. Existem duas possibilidades de utilização de dicionários na análise por síntese: dicionário fixo e dicionário adaptativo.

Uma das vantagens do codificador CELP sobre o LPC é considerar um bloco (ou sub-bloco) como tendo uma parte sonora e outra surda. O dicionário adaptativo tem como objetivo estimar tanto a parte sonora quanto a surda com certa precisão, uma vez que se adapta às características do segmento a ser analisado. O dicionário fixo é responsável por estimar a parte surda que o adaptativo não conseguiu. Esta estimativa é feita a partir da busca da excitação armazenada no dicionário que geraria a resposta mais próxima do sinal a ser estimado.

No dicionário adaptativo as K seqüências armazenadas x_n são inicialmente zeradas e depois são atualizadas uma vez para cada sub-bloco com base na soma das melhores excitações dos dois dicionários para blocos anteriores. Já o dicionário fixo, como o nome já sugere, não altera o seu conteúdo.

No momento da busca, todas as excitações são filtradas pelo filtro de síntese, pois o que interessa para o codificador é a resposta gerada pela passagem de cada excitação pelo filtro, com o intuito de comparar cada uma destas respostas com o sinal original do sub-bloco em questão, como melhor explicado mais a diante. Para agilizar este processo pode-se filtrar todo o dicionário de uma só vez e depois realizar as buscas.

Para obter-se o sinal estimado que será utilizado para reconstruir a voz, calculam-se o ganhos a serem multiplicados em cada resposta obtida pela filtragem do dicionário para que as respostas possam ser comparadas com o sinal-alvo (o sinal que está sendo estimado). Este ganho é calculado da seguinte maneira:

$$G = \frac{R_{s,d}}{R_{d,d}},$$

onde $R_{s,d}$ é a correlação entre o sinal de voz a ser analisado, chamado também de sinal-alvo, e as respostas correspondentes a cada excitação contida no dicionário em questão. A $R_{d,d}$ é a autocorrelação entre as respostas correspondentes a cada excitação contida no dicionário em questão.

2.4.5 Análise por síntese

A análise por síntese é o processo de obter a excitação a ser utilizada para reconstruir a voz do locutor na saída do codificador a partir da comparação entre as respostas obtidas para cada excitação contida nos dicionários e o sinal original.

Como já mencionado anteriormente, cada bloco do sinal de voz, obtido através do janelamento, é subdividido em quatro sub-blocos de tamanhos idênticos como mostra a figura 2.5. É feita, então, a análise por síntese para cada um destes sub-blocos. Tal processo ocorre da seguinte maneira:

1. Para cada sub-bloco de voz, as excitações contidas no dicionário são submetidas ao filtro de síntese calculado pela análise LPC, gerando uma resposta para cada excitação;
2. Cada uma dessas respostas é subtraída do sinal contido no sub-bloco presente na entrada do sistema, gerando um sinal de erro;
3. Guarda-se esse sinal de erro;

4. A excitação que gerar o menor erro médio quadrático entre todas será utilizada para a reconstrução daquele sub-bloco do sinal de voz na saída do codificador.

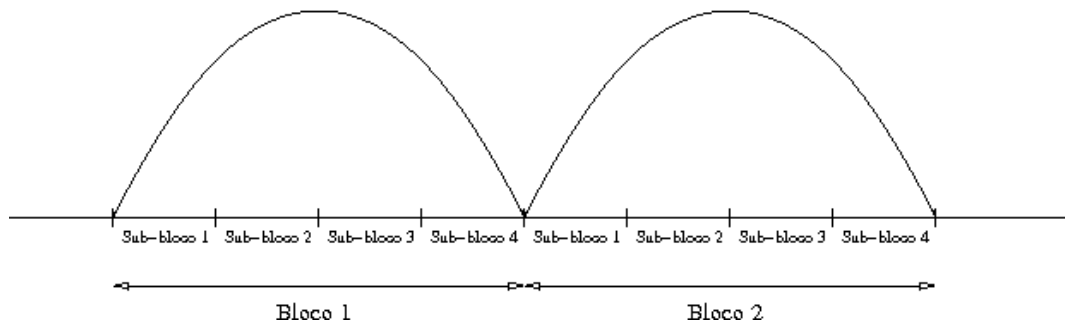


Figura 2. 5 - Divisão de um bloco em 4 sub-blocos.

2.4.6 Resumo do sistema CELP

O sistema CELP é ilustrado na figura 2.6, onde foram feitas marcações para auxiliar a identificação de cada etapa do sistema. Pressupõe-se que o sinal de voz já está amostrado e dividido em blocos (e sub-blocos). Na figura existe uma chave, que não existe no programa, com o intuito de facilitar o entendimento do sistema.

No codificador:

1. O sub-bloco do sinal de voz entra no sistema e calculam-se os coeficientes LPC do filtro de síntese;
2. Os coeficientes LPC são convertidos em coeficientes LSF;
3. Interpolam-se os coeficientes LSF utilizando os coeficientes do bloco atual e do bloco anterior, o que fornece uma melhor estimativa de como está o trato vocal para cada sub-bloco, ao invés de apenas para cada bloco.
4. Para cada sub-bloco, os coeficientes LSF interpolados (serão chamados apenas de LSF) são convertidos de volta para coeficientes LPC, para que o filtro de síntese possa ser gerado;
5. Os coeficientes LSF relativos a cada sub-bloco são transformados em coeficientes DLSF para que possam ser quantizados e transmitidos;

6. Inicia-se o processo de análise por síntese com a procura da melhor excitação do dicionário adaptativo. Este dicionário tem as suas excitações filtradas pelo filtro de síntese em cascata com o filtro perceptivo, gerando um dicionário de respostas.
7. Para cada seqüência do dicionário adaptativo, calcula-se o ganho correspondente e multiplica-se a resposta do filtro de síntese pelo ganho calculado gerando-se o sinal estimado;
8. Coloca-se a chave A/F na posição A;
9. Subtrai-se o sinal estimado do sinal-alvo, gerando-se o sinal de erro;
10. A seqüência que gerar o menor EMQ (erro médio quadrático) será a seqüência ótima, tendo seu índice e ganho relativo guardados;
11. Atualiza-se o sinal-alvo, subtraindo-se do mesmo a melhor seqüência do dicionário adaptativo multiplicada pelo devido ganho;
12. Já considerando o novo sinal-alvo, realiza-se a busca no dicionário fixo. Este dicionário tem as suas excitações filtradas pelo filtro de síntese em cascata com o filtro perceptivo, gerando um dicionário de respostas.
13. Para cada seqüência do dicionário fixo, calcula-se o ganho correspondente e multiplica-se a resposta do filtro de síntese pelo ganho calculado gerando-se o sinal estimado;
14. Coloca-se a chave A/F na posição F;
15. Subtrai-se o sinal estimado do sinal-alvo, gerando-se o sinal de erro;
16. A seqüência que gerar o menor EMQ (erro médio quadrático) será a seqüência ótima, tendo seu índice e ganho relativo guardados;
17. Somam-se as excitações de ambos os dicionários multiplicados pelos respectivos ganhos obtendo, assim, a excitação completa;

18. Atualiza-se o dicionário adaptativo colocando-se ao fim do mesmo a resposta ótima completa;
19. São transmitidos então os coeficientes DLSF, os índices das excitações de ambos os dicionários e os respectivos ganhos.

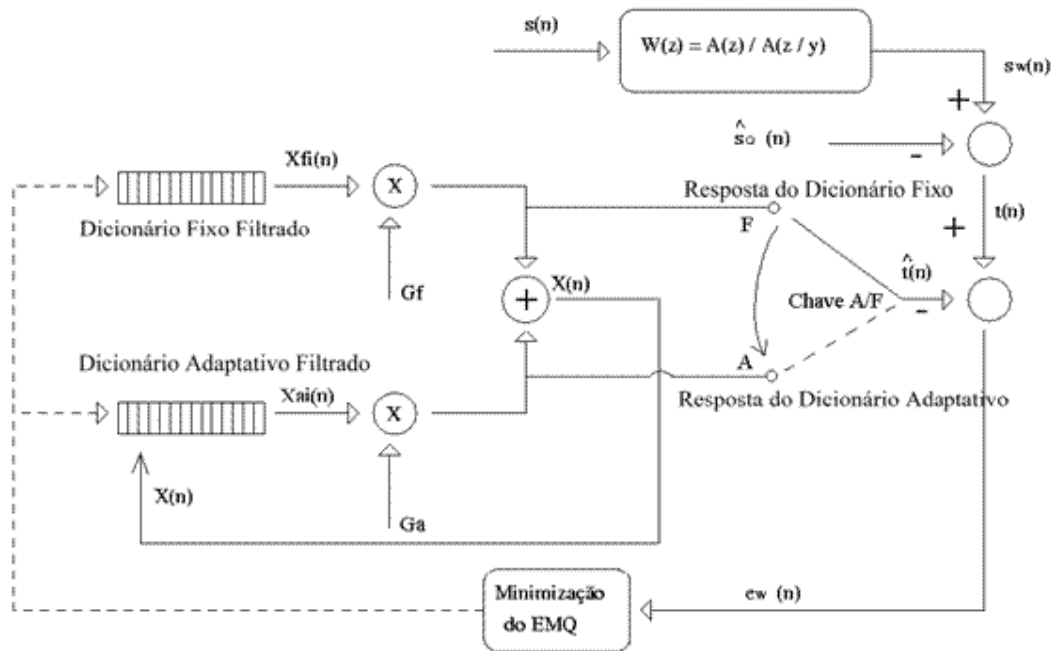


Figura 2. 6 - Esquemático do codificador de um sistema CELP [4].

No decodificador:

1. Os coeficientes DLSF recebidos são convertidos em LSF e depois LPC;
2. Com os coeficientes LPC formam-se o filtro de síntese e o filtro perceptivo;
3. Obtêm-se as respostas dos filtros fixo e adaptativo através da filtragem;
4. Aplicam-se os devidos ganhos e obtêm-se a estimativa do sinal original.

2.5. Conclusão

Neste capítulo, foi apresentada uma visão geral de codificação de voz, abordando os codificadores de forma de onda, os codificadores paramétricos, os codificadores híbridos e uma comparação entre eles.

O codificador híbrido CELP teve um maior destaque por ser o codificador utilizado neste trabalho. É necessário o seu entendimento para que as mudanças propostas no Capítulo 4 sejam compreendidas.

Capítulo 3

Banco de voz

3.1 Introdução

Este capítulo tem como objetivo explicar detalhadamente o banco de voz utilizado, desde a sua obtenção até a forma como foi organizado e algumas características de seus arquivos.

A seção 3.2 explica o que é o banco de voz e qual a sua importância para este trabalho, além de apresentar um roteiro da obtenção dos arquivos que o compõem, com detalhes da forma como foi feita esta obtenção e o motivo para tal.

Características do banco de voz tais como idioma, percentual de silêncio, duração, dentre outras são apresentadas na seção 3.3.

A divisão feita no banco em treinamento e teste será explicada na seção 3.4.

3.2 O que é um banco de voz

Define-se, neste trabalho, um banco de voz como sendo um conjunto de arquivos, cuja extensão é .wav, contendo arquivos codificados pelo método PCM (*Pulse-Code Modulation*). Originalmente eram utilizados 58 arquivos do OSR (*Open Speech Repository*) [13] com a distribuição por idioma segundo a Tabela 3.1.

Tabela 3.1 - Frases do OSR utilizadas.

Legenda	Idioma (locutor)	Frases
CH	Chinês (masculino)	4
FR	Francês (masculino)	6
IN	Indiano (masculino)	8
UK	Inglês Britânico (masculino)	15
US	Inglês EUA (masculino)	15

US	Inglês EUA (feminino)	10
----	-----------------------	----

Cada um destes arquivos contém mais de uma frase, algo que não é desejado. Sendo assim estes 58 arquivos foram recortados, então, em outros 596 arquivos com o objetivo de que cada novo arquivo possuisse uma única frase.

3.2.1 O recorte

Para recortar os arquivos do OSR, utilizou-se o programa audacity, disponibilizado gratuitamente no site www.audacity.com, de acordo com o seguinte roteiro:

1. Abre-se o arquivo contendo as frases a serem recortadas (será chamado de arquivo original daqui em diante), como mostra a figura Fig. 3.1;
2. Demarca-se a região que se deseja recortar, em que se definiu o limite de cada frase como sendo a metade silêncio entre as frases ou toda a região de silêncio no início ou final do arquivo, como mostram as figuras Fig. 3.2, Fig. 3.4 e Fig. 3.5;
3. Seleciona-se a opção Arquivo/Exportar Seleção como WAV, gerando-se o novo arquivo que contém a frase recortada (será chamado de arquivo recortado daqui em diante).



Figura 3. 1 - Arquivo original contendo 5 frases.



Figura 3. 2 - Região escolhida para recortar a primeira frase.

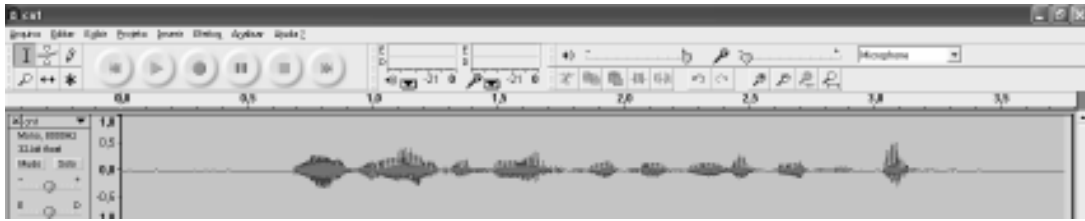


Figura 3. 3 - Arquivo gerado para a primeira frase recortada.



Figura 3. 4 - Região de silêncio no final do arquivo original.

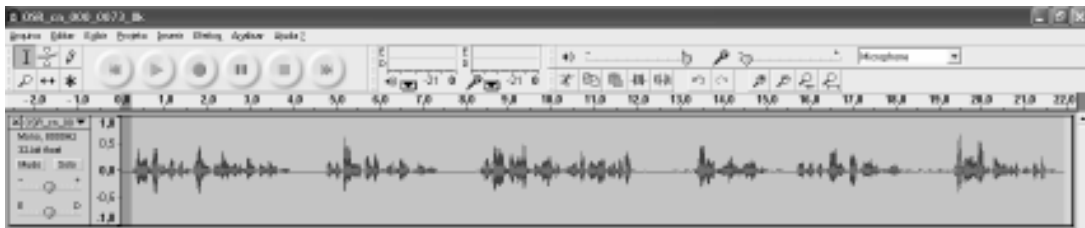


Figura 3. 5 - Região de silêncio no início do arquivo original.

3.3 Características do banco de voz

Algumas características das frases contidas no banco de voz foram levantadas e colocadas nas tabelas 3.2 e 3.3. Tais características são:

Nome: Nome do arquivo recortado.

Idioma: Idioma da frase contida no arquivo recortado.

Sexo: Sexo do locutor da frase.

Fs (Hz): Frequência de amostragem do sinal da frase. Todos os arquivos utilizados neste trabalho foram amostrados com 8 kHz.

Nbits: Número de bits por amostra do sinal que representa a frase. Todos os arquivos utilizados neste trabalho possuem 16 bits/amostra.

Energia: Somatório da Energia de cada amostra da frase.

% silêncio: Percentual do tempo de silêncio em relação ao tempo total da frase.

Duração: Tempo de duração, em segundos, da frase.

Máximo: Valor máximo de uma amostra do sinal da frase.

Mínimo: Valor mínimo de uma amostra do sinal da frase.

Te/Tr: Esta coluna indica se o arquivo faz parte do banco de teste (Te) ou do banco de treinamento (Tr).

3.4 Bancos de treinamento e teste

O banco de voz foi dividido em dois bancos: banco de treinamento e banco de teste. O banco de treinamento foi utilizado para treinar o sistema, isto é, para se obter o *codebook* de quantização para os coeficientes DLSF e de quantização dos ganhos dos dicionários adaptativo e fixo, além de validar todas as alterações propostas no capítulo 5. O banco de teste foi utilizado para se testar o sistema obtido através do banco de treinamento, tendo o seu MOS médio comparado com o MOS médio do banco de treinamento.

Uma boa escolha dos arquivos que farão parte do banco de treinamento é imprescindível para se ter um bom sistema de codificação, pois estes arquivos devem representar o melhor possível os arquivos que serão codificados na prática. Desta forma, queremos garantir que um bom resultado da codificação do banco de treinamento se reflita num bom resultado de modo geral na prática. Foram escolhidas 200 frases de forma aleatória, o que representa cerca de 30% das 596 frases totais, buscando que ambos os bancos tivessem características semelhantes, especialmente o % silêncio, pois este parâmetro interfere bastante na nota MOS estimada pelo PESQ. Todos os arquivos de ambos os bancos estão codificados em PCM com $N_{bits} = 16$ bits e $F_s = 8$ kHz.

Tabela 3. 2 - Distribuição do banco de voz por tipo, idioma e sexo.

Idioma	Sexo	Treinamento	Teste	Total
CH	M	9	11	20
FR	M	43	61	104
IN	M	36	44	80
UK	M	61	91	152
US	F	24	76	100
	M	27	113	140
Total	M/F	200	396	596

Tabela 3.3 - Médias dos Parâmetros por tipo.

Te/Tr	Parâmetro	Média
Teste	Mínimo	-0,273264172
	Máximo	0,390349788
	Duração	4,45718024
	% silêncio	54,91188266
	Energia	74,78215542
Treinamento	Mínimo	-0,273009625
	Máximo	0,378399775
	Duração	4,700506875
	% silêncio	56,19424131
	Energia	69,26830886
Total de Mínimo		-0,273178753
Total de Máximo		0,386339716
Total de Duração		4,538833473
Total de % silêncio		55,34220435
Total de Energia		72,93187134

A distribuição em idiomas e sexo está ilustrada na tabela 3.2, enquanto a tabela 3.3 mostra as médias dos parâmetros de cada banco, mostrando que realmente as características de ambos são similares, com uma diferença de 2% para o % silêncio, fato este que é satisfatório. A caracterização de cada argumento é colocado no Apêndice A deste trabalho, para uma referência mais completa.

3.5 Conclusão

Neste capítulo foi explicado o que é um banco de voz e foram descritos os dois bancos utilizados neste trabalho, o banco de treinamento e o banco de teste, assim como a finalidade de cada um e a importância de se ter um banco de voz de qualidade.

No capítulo 4 será descrito o codificador inicial, este que serviu de ponto de partida para o codificador CELP deste trabalho.

Capítulo 4

Codificador inicial

4.1 Introdução

O codificador utilizado como base para este trabalho foi desenvolvido por alunos do DEL/UFRJ – Departamento de Eletrônica e de Computação da Universidade Federal do Rio de Janeiro – em que a versão anterior a este trabalho foi implementada por Bruno Catarino Bispo [4].

Esta versão do codificador foi implementada na linguagem de programação C na plataforma UNIX e tem como programas um codificador e um decodificador. O codificador tem como entrada um arquivo de áudio do tipo .wav com codificação PCM, 16 bits por amostra e frequência de amostragem de 8 kHz. A saída é um arquivo do tipo binário com codificação CELP. Já o decodificador tem como entrada um arquivo do tipo binário gerado pelo codificador e a saída é um arquivo do tipo .wav com 16 bits por amostra e frequência de amostragem 8 kHz. Para facilitar a identificação dos arquivos mencionados no decorrer do trabalho o arquivo de entrada do codificador será chamado de frase original e o arquivo de saída do decodificador de frase codificada, apesar de ambos os arquivos serem do tipo .wav com a codificação PCM.

Este capítulo visa informar os parâmetros utilizados no codificador que serviu como base para este trabalho (será chamado de codificador inicial daqui em diante), tais como o número de bits utilizados para quantizar os coeficientes DLSF e para quantizar os ganhos dos dicionários adaptativo e fixo, valor utilizado de γ , interpolação dos coeficientes LSF, dentre outros, explicados com detalhe na seção 4.2. As modificações destes parâmetros propostas neste trabalho serão apresentadas nesta seção e explicadas no capítulo 5.

Na seção 4.3 encontram-se os resultados obtidos ao utilizar o codificador inicial para codificar e decodificar os arquivos do banco de treinamento e do banco de teste.

4.2 Parâmetros do codificador inicial

4.2.1 Quantização

Quantização é o processo de aproximar um conjunto de valores, geralmente infinitos valores contínuos ou muitos valores discretos, por um conjunto finito de valores discretos, de preferência pequenos para a aplicação. Como toda aproximação, a quantização insere um erro, uma distorção dos valores originais. Um bom quantizador é aquele que representa o sinal original com uma distorção mínima.

Existem dois tipos de quantização: quantização escalar e quantização vetorial. Na quantização escalar, cada dimensão do sinal é tratada de maneira independente, enquanto na quantização vetorial as dimensões são agrupadas em vetores. Podemos considerar que um sinal de voz é unidimensional (amplitude). Devido a isto e ao fato de a quantização vetorial ser mais complexa, as quantizações feitas no codificador inicial são todas escalares.

Uma outra classificação para quantizadores é quanto aos intervalos de quantização, sendo estes definidos como o intervalo em que os valores da entrada serão levados a um único valor de saída. Sendo este intervalo de quantização um valor fixo, chama-se o quantizador de Uniforme e de não-uniforme os que possuem tal intervalo com diferentes tamanhos. Um quantizador uniforme pode ser facilmente especificado pelo seu limite inferior e o intervalo de quantização. Já os não-uniformes necessitam de um dicionário contendo os limites de quantização, chamado neste trabalho de dicionário de quantização. Como a intenção é a de utilizar a menor quantidade de bits possível, uma quantização baseada em estatística faz-se necessária, no caso deste trabalho foi dada maior precisão para os valores mais significativos segundo critérios perceptivos, na medida do possível.

Os dez coeficientes DLSF e os ganhos dos dicionários adaptativo e fixo foram quantizados de maneira escalar e não-uniforme, sendo cada coeficiente e tipo de ganho quantizado a partir de um dicionário de quantização específico, isto é, o coeficiente DLSF(0) foi quantizado pelo dicionário_DLSF(0), o coeficiente DLSF(1) pelo dicionário_DLSF(1) etc. Os dez dicionários foram codificados com 4 bits cada, isto é, cada coeficiente DLSF contém 16 possíveis valores. Para não criar uma confusão entre dicionários de quantização e dicionários de respostas adaptativo e fixo, os ganhos dos dicionários de respostas adaptativo e fixo serão chamados de G_a e G_f , respectivamente.

Tabela 4.1 – Dicionários de quantização dos coeficientes DLSF.

DLSF(0)	Dicionário	{0.0409, 0.0512, 0.0659, 0.0810, 0.0971, 0.1121, 0.1268, 0.1408, 0.1533, 0.1662, 0.1801, 0.1951, 0.2131, 0.2352, 0.2671, 0.3189}
	Partição	{0.0447, 0.0585, 0.0733, 0.0892, 0.1045, 0.1195, 0.1340, 0.1470, 0.1597, 0.1731, 0.1874, 0.2034, 0.2236, 0.2494, 0.2884}
DLSF(1)	Dicionário	{0.0401, 0.0518, 0.0669, 0.0798, 0.0917, 0.1037, 0.1173, 0.1332, 0.1514, 0.1700, 0.1886, 0.2109, 0.2351, 0.2635, 0.3030, 0.3813}
	Partição	{0.0446, 0.0596, 0.0735, 0.0858, 0.0976, 0.1104, 0.1249, 0.1423, 0.1605, 0.1790, 0.1994, 0.2228, 0.2484, 0.2814, 0.3348}
DLSF(2)	Dicionário	{0.0595, 0.0882, 0.1108, 0.1308, 0.1502, 0.1704, 0.1911, 0.2136, 0.2383, 0.2656, 0.2947, 0.3268, 0.3523, 0.3898, 0.4481, 0.5504}
	Partição	{0.0754, 0.0997, 0.1209, 0.1404, 0.1600, 0.1807, 0.2019, 0.2257, 0.2515, 0.2803, 0.3104, 0.3412, 0.3686, 0.4156, 0.4890}
DLSF(3)	Dicionário	{0.0979, 0.1407, 0.1738, 0.2035, 0.2302, 0.2552, 0.2791, 0.3027, 0.3283, 0.3561, 0.3856, 0.4185, 0.4534, 0.4845, 0.5257, 0.6192}
	Partição	{0.1210, 0.1573, 0.1888, 0.2174, 0.2430, 0.2671, 0.2904, 0.3154, 0.3418, 0.3704, 0.4015, 0.4352, 0.4708, 0.5004, 0.5623}
DLSF(4)	Dicionário	{0.0714, 0.1146, 0.1542, 0.1894, 0.2202, 0.2471, 0.2712, 0.2917, 0.3117, 0.3399, 0.3741, 0.4147, 0.4648, 0.5230, 0.6023, 0.7403}
	Partição	{0.0945, 0.1347, 0.1724, 0.2053, 0.2340, 0.2593, 0.2827, 0.3003, 0.3250, 0.3561, 0.3932, 0.4391, 0.4922, 0.5593, 0.6583}
DLSF(5)	Dicionário	{0.0633, 0.0915, 0.1162, 0.1412, 0.1656, 0.1909, 0.2166, 0.2417, 0.2664, 0.2913, 0.3194, 0.3486, 0.3742, 0.4133, 0.4812, 0.6110}
	Partição	{0.0787, 0.1041, 0.1286, 0.1533, 0.1778, 0.2042, 0.2291, 0.2540, 0.2788, 0.3049, 0.3345, 0.3616, 0.3902, 0.4436, 0.5320}
DLSF(6)	Dicionário	{0.0927, 0.1369, 0.1688, 0.1987, 0.2249, 0.2475, 0.2693, 0.2880, 0.3087, 0.3337, 0.3629, 0.3961, 0.4365, 0.4867, 0.5514, 0.6647}
	Partição	{0.1177, 0.1529, 0.1845, 0.2120, 0.2365, 0.2591, 0.2790, 0.2978, 0.3210, 0.3480, 0.3791, 0.4157, 0.4608, 0.5173, 0.5959}
DLSF(7)	Dicionário	{0.0579, 0.0868, 0.1100, 0.1319, 0.1527, 0.1727, 0.1922, 0.2123, 0.2331, 0.2569, 0.2839, 0.3126, 0.3340, 0.3705, 0.4260, 0.5289}
	Partição	{0.0737, 0.0987, 0.1210, 0.1425, 0.1630, 0.1825, 0.2021, 0.2224, 0.2448, 0.2700, 0.2979, 0.3239, 0.3497, 0.3967, 0.4660}

DLSF(8)	Dicionário	{0.0935, 0.1299, 0.1570, 0.1806, 0.2028, 0.2245, 0.2446, 0.2646, 0.2841, 0.3042, 0.3204, 0.3382, 0.3640, 0.3987, 0.4502, 0.5473}
	Partição	{0.1138, 0.1440, 0.1688, 0.1919, 0.2138, 0.2342, 0.2549, 0.2744, 0.2938, 0.3131, 0.3283, 0.3502, 0.3796, 0.4219, 0.4885}
DLSF(9)	Dicionário	{0.0504, 0.0756, 0.0947, 0.1125, 0.1302, 0.1486, 0.1683, 0.1893, 0.2126, 0.2380, 0.2676, 0.3014, 0.3293, 0.3541, 0.3981, 0.4837}
	Partição	{0.0645, 0.0853, 0.1040, 0.1212, 0.1393, 0.1582, 0.1787, 0.2009, 0.2249, 0.2524, 0.2840, 0.3202, 0.3385, 0.3739, 0.4315}

Tabela 4. 2 - Ganhos dos dicionários adaptativo e fixo.

Ga	Dicionário	{0.0454, 0.1244, 0.1972, 0.2672, 0.3379, 0.4034, 0.4677, 0.5273, 0.5845, 0.6385, 0.6909, 0.7392, 0.7850, 0.8280, 0.8707, 0.9118, 0.9513, 0.9924, 1.0386, 1.0964, 1.1682, 1.2600, 1.3835, 1.5509, 1.7929, 2.1475, 2.6637, 3.4627, 4.7737, 7.0061, 11.5963, 25.9594}
	Partição	{0.0875, 0.1615, 0.2317, 0.3025, 0.3709, 0.4360, 0.4982, 0.5567, 0.6117, 0.6655, 0.7155, 0.7622, 0.8065, 0.8496, 0.8913, 0.9318, 0.9713, 1.0141, 1.0659, 1.1301, 1.2111, 1.3165, 1.4612, 1.6601, 1.9556, 2.3838, 3.0269, 4.0412, 5.7611, 8.7479, 16.0950}
Gf	Dicionário	{0.0, 5.4, 11.1, 17.2, 23.3, 29.4, 35.8, 42.7, 50.2, 58.4, 67.5, 77.7, 88.7, 100.8, 114.4, 128.8, 144.8, 163.1, 183.2, 205.3, 230.0, 258.8, 291.4, 329.8, 375.8, 434.9, 507.1, 604.3, 736.2, 934.8, 1263.9, 2129.5}
	Partição	{0.5, 8.3, 14.2, 20.3, 26.3, 32.5, 39.2, 46.4, 54.3, 62.9, 72.5, 83.1, 94.6, 107.4, 121.5, 136.7, 153.8, 172.9, 193.9, 217.2, 243.8, 274.4, 309.8, 352.2, 403.6, 468.6, 551.6, 665.6, 825.6, 1074.8, 1544.7}

Uma modificação proposta neste trabalho é a verificação da resolução de bits para estes parâmetros, a fim de escolher um melhor compromisso taxa de bits/qualidade de codificação.

4.2.2 Duplo ciclo

Chama-se Duplo Ciclo o processo de se procurar as K excitações que geram o menor EMQ (Erro Médio Quadrático), ao invés de apenas uma única excitação, como descrito na seção 2.4.4. Estas K excitações são então comparadas fazendo-se a busca tradicional, mais elaborada. Desta forma, reduz-se significativamente a complexidade geral na busca pela melhor excitação. Faz-se isto com o objetivo de tornar o sistema CELP mais fiel na sua codificação. Para entender o motivo para tal, deve-se entender como se dá o processo de busca dessas excitações.

Tanto o dicionário fixo quanto o adaptativo podem ser vistos como conjuntos de N excitações (vetores) com 40 amostras cada (5 ms de duração). Cada conjunto é representado por um único vetor de $N+39$ amostras e as excitações são vetores obtidos a partir deste vetor maior da seguinte maneira:

1. O dicionário é representado por um único vetor $D = [d_1, d_2, d_3, \dots, d_{N+41}]$;
2. A primeira excitação do dicionário é $v_1 = [d_1, d_2, d_3, \dots, d_{40}]$;
3. As demais excitações do dicionário são dadas por $v_n = [d_n, d_{n+1}, d_{n+2}, \dots, d_{n+39}]$;

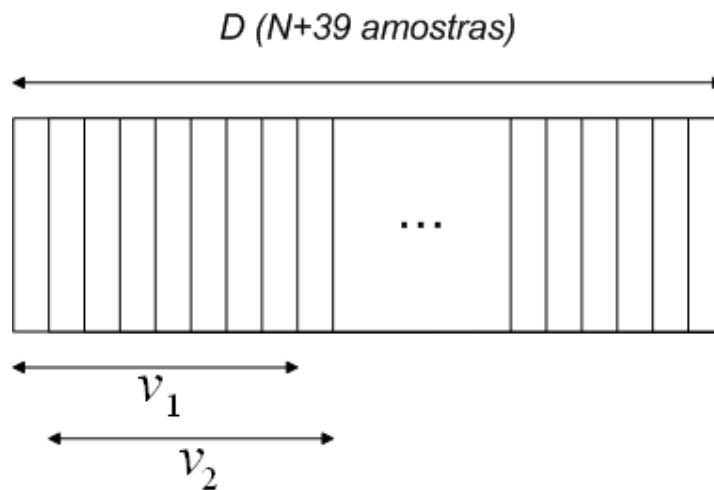


Figura 4. 1 - Dicionário com N excitações.

No caso do codificador inicial $N = 256$ para o dicionário fixo e $N = 1024$ para o dicionário adaptativo, como mostra a tabela. A filtragem de cada dicionário é feita aplicando-se separadamente cada vetor v_n ao filtro de síntese modificado pelo filtro perceptivo e armazenando sua resposta r_n ,

tendo um custo de 102.400 (256 excitações x 40 amostras x 10 coeficientes LPC) multiplicações para o dicionário fixo e de 409.600 multiplicações para o dicionário adaptativo.

Um modo de diminuir esta complexidade computacional é filtrar todas as excitações de uma só vez, isto é, aplicando-se o vetor D ao filtro de síntese modificado pelo filtro perceptivo e armazenando sua resposta R . A estimativa \hat{r}_n relativa a resposta r_n , é obtida da seguinte maneira:

1. A resposta R é representada por um único vetor $R = [m_1, m_2, m_3, \dots, m_{N+41}]$;
2. A estimativa relativa a resposta da primeira excitação do dicionário é $\hat{r}_1 = [m_1, m_2, m_3, \dots, m_{40}]$;
3. As demais estimativas são dadas por $\hat{r}_n = [m_n, m_{n+1}, m_{n+2}, \dots, m_{n+39}]$;

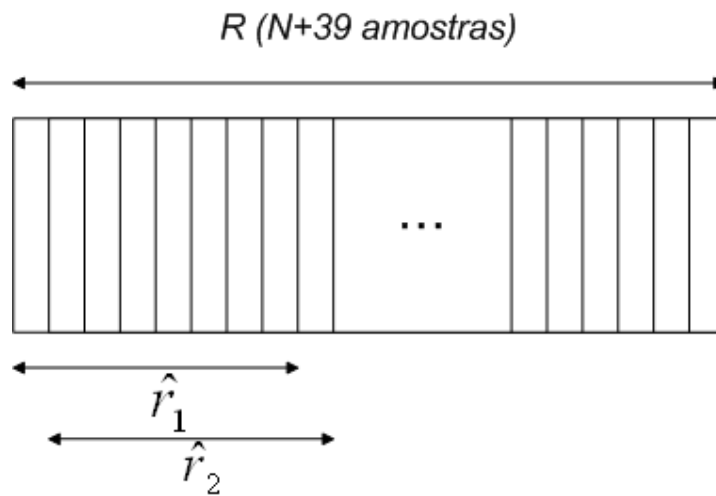


Figura 4. 2 - Resposta R , contendo as estimativas \hat{r}_n , relativa ao dicionário D .

Isto diminui a complexidade para $(N + 39) * 10$ multiplicações, aproximadamente 40 vezes menor que o método anterior, cuja complexidade é $N * 40 * 10$, porém se obtém as estimativas \hat{r}_n das respostas r_n , visto que no caso de menor complexidade computacional as condições iniciais do filtro não serão nulas, o que acontece no caso de maior complexidade computacional. Isto quer dizer que a excitação v_n que gera a resposta r_n com o menor EMQ não será, necessariamente, a mesma excitação que gera a estimativa \hat{r}_n com o menor EMQ, algo que não é desejado.

A fim de aumentar a probabilidade de a excitação que gera a resposta r_n com o menor EMQ ser encontrada, se armazena então, não só a excitação v_n que gerou a estimativa \hat{r}_n com o menor

EMQ, mas as K excitações que geraram as estimativas com os K menores erros. Estas K excitações são novamente filtradas, mas desta vez separadamente, obtendo-se as respostas r_n , e não mais as estimativas \hat{r}_n . Escolhe-se, entre estas K excitações, aquela que der o menor erro e a probabilidade de esta excitação ser realmente a que possui a resposta com o menor EMQ dentre todas do dicionário é tão maior quanto maior for o valor de K . Note que a complexidade do Duplo Ciclo será dada por:

$$Complexidade = 10 * (N + 39) + 400K$$

No caso do codificador inicial $K_a = K_f = 20$ excitações (em que K_a e K_f são os ganhos dos dicionários adaptativo e fixo). São propostos neste trabalho testes para diferentes valores de K para ambos os dicionários, a fim de obter um melhor compromisso complexidade/qualidade de codificação.

4.2.3 Interpolação dos coeficientes LSF

Conforme mencionado no Capítulo 3, os blocos de 20 ms são divididos em quatro sub-blocos de 5 ms cada, com o objetivo de acompanhar a velocidade do trato vocal através da busca da melhor excitação dos dicionários adaptativo e fixo. Para explorar melhor esta divisão, garantindo uma suavidade na transição entre os blocos consecutivos, faz-se uma interpolação entre os coeficientes LSF do bloco atual e do bloco anterior gerando coeficientes LSF interpolados, em que a interpolação é feita através de uma média ponderada entre os coeficientes dos blocos anterior e atual. Chamando os coeficientes LSF do bloco anterior de LSF_{an} e do atual de LSF_{at} e o coeficiente LSF interpolado de LSF_{in} , que é dado segundo a equação:

$$LSF_{in} = p_{an} * LSF_{an} + p_{at} * LSF_{at}$$

Tabela 4.3 - Distribuição dos pesos por sub-bloco.

Sub-bloco	p_{an}	p_{at}
1	0,75	0,25
2	0,50	0,50

Sub-bloco	p_{an}	p_{at}
3	0,25	0,75
4	0,00	1,00

Propôs-se então a verificação dos pesos para cada bloco, com a finalidade de melhorar a qualidade de codificação.

4.2.4 Coeficiente de ponderação γ

O valor utilizado para o coeficiente de ponderação foi de $\gamma = 0,8$, valor indicado por diversos trabalhos. Propôs-se, então, fazer a verificação do valor para este coeficiente que dê os melhores resultados de qualidade de codificação.

4.2.5 Detecção de silêncio

O codificador inicial não apresenta a detecção de silêncio, algo que leva a uma taxa de transmissão muitas vezes maior do que a necessária, uma vez que boa parte de uma conversa, por exemplo, pelo menos um dos falantes está em silêncio, além do fato de haver bastante silêncio nas próprias frases, como é mostrado no capítulo 3. Este fato motivou a inclusão de um bloco de detecção de silêncio no codificador deste trabalho.

4.2.6 Resumo

A Tabela 4.4 é um resumo das características do codificador inicial, com destaque para a taxa de transmissão de 7,6 kbps, que é cerca de oito vezes menor que os 64 kbps do codificador PCM G.711.

Tabela 4. 4 - Resumo das Características do Codificador Inicial.

Parâmetro	Característica
DLSF(0) a DLSF(9)	10 x 4 bits

Ga	5 bits
Gf	5 bits
Ka	20 excitações
Kf	20 excitações
γ	0,8
Dicionário Adaptativo	1024 excitações
Dicionário Fixo	256 excitações
Taxa de Transmissão	7,6 kbps

4.3 Resultados obtidos

Tabela 4.5 - MOS do Codificador Inicial por tipo.

Te/Tr	MOS
Tr	3,182
Te	3,133

A tabela 4.5 mostra os valores MOS obtidos pelo codificador inicial com as bases de treinamento e de teste. Comparando estes resultados com os da Figura 2.3, vemos que o codificador inicial compara-se em qualidade com o codificador GSM, com taxa de aproximadamente 13 kbps e MOS médio de 3,2, mas fica aquém dos codificadores G.729.1 e G.723.1, ambos do tipo CELP, com taxas de transmissão de cerca de 10 kbps e 6 kbps e MOS médios de 3,9 e 3,6, respectivamente. Tais resultados mostram que ainda há melhoras a serem feitas tanto na taxa de transmissão, quanto na qualidade de codificação, melhoras estas que foram mencionadas nas seções anteriores do Capítulo 4 e serão descritas no Capítulo 5.

4.4 Conclusão

O codificador inicial foi descrito neste capítulo, com o objetivo de mostrar em que situação se encontrava o codificador CELP antes de serem feitos os estudos propostos por este trabalho e que serão descritos com bastante detalhe no capítulo 5.

No próximo capítulo, serão detalhadas e explicadas as alterações no codificador inicial, a fim de aumentar a qualidade de codificação, ao mesmo tempo diminuindo a taxa de transmissão, com uma pequena preocupação na complexidade computacional.

Capítulo 5

Alterações no Codificador

5.1 Introdução

Este capítulo tem por objetivo enumerar as alterações que foram feitas no codificador inicial, detalhando cada alteração e mostrando os resultados obtidos. É importante ter em mente que os resultados obtidos utilizando o banco de treinamento são os utilizados como medidas para optar o que é melhor e o que é pior, mas a utilização do banco de teste é importante para fazer algumas confirmações, quando se julgar necessário e para avaliar o desempenho da versão final do codificador CELP.

Antes das modificações citadas no Capítulo 4, foram implementados um pré-processamento e um pós-processamento com o intuito de melhorar a qualidade de codificação, que será explicado na seção 5.2. Os pesos da interpolação dos coeficientes LSF foram reavaliados e o processo para tal é descrito na seção 5.3, enquanto a requantização dos coeficientes DLSF é descrita na seção 5.4.

A reavaliação do coeficiente de ponderação γ foi feita na seção 5.5. Já a requantização dos ganhos G_a e G_f dos dicionários Adaptativo e Fixo, respectivamente, é descrita na seção 5.6. A reavaliação das quantidades K_a e K_f das melhores excitações dos dicionários adaptativo e fixo, respectivamente, é descrita na seção 5.7. Na seção 5.8 faz-se uma comparação entre o codificador inicial e a versão do codificador CELP deste trabalho anterior à detecção de silêncio

O detalhamento do estudo sobre a inclusão detecção de silêncio é feito na seção 5.9, em que o objetivo principal da inclusão deste processo é o de diminuir a taxa de transmissão média, pois sua inserção no codificador CELP tende a diminuir a qualidade de codificação.

Na seção 5.10 encontra-se o resultado final do codificador CELP deste trabalho, assim como um resumo comparativo entre o codificador inicial e o codificador CELP deste trabalho.

5.2 Pré-processamento e pós-processamento

Foram implementados um pré- e um pós-processamentos, pois se percebeu que ocorria um *overflow* de memória ao serem efetuados cálculos durante o processo de codificação. O pré-processamento (feito no codificador), então, divide-se pela metade a amplitude do sinal antes de ser aplicado à etapa inicial de codificação e o pós-processamento (feito no decodificador) multiplica o sinal por dois, restaurando a divisão feita no pré-processamento. Estes pré- e pós-processamentos foram colaborações do doutorando Vagner Latsch, pelo projeto Maritaca [5] e fazem parte do padrão G.729 [8].

A figura 5.1 mostra o diagrama esquemático do funcionamento destes processamentos e a tabela 5.1 mostra o resultado obtido, destacando que o desempenho teve uma melhora de 0,068 MOS para o banco de treinamento. Esta versão do codificador será chamada de CELP Processado e o codificador inicial será chamado de CELP Inicial.

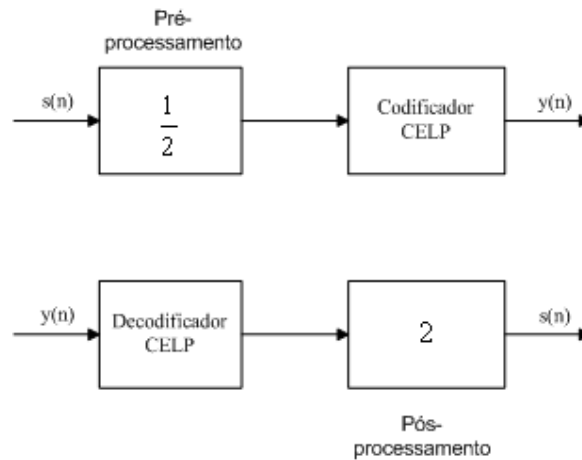


Figura 5. 1 - Pré- e Pós-processamentos.

Tabela 5. 1 - Comparação de qualidade de codificação entre o CELP Inicial e o CELP Processado.

Versão	Tr	Te
CELP Inicial	3,182	3,133
CELP Processado	3,250	3,207

5.3 Interpolação dos coeficientes LSF

Antes de começar os estudos propriamente ditos, foi feito um estudo dos códigos do codificador e do decodificador, com o intuito de entender o funcionamento destes e poder modificar os pontos de interesse. Durante este estudo preliminar, foi feita a requantização dos coeficientes DLSF com 6 bits cada (com um total de 60 bits), processo este descrito na seção 5.4.1. Com o CELP 60 (para reduzir o texto e facilitar a identificação, as versões do CELP desenvolvidas neste trabalho serão chamadas de CELP seguida do número de bits total dos coeficientes DLSF), foi então feito o estudo dos pesos da interpolação dos coeficientes LSF. Este estudo foi realizando fazendo os pesos variarem da seguinte maneira:

1. Os pesos de um mesmo sub-bloco são complementares, isto é $p_{an} = 1 - p_{at}$;
2. $\max p_{an} = \max p_{at} = 1,00$;
3. $\left\{ \begin{array}{l} p_{an}(n) = p_{an}(n-1) + 0,25 \\ p_{at}(n) = p_{at}(n-1) + 0,25 \end{array} \right\}$, onde $n = 1, 2, 3, 4$ é o sub-bloco;
4. Os testes foram feitos para $p_{at}(1) = 0,25; 0,50; 0,75$ e $1,00$. A cada modificação do valor de $p_{at}(1)$, modificaram-se também os valores dos outros pesos conforme o passo 3.

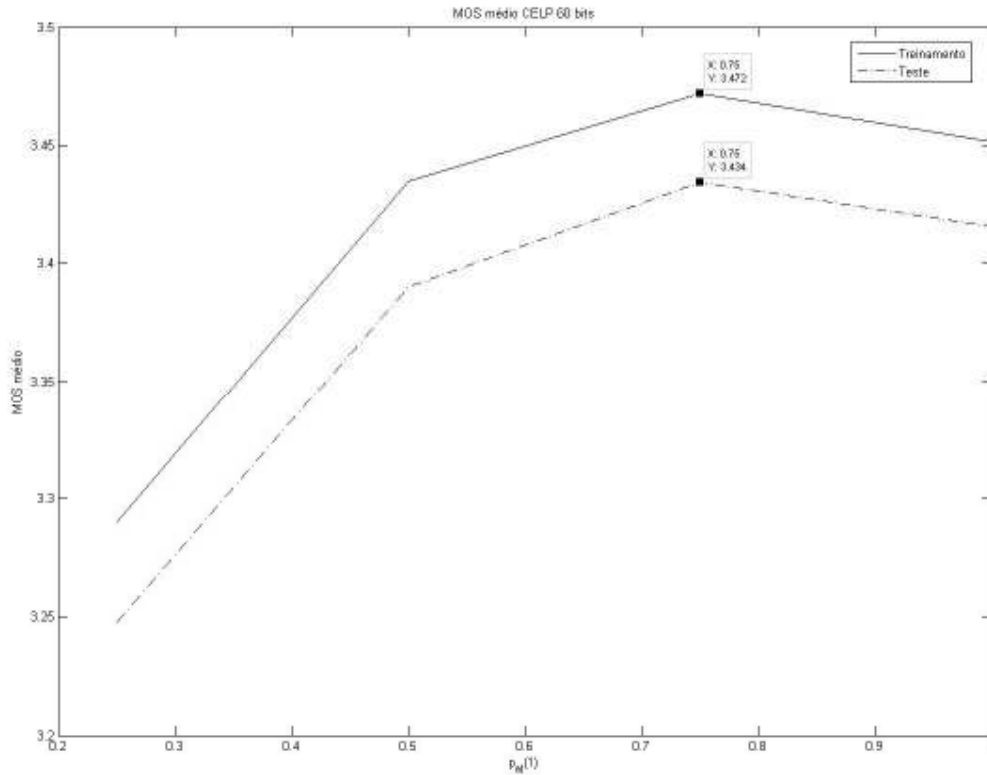


Figura 5. 2 - MOS médio para CELP 60 em função de $p_{at}(1)$.

A figura 5.2 mostra o resultado obtido com o estudo da interpolação dos coeficientes LPC e nota-se que o melhor valor para $p_{at}(1)$ é 0,75. A nova distribuição dos pesos ficou então da maneira mostrada na tabela 5.2. Todos os testes das demais seções foram feitos com os novos valores para os pesos da interpolação dos coeficientes LSF.

Tabela 5. 2 - Nova distribuição dos pesos por sub- bloco.

Sub-bloco	p_{at}	p_{an}
1	0,75	0,25
2	1,00	0,00
3	1,00	0,00
4	1,00	0,00

A tabela 5.3 compara as versões do CELP 60 ($p_{at}(1)=0,25$) e o CELP 60 0,75 ($p_{at}(1)=0,75$). O resultado mostra que houve uma melhoria significativa na qualidade de codificação sem qualquer aumento de complexidade computacional ou da taxa de transmissão.

Tabela 5.3 - Comparação de qualidade entre o CELP 60 e o CELP 60 0,75.

Versão	Tr	Te
CELP 60	3,290	3,248
CELP 60 0,75	3,472	3,434

5.4 Requantização dos coeficientes DLSF

Como mencionado na seção 4.2.1, este trabalho propõe o estudo da resolução de bits para os dez coeficientes DLSF, pois foi utilizado um total de 40 bits para estes coeficientes (com 4 bits por coeficiente).

Com o objetivo de diminuir a taxa de transmissão e ao mesmo tempo melhorar a qualidade de codificação, foi feita uma ampla quantidade de testes para valores decrescentes de resolução de bits para os coeficientes DLSF.

5.4.1 O processo de requantização dos coeficientes DLSF

A quantização destes coeficientes é realizada da seguinte maneira:

1. Executa-se o programa de codificação para o banco de treinamento, salvando em um arquivo do tipo texto, todos os coeficientes LSF gerados pela função `lpc2lsf` contida no arquivo `lpc.c`, sendo estes coeficientes do tipo `double`, considerados grosseiramente não quantizados;
2. Utilizando-se um script no programa MATLAB, calcula-se o dicionário ótimo para uma dada resolução de bits. Este script utiliza a função `lloyds` do MATLAB;
3. Carrega-se este novo dicionário no arquivo `quant_lsf.h` através do programa `carregar_cb_part` em C++ e recompila-se o codificador e decodificador;
4. Executa-se o programa de codificação e o de decodificação para os bancos de treinamento e de teste e comparando-se a saída do decodificador com os arquivos originais utilizando uma implementação do *PESQ*, guardando a nota MOS relativa a cada frase em um arquivo texto e depois se calculando as médias para cada banco.

5.4.2 Testes

Para saber qual versão do CELP seria utilizada para iniciar a otimização das resoluções de bits para os coeficientes DLSF, foi feito um teste entre as versões CELP 60, CELP 50 e CELP 40, com 6, 5 e 4 bits por coeficientes, respectivamente, como mostra a tabela 5.4. Apesar de um dos objetivos ser a diminuição da taxa de transmissão, aumentou-se a resolução de bits para um total de 60 bits e 50 bits, pois não se sabia a priori, se a distribuição com 4 bits por coeficiente DLSF era a melhor para um total de 40 bits, que não se mostrou a melhor com o decorrer dos testes.

Tabela 5. 4 - Escolha da versão inicial para a otimização dos coeficientes DLSF.

Versão	MOS médio Tr	MOS médio Te
CELP 60	3,472	3,434
CELP 50	3,467	3,421
CELP 40	3,431	3,396

Por apresentar um melhor compromisso o CELP 50 (5 bits para cada coeficiente DLSF) foi utilizado como ponto de partida dos testes e foi-se diminuindo a resolução de bits até que o MOS tivesse uma queda brusca, sendo então utilizado como nova resolução aquela anterior a que originou esta queda brusca. É necessário lembrar que o objetivo deste trabalho é ter uma qualidade melhor em pelo menos 0,2 MOS com uma taxa menor que o Codificador Inicial, algo que será vital na consideração do que é ou não uma queda brusca.

A tabela 5.5 ilustra todo o processo de otimização dos coeficientes DLSF, em que a coluna versão indica qual a quantidade total de bits para os 10 coeficientes DLSF. A coluna distribuição de bits mostra como se encontrava a distribuição de bits para os 10 coeficientes no momento do estudo em questão e as colunas treinamento e teste indicam o MOS médio para cada banco de voz. A versão escolhida foi a última da tabela 5.5, tendo a distribuição [4 3 4 4 4 3 3 3 3 1], com um total de 32 bits e um MOS de 3,379 para o Treinamento, aproximadamente 0,2 maior que o MOS de 3,181 para o codificador inicial. Estão destacadas, também, as linhas referentes ao CELP 40 com duas diferentes distribuições, para efeito de comparação com o codificador inicial. O método de escolha da melhor distribuição de bits está ilustrado na figura 5.3 e foi feito da seguinte maneira:

1. Traçou-se uma reta passando pelos pontos referentes às linhas 2 (CELP 49) e 4 (CELP 48) da tabela 5.5;
2. Os pontos abaixo da reta foram considerados inaceitáveis e o ponto mais acima da reta para uma mesma taxa de transmissão foi escolhido como base para uma taxa de transmissão menor.

Tabela 5. 5 - Otimização dos Coeficientes DLSF.

Versão	Distribuição de Bits	Treinamento	Teste
CELP 49	[5 5 5 5 5 5 5 5 4]	3,471	3,426
CELP 48	[5 5 5 5 5 5 5 5 3]	3,457	3,421
CELP 47	[5 5 5 5 5 5 5 5 2]	3,457	3,424
CELP 46	[5 5 5 5 5 5 5 5 1]	3,444	3,404
CELP 46	[5 5 5 5 5 5 5 4 2]	3,455	3,424
CELP 45	[5 5 5 5 5 5 5 3 2]	3,453	3,422
CELP 44	[5 5 5 5 5 5 5 2 2]	3,437	3,411
CELP 44	[5 5 5 5 5 5 4 3 2]	3,454	3,416
CELP 43	[5 5 5 5 5 5 3 3 2]	3,449	3,410
CELP 42	[5 5 5 5 5 5 2 3 2]	3,444	3,407
CELP 42	[5 5 5 5 5 4 3 3 2]	3,443	3,411
CELP 42	[5 5 5 5 4 5 3 3 2]	3,451	3,410
CELP 41	[5 5 5 5 3 5 3 3 2]	3,435	3,401
CELP 41	[5 5 5 4 4 5 3 3 2]	3,451	3,405
CELP 40	[5 5 5 3 4 5 3 3 2]	3,433	3,392
CELP 40	[5 5 5 4 4 5 3 3 2]	3,440	3,404
CELP 39	[5 5 5 3 4 4 5 3 3 2]	3,256	3,237
CELP 39	[5 5 4 4 4 4 5 3 3 2]	3,434	3,394
CELP 38	[5 5 3 4 4 4 5 3 3 2]	3,420	3,381
CELP 38	[5 4 4 4 4 4 5 3 3 2]	3,429	3,393
CELP 37	[4 4 4 4 4 4 5 3 3 2]	3,417	3,377
CELP 36	[4 4 4 4 4 4 3 3 2]	3,417	3,379
CELP 35	[4 4 4 4 4 3 3 3 2]	3,414	3,372
CELP 34	[4 4 4 4 4 3 3 3 1]	3,403	3,357
CELP 33	[4 4 4 4 4 3 3 2 1]	3,387	3,346

CELP 33	[4 4 4 4 4 3 2 3 1]	3,193	3,164
CELP 33	[4 4 4 4 4 2 3 3 1]	3,381	3,345
CELP 33	[4 4 4 4 3 3 3 3 1]	3,393	3,352
CELP 33	[4 4 4 3 4 3 3 3 1]	3,388	3,350
CELP 33	[4 4 3 4 4 3 3 3 1]	3,370	3,339
CELP 33	[4 4 3 4 4 3 3 3 1]	3,380	3,347
CELP 33	[4 3 4 4 4 4 3 3 3 1]	3,389	3,351
CELP 33	[3 4 4 4 4 4 3 3 3 1]	3,382	3,345
CELP 32	[4 3 4 4 4 3 3 3 3 1]	3,379	3,345

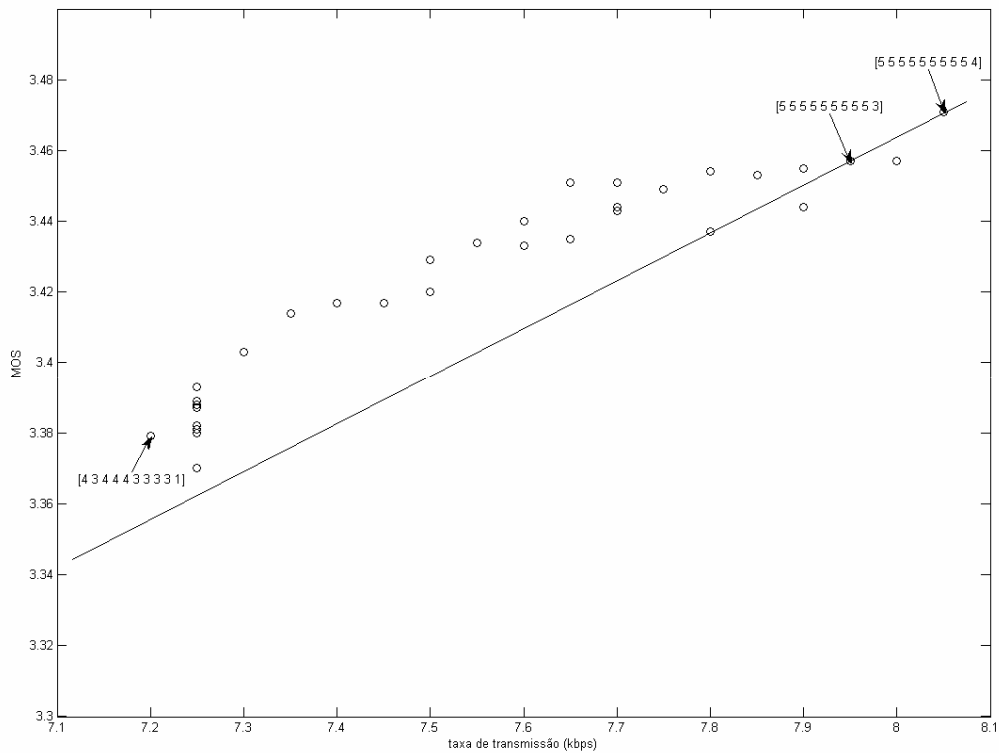


Figura 5.3 - Método de decisão da melhor distribuição de bits.

A tabela 5.6 mostra como ficou o dicionário de quantização do CELP 32 que será utilizado nos testes das próximas seções.

Tabela 5. 6 - Dicionário de quantização dos coeficientes DLSF do CELP 32.

DLSF(0)	Dicionário	{6.1247000e-002, 7.7576987e-002, 9.3576000e-002, 1.0195233e-001, 1.1182552e-001, 1.3348394e-001, 1.3998146e-001, 1.4590300e-001, 1.5426300e-001, 1.5956200e-001, 1.6569026e-001, 1.7700800e-001, 1.8070709e-001, 1.8397100e-001, 2.0141724e-001, 2.4864175e-001};
	Partição	{6.1247000e-002, 8.8125754e-002, 9.7764164e-002, 1.0611600e-001, 1.2338993e-001, 1.3703800e-001, 1.4294223e-001, 1.4992777e-001, 1.5691250e-001, 1.6219600e-001, 1.7321060e-001, 1.7885755e-001, 1.8233905e-001, 1.8994742e-001, 2.1934965e-001};
DLSF(1)	Dicionário	{3.8682651e-002, 5.9049205e-002, 7.7767889e-002, 9.9221859e-002, 1.2643719e-001, 1.6065919e-001, 2.0675693e-001, 3.0289328e-001};
	Partição	{4.9029714e-002, 6.8109317e-002, 8.7578455e-002, 1.1175083e-001, 1.4287143e-001, 1.8157201e-001, 2.4205150e-001};
DLSF(2)	Dicionário	{7.0987699e-002, 1.0836179e-001, 1.4195338e-001, 1.7589146e-001, 2.1087368e-001, 2.4507512e-001, 2.7858848e-001, 3.0916767e-001, 3.3684762e-001, 3.6273585e-001, 3.8803136e-001, 4.1391454e-001, 4.4169477e-001, 4.7441906e-001, 5.1757902e-001, 6.2445798e-001};
	Partição	{9.0921368e-002, 1.2534437e-001, 1.5891746e-001, 1.9277307e-001, 2.2790196e-001, 2.6230790e-001, 2.9388510e-001, 3.2362049e-001, 3.4987872e-001, 3.7549196e-001, 4.0068484e-001, 4.2761604e-001, 4.5728548e-001, 4.9344103e-001, 5.5257734e-001}
DLSF(3)	Dicionário	{9.7213994e-002, 1.5656777e-001, 1.9701364e-001, 2.2679143e-001, 2.5187227e-001, 2.7480697e-001, 2.9694469e-001, 3.1963783e-001, 3.4455816e-001, 3.7296450e-001, 4.0663973e-001, 4.4945384e-001, 5.0430599e-001, 5.7254587e-001, 6.7743550e-001, 8.6497810e-001}
	Partição	{1.3032903e-001, 1.7882770e-001, 2.1281091e-001, 2.3979936e-001, 2.6348038e-001, 2.8590718e-001, 3.0820661e-001, 3.3161535e-001, 3.5813308e-001, 3.8906122e-001, 4.2673665e-001, 4.7542854e-001, 5.3600849e-001, 6.1881063e-001, 7.5506466e-001}
DLSF(4)	Dicionário	{6.3290627e-002, 1.0850374e-001, 1.4554552e-001, 1.8149783e-001, 2.1477397e-001, 2.4435814e-001, 2.7098761e-001, 2.9504532e-001, 3.1805165e-001, 3.4125617e-001, 3.6606120e-001, 3.9367872e-001, 4.2637976e-001, 4.7241157e-001, 5.4443733e-001, 6.9816511e-001}

	Partição	{8.8279921e-002, 1.2748365e-001, 1.6381912e-001, 1.9880144e-001, 2.2989863e-001, 2.5779691e-001, 2.8323657e-001, 3.0657252e-001, 3.2961812e-001, 3.5343735e-001, 3.7954793e-001, 4.0914524e-001, 4.4679878e-001, 5.0375477e-001, 6.0410909e-001}
DLSF(5)	Dicionário	{1.0628123e-001, 1.8836559e-001, 2.4230497e-001, 2.8538239e-001, 3.2899176e-001, 3.8221231e-001, 4.6850286e-001, 6.6368576e-001}
	Partição	{1.5207273e-001, 2.1734961e-001, 2.6424707e-001, 3.0642294e-001, 3.5367152e-001, 4.1862507e-001, 5.4296499e-001}
DLSF(6)	Dicionário	{1.2315748e-001, 2.1156299e-001, 2.6439728e-001, 3.0584653e-001, 3.4670857e-001, 3.9574152e-001, 4.6819360e-001, 6.2210679e-001}
	Partição	{1.7509915e-001, 2.4027236e-001, 2.8569343e-001, 3.2596323e-001, 3.6982735e-001, 4.2707464e-001, 5.2842518e-001}
DLSF(7)	Dicionário	{9.2036160e-002, 1.7093928e-001, 2.2291712e-001, 2.6334335e-001, 3.0082440e-001, 3.4206156e-001, 3.9642711e-001, 5.2407808e-001}
	Partição	{1.3655640e-001, 1.9895252e-001, 2.4425045e-001, 2.8183379e-001, 3.2048773e-001, 3.6625915e-001, 4.4005276e-001}
DLSF(8)	Dicionário	{1.2938033e-001, 2.0874312e-001, 2.5920534e-001, 3.0032312e-001, 3.3991393e-001, 3.8693917e-001, 4.5683670e-001, 6.0581601e-001}
	Partição	{1.7471569e-001, 2.3578360e-001, 2.8030921e-001, 3.1980747e-001, 3.6189514e-001, 4.1724792e-001, 5.1332767e-001}
DLSF(9)	Dicionário	{1.9451095e-001, 3.6456870e-001}
	Partição	{2.8004452e-001}

5.5 Reavaliação do coeficiente de ponderação γ

Após a requantização dos coeficientes DLSF, foi feita a reavaliação do valor do coeficiente de ponderação γ . A fim de testar se o valor ótimo para γ é realmente $\gamma = 0,8$ como indicado na literatura [4], foram feitos testes com valores entre 0,05 e 1,00 com incrementos de 0,05. As versões utilizadas foram a CELP 32 e a CELP 40, destacadas na tabela 5.5.

O resultado destes testes encontra-se na Figura 5. 4, que mostra que a região com os maiores valores MOS é $\gamma=[0.6,0.8]$ ao invés da região $\gamma=[0.8,0.9]$ [2]. O valor utilizado para o coeficiente de ponderação para os demais testes foi o de $\gamma = 0,75$ para o CELP 32.

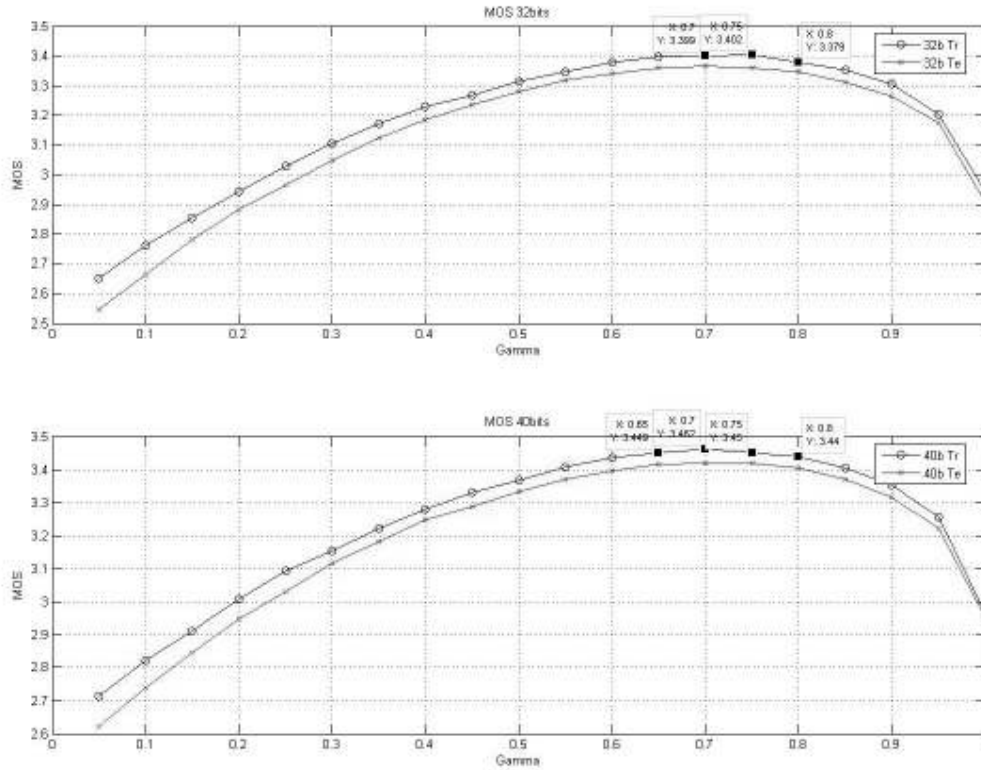


Figura 5. 4 - Testes do valor ótimo de γ .

O resultado da tabela 5.7 mostra a comparação entre as versões do CELP 32 com $\gamma = 0,8$ e $\gamma = 0,75$, mostra uma melhora de aproximadamente 0,02 MOS para o desempenho relativo ao banco de treinamento.

Tabela 5. 7 - Comparação de qualidade entre versões com diferentes valores de γ para o CELP 32.

γ	Tr	Te
0,8	3,379	3,345
0,75	3,402	3,360

5.6 Requantização dos ganhos G_a e G_f

Após a escolha do coeficiente de ponderação para o CELP 32 $\gamma = 0,75$, foram feitos testes com a finalidade de otimizar os ganhos G_a e G_f dos Dicionários Adaptativo e Fixo, respectivamente.

Como o ganho G_f depende do ganho G_a , mas o inverso não é verdadeira, os testes começaram requantizando o ganho G_a com a resolução de bits variando de 8 bits a 1 bit, utilizando o dicionário de quantização para o ganho G_f do CELP inicial, cuja resolução de bits é 5 bits. O resultado destes testes encontra-se na Figura 5. 5.

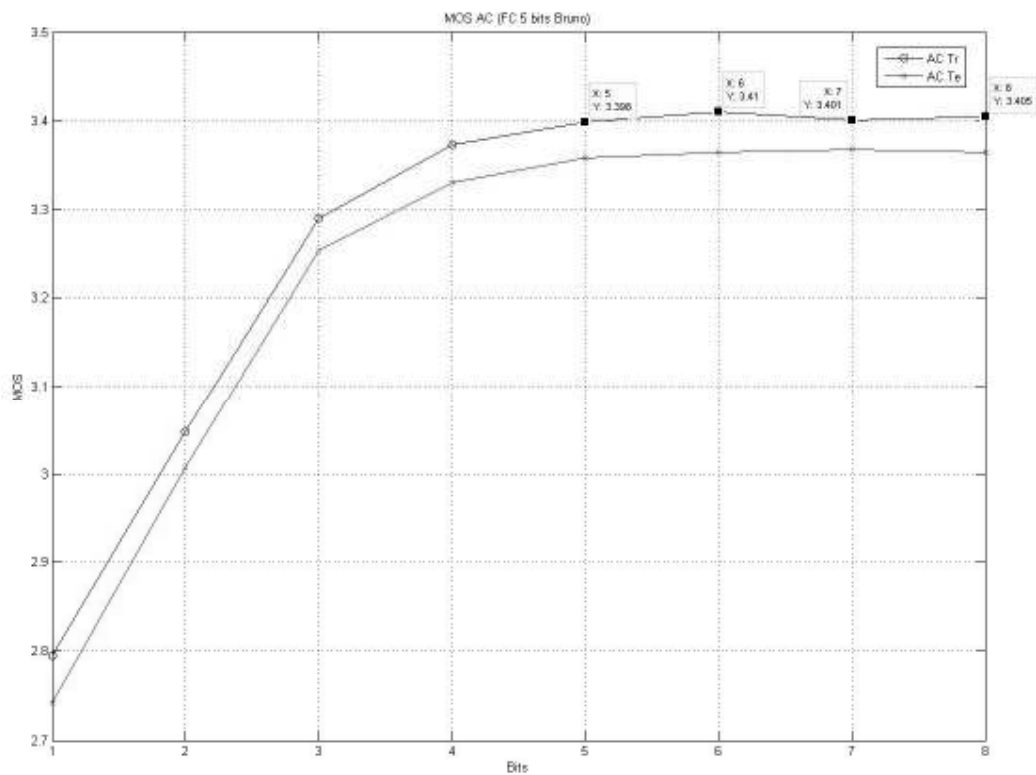


Figura 5. 5 - Teste para a resolução de bits de G_a .

Como mostrado na Figura 5. 5, o valor da resolução de bits ótima para o CELP 32 é de 6 bits, levando em consideração o compromisso entre qualidade de codificação e taxa de transmissão. A tabela 5.8 mostra a comparação entre o desempenho codificação das versões do CELP 32 com as quantizações de 5 bits para G_a e G_f originais do CELP Inicial e o CELP 32 com G_a requantizado com 6 bits e G_f quantizado com os 5 bits da maneira original do CELP Inicial, onde vemos que houve uma melhora de aproximadamente 0,01 MOS relativo ao banco de treinamento.

Tabela 5. 8 - Comparação do desempenho entre o CELP 32 com Ga quantizado com 5 e 6 bits.

Quantização de Ga	Tr	Te
5 bits	3,402	3,360
6 bits	3,410	3,364

Com o ganho Ga do dicionário adaptativo quantizado com 6 bits, foram feitos os mesmo testes para o ganho Gf do dicionário fixo, isto é, variando sua resolução de bits de 8 bits a 1 bit e seu resultado é exibido na Figura 5. 6.

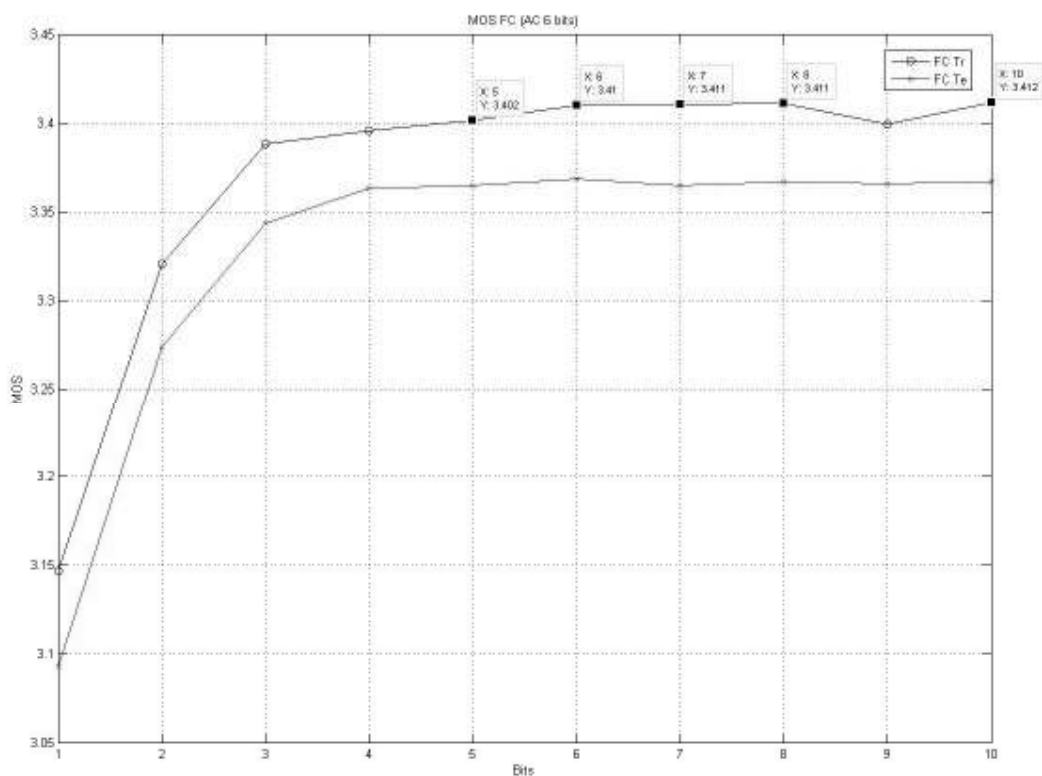


Figura 5. 6 - Teste da resolução de bits para Gf.

Vemos na Figura 5. 6 que o MOS varia pouco a partir de 5 bits de resolução, sendo a resolução de 6 bits seria escolhida para Gf, a fim de se manter o compromisso qualidade de codificação e taxa de transmissão, caso o desempenho não tivesse sido igual ao do codificador com Ga requantizado com 6 bits e Gf quantizado com 5 bits com o dicionário de quantização do CELP Inicial, o que fez com que se mantivesse o ganho Gf do dicionário fixo quantizado com 5 bits.

5.7 Reavaliação de Ka e Kf

Após a escolha da resolução de 6 bits para Ga e 5 bits para Gf, as quantidades de melhores excitações armazenadas Ka e Kf, para o duplo ciclo dos dicionários adaptativo e fixo, respectivamente, foram reavaliadas a fim de verificar seus valores ótimos levando em consideração a qualidade de codificação e a complexidade computacional que varia diretamente com os valores de Ka e Kf, como é explicado na seção 4.2.2.

Foram feitos, então, testes variando o valor de Ka de 0 a 100, com variação de 5 unidades, como mostram a Figura 5. 7 e a Figura 5. 8. Escolheu-se então a região de Ka variando de 5 a 15, correspondente aos maiores valores MOS e menor complexidade computacional (menor valor de Ka), em que o valor MOS máximo foi de 3,413 para Ka=45 e para Ka=10 o valor MOS foi de 3,412, sendo a diferença imperceptível para o ouvido humano, mas a diferença na complexidade é de 9.910 para Ka=10 contra 23.510 para Ka=45, 237% superior que o primeiro caso.

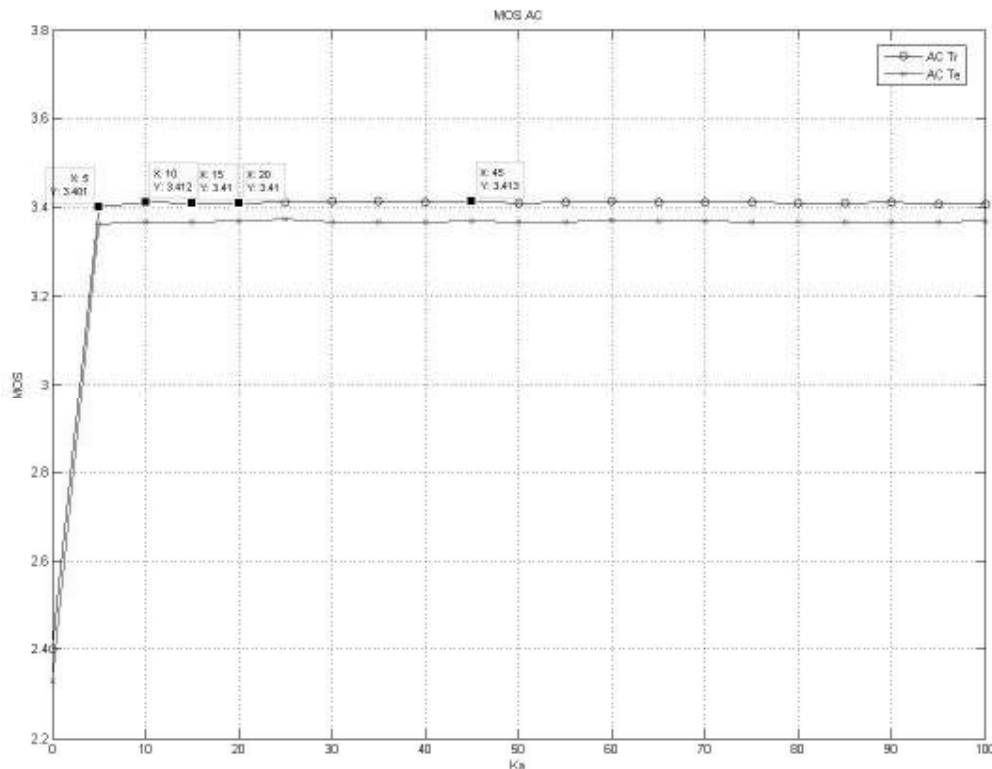


Figura 5. 7 - Resultado dos testes de qualidade para Ka entre 0 e 100.

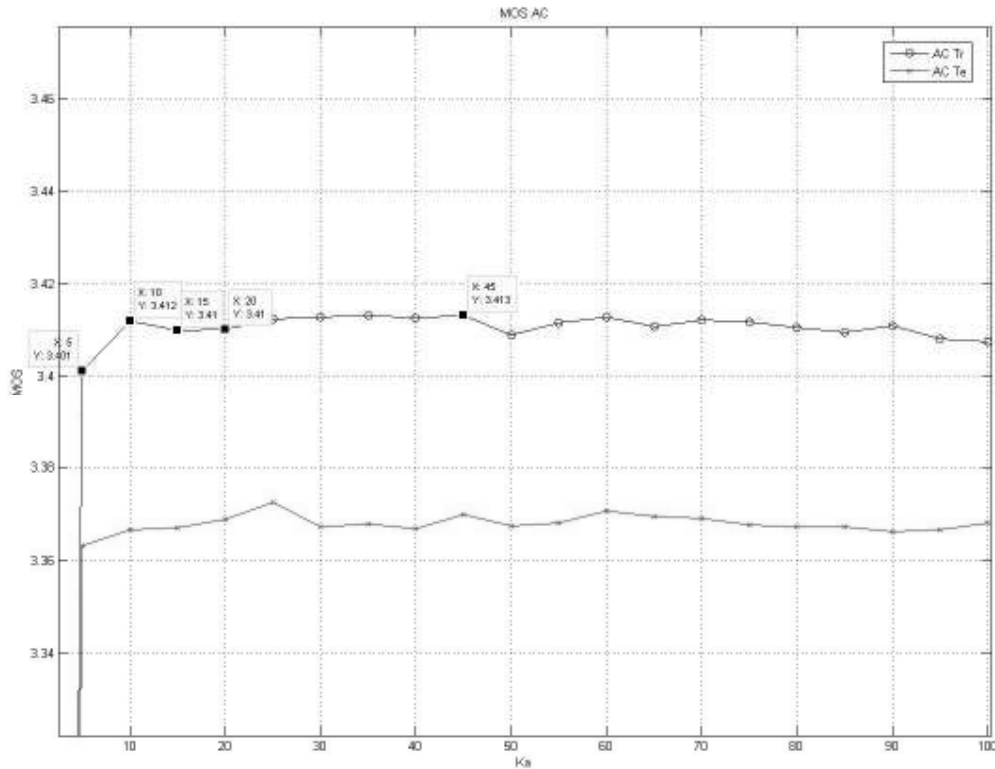


Figura 5. 8 - Zoom do resultado dos testes de qualidade para Ka entre 0 e 100.

Variou-se, então, os valores de Ka dentro desta região com passos de 1 unidade, cujo resultado é ilustrado pela Figura 5. 9, que mostra uma curva com aspecto de ruído, o que significa que o valor de Ka gerador da melhor qualidade da região pode ser qualquer um destes, mas como Ka =10 foi o que obteve o maior MOS, este valor de Ka foi escolhido. A tabela 5.9 mostra a comparação entre a qualidade de codificação e complexidade computacional para Ka = 20 (utilizado no CELP Inicial) e Ka = 10.

Tabela 5. 9 - Comparação de desempenho e complexidade computacional entre Ka=20 e Ka=10.

Ka	Tr	Te	Multiplicações
20	3,410	3,369	23.510
10	3,412	3,367	9.910

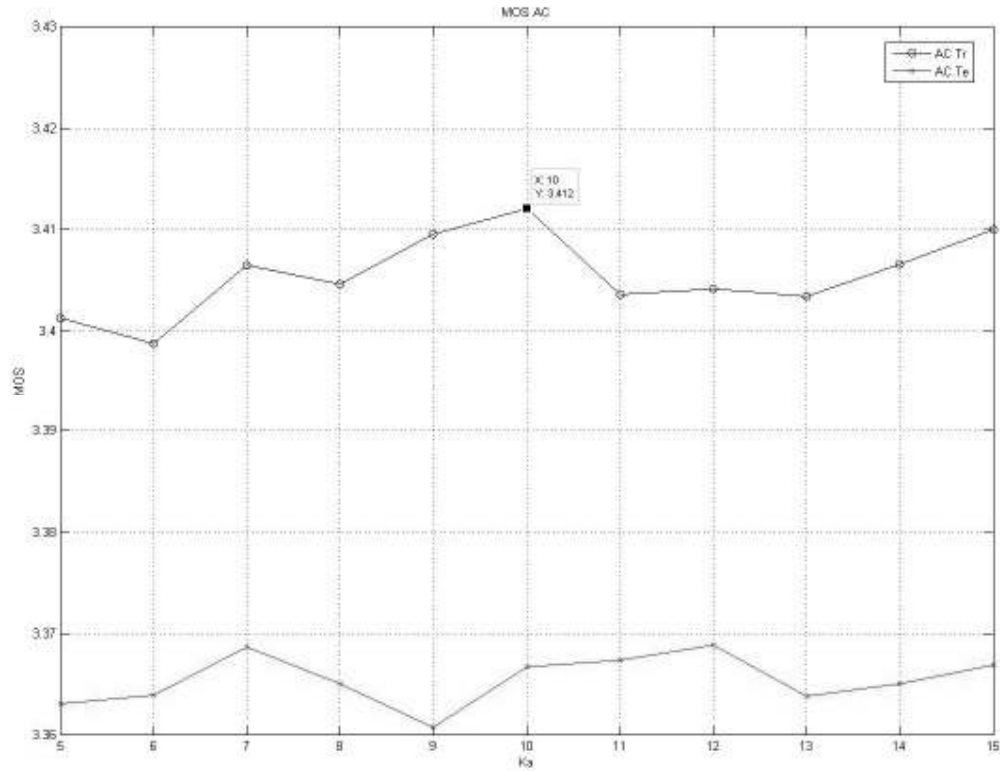


Figura 5. 9 - Testes de qualidade para a região de Ka entre 5 e 15.

Fixando então em 10 o valor de Ka, fez-se o mesmo teste para Kf, isto é, variando seu valor entre 0 e 100, com passos de 5 unidades. As figuras 5.10 e 5.11 mostram o resultado destes testes, em que o valor MOS máximo se dá para Kf=20 e Kf=30, o primeiro sendo menos complexo computacionalmente, o que é desejado. A região escolhida para testes mais detalhados, foi a de Kf entre 15 e 25.

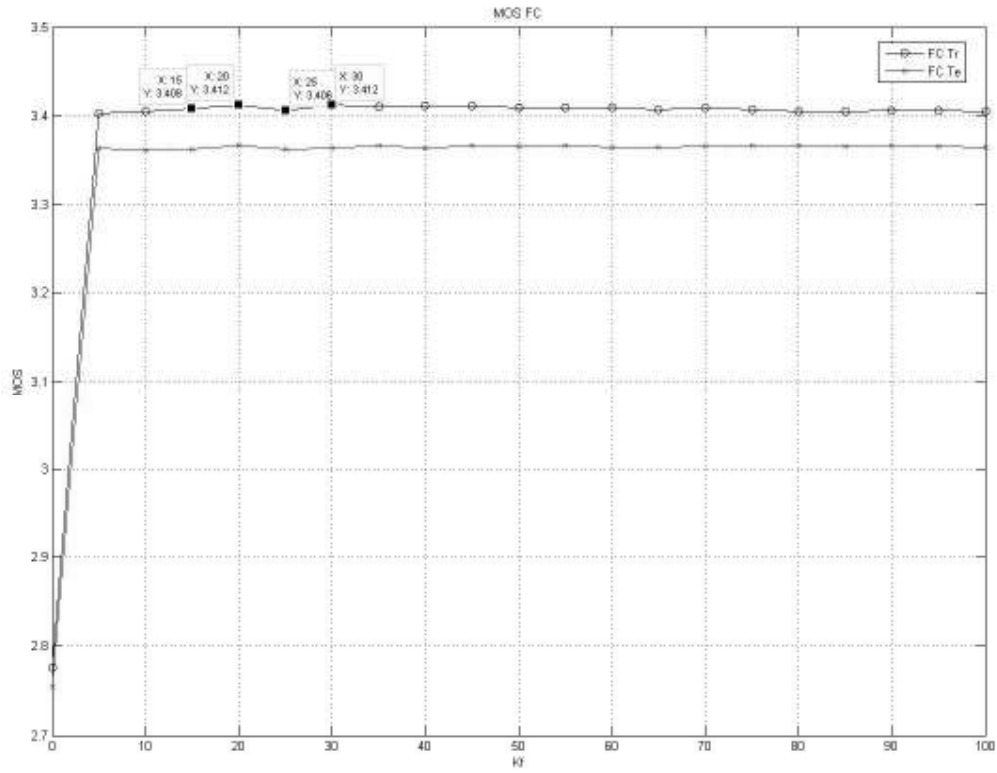


Figura 5. 10 - Resultado dos testes de qualidade para Kf entre 0 e 100.

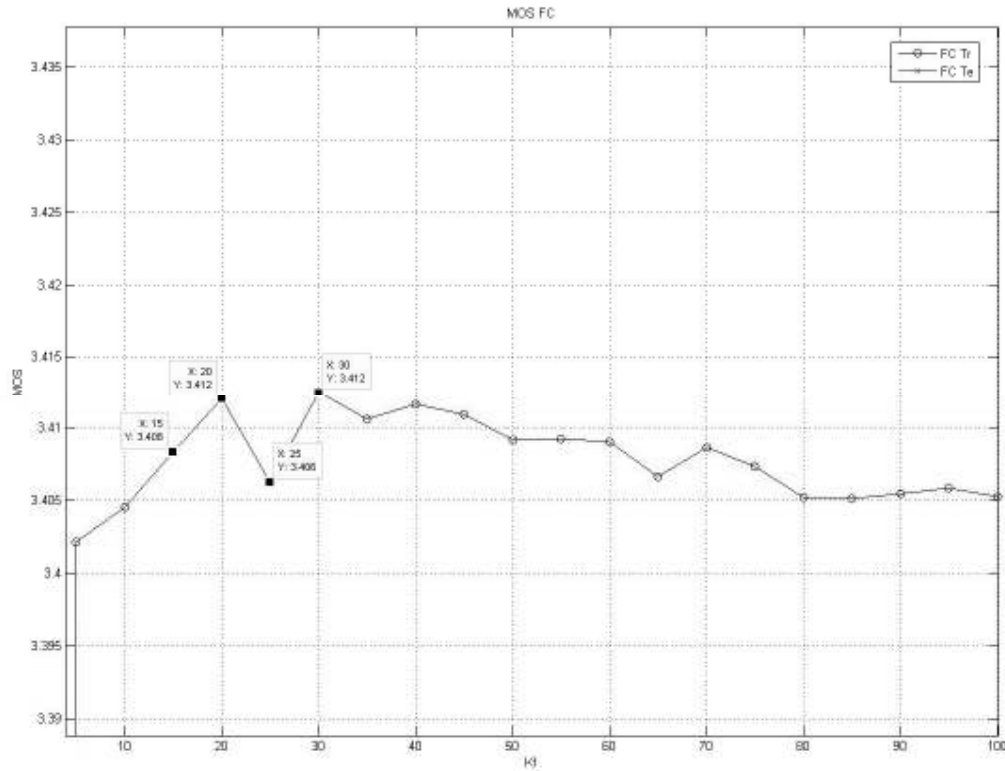


Figura 5.11 - Zoom do resultado dos testes de qualidade para Kf entre 0 e 100.

Na região delimitada por $K_f=15$ e $K_f=25$, foram feitos testes variando K_f com passos de 1 unidade, com a finalidade de determinar o valor de K_f que gera o maior MOS como mostra a figura 5.12. Assim como ocorrido para K_a , a curva mostrada na figura 5.12 tem um comportamento ruidoso, sendo o valor ótimo para K_f difícil de ser determinado, mas como o valor de $K_f=16$ foi o que apresentou o maior MOS, foi o escolhido para ser incorporado ao sistema CELP. A comparação da qualidade de codificação e a complexidade computacional do codificador CELP para $K_f=20$ e $K_f=16$ encontra-se na tabela 5.10, que mostra a grande diferença do número de multiplicações para uma mesma qualidade de codificação.

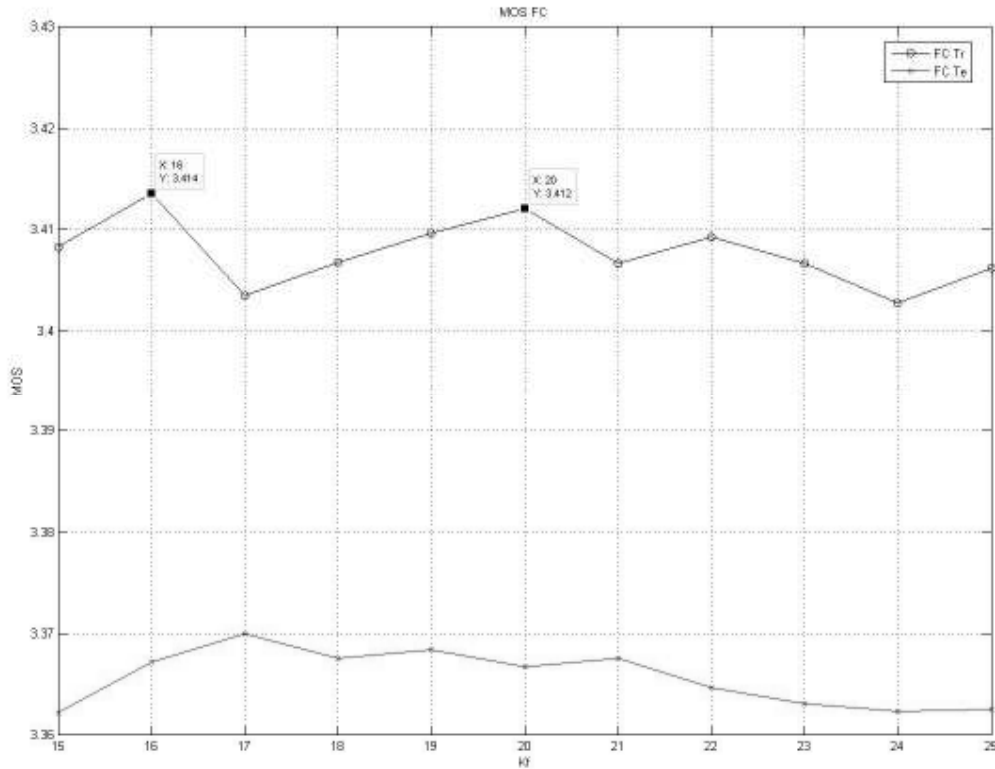


Figura 5. 12 - Resultado dos testes de qualidade para a região de Kf entre 15 e 25.

Tabela 5. 10 - Comparação de desempenho e complexidade computacional entre Kf=20 e Kf=15.

Kf	Tr	Te	Multiplicações
20	3,412	3,367	23.510
16	3,414	3,367	11.910

5.8 Codificador final sem detecção de silêncio

Esta seção é um resumo dos resultados obtidos pelos testes feitos nas seções anteriores deste capítulo 5. A tabela 5.11 mostra a comparação entre o codificador inicial e o codificador final, em que as taxas de transmissão de 7,6 kbps e 7,4 kbps, respectivamente, são similares mas a complexidade computacional é menor, visto que tanto Ka quanto Kf são menores para o codificador final e sua qualidade de codificação é de 0,232 para o banco de treinamento e de 0,234 para o banco de teste.

Tabela 5. 11 - Comparação entre o Codificador Inicial e o Codificador Final.

Parâmetro	CELP Inicial	CELP Final
DLSF(0) a DLSF(9)	40 bits	32 bits
Ga	5 bits	6 bits
Gf	5 bits	5 bits
Ka	20 excitações	10 excitações
Kf	20 excitações	16 excitações
γ	0,8	0,75
Dicionário Adaptativo	1024 excitações	512 excitações
Dicionário Fixo	256 excitações	512 excitações
Taxa de Transmissão	7,6 kbps	7,4 kbps
MOS médio Tr	3,182	3,414
MOS médio Te	3,133	3,367

5.9 Detecção de silêncio

A detecção de silêncio é o nome dado ao processo de identificar se um determinado bloco representa um som do tipo silêncio e, ao invés de transmitir tal bloco da maneira comum, envia-se apenas um identificador de silêncio, deixando a cargo de o decodificador reconstituir este bloco, seja inserindo amostras de valor nulo, ruído branco, ou algum tipo de silêncio pré-gravado.

O primeiro passo a ser dado, foi o de determinar como seria feita a identificação de um bloco de silêncio. Para isso, gravou-se em um arquivo do tipo texto todas as amostras dos sinais do Banco de Treinamento, montando-se um único sinal bastante extenso.

Selecionou-se, então, uma região aleatória deste arquivo contendo 1024 blocos de 20 ms para teste como mostra a figura 5.13. As primeiras 23.110 amostras desta região foram selecionadas para caracterizar o silêncio, sendo então aplicadas janelas retangulares de 20 ms nesta segunda região e calculando a energia e a taxa de cruzamento por zero de cada bloco e calculando-se as médias da energia e da taxa de cruzamento por zero, sendo $E_{lim} = 50 \text{ dB}$ e $TCZ_{lim} = 57$, a Energia limite e a taxa de cruzamento por zero limite.

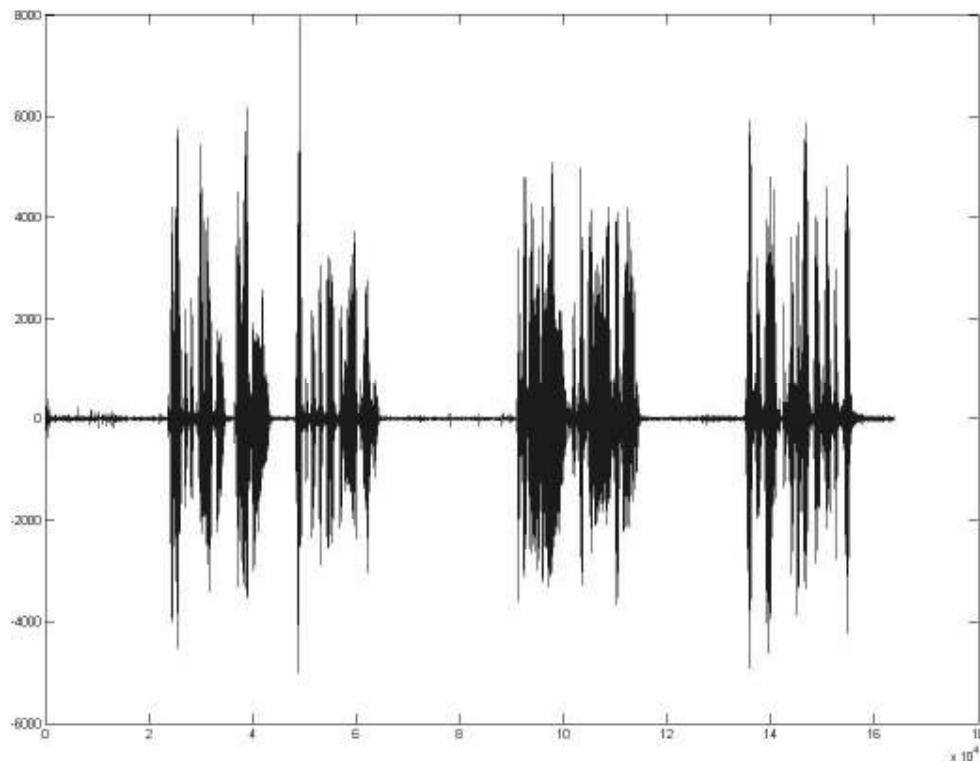


Figura 5. 13 - Região selecionada para os testes de detecção de silêncio.

Fez-se então um teste de detecção de silêncio em toda a região, da seguinte forma:

1. Calculou-se a Energia E de cada bloco;
2. Se $E > E_{lim}$, o bloco é classificado como sendo do tipo Sonoro. Caso contrário calcula-se a taxa de cruzamento por zero TCZ ;
3. Se $TCZ > TCZ_{lim}$, o bloco é classificado como sendo do tipo Surdo. Caso contrário é classificado como sendo do tipo Silêncio;
4. Os blocos do tipo Silêncio são substituídos por blocos contendo todas as amostras de valor nulo, e os outros blocos tem suas amostras mantidas.

O resultado deste teste é mostrado na figura 5.14. Destaca-se a classificação errada de blocos do tipo Plosivo como Silêncio, mas para isso deveria ser feito um teste mais sofisticado de detecção de silêncio, isto é, não somente fazendo a comparação de limiares, o que aumentaria bastante a complexidade computacional.

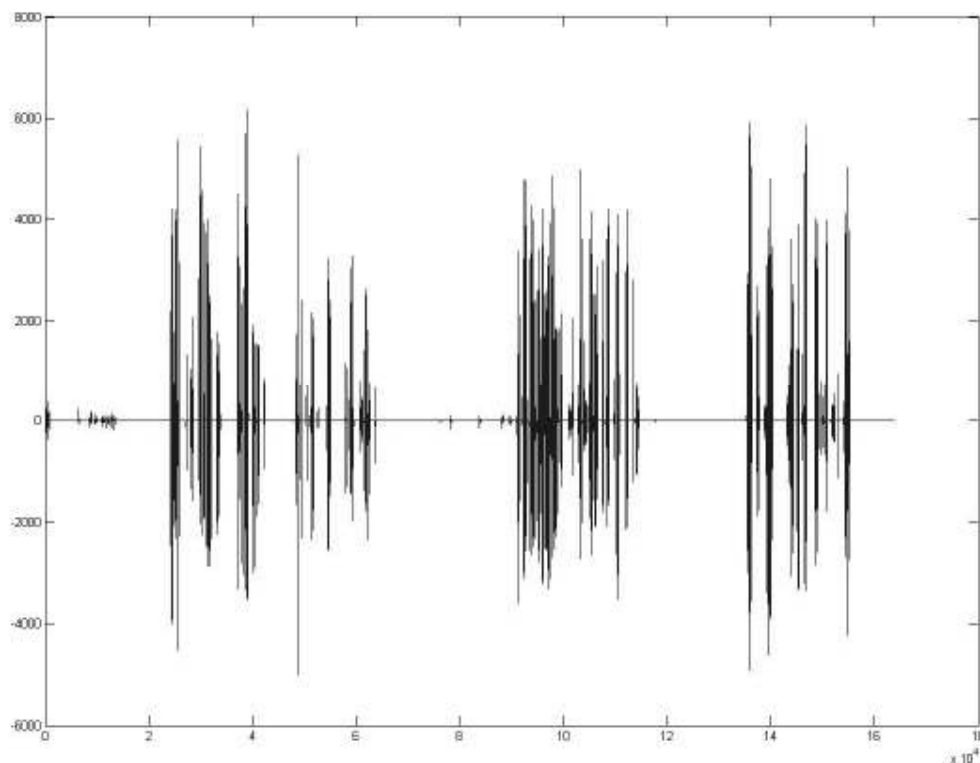


Figura 5.14 - Resultado do teste de detecção de silêncio na região de teste.

Para verificar se a detecção de silêncio feita desta maneira é satisfatória, fez-se então o mesmo teste da figura 5.14 para ambos os Bancos de Treinamento e Teste, armazenando-se a classificação de todos os blocos de sinal das frases do Banco de Treinamento, verificando a porcentagem de blocos de Silêncio em relação ao número total de blocos e a taxa média de transmissão, visto que ela se torna variável, uma vez que depende agora da quantidade de silêncio presente no sinal a ser codificado.

A tabela 5.12 mostra o resultado dos testes de detecção de silêncio feita nos Bancos de Treinamento e Teste, com diversos valores de E_{lim} , com a finalidade de mostrar a relação entre quantidade de blocos considerados Silêncio (coluna % silêncio Tr), a qualidade de codificação (colunas Tr e Te) e a taxa média de transmissão (coluna Tx (kbps)). A linha em negrito correspondente ao limiar $E_{lim} = 50 \text{ dB}$, calculado como explicado anteriormente, mostra que a qualidade de codificação é insatisfatória, pois há uma diminuição de 0,981 ou 28,73% do MOS médio referente ao Banco de Treinamento, apesar de uma queda de 39,50% da taxa média de transmissão.

Tabela 5. 12 - Resultado dos testes de detecção de silêncio para diferentes valores de E_{lim} .

E_{lim} (dB)	Tr	Te	% silêncio Tr	Tx (kbps)
18,0	3,409	3,362	0,51	7,36
28,0	3,409	3,362	0,52	7,36
38,0	3,252	3,227	5,43	7,00
39,0	3,165	3,146	8,95	6,74
39,5	3,123	3,109	10,88	6,60
40,0	3,074	3,072	13,01	6,44
48,0	2,608	2,547	36,47	4,70
50,0	2,433	2,363	39,50	4,48
58,0	1,745	1,709	53,17	3,47
68,0	1,381	1,372	62,38	2,78

A fim de melhorar a qualidade de codificação, foram feitos dois novos testes para $E_{lim} = 38 \text{ dB}$, $E_{lim} = 39 \text{ dB}$ e $E_{lim} = 39,5 \text{ dB}$, um substituindo-se agora o bloco de silêncio por um bloco de ruído branco com energia de bloco $E = 43 \text{ dB}$ e o outro por um bloco de silêncio com energia de bloco $E = 43 \text{ dB}$, ao invés de um bloco contendo amostras nulas. Ambos os blocos de ruído branco e de silêncio foram pré-gravados, com o primeiro sendo gerado no MATLAB com a função rand() e o segundo sendo retirado da região de testes ilustrada na figura 5.13.

A tabela 5.13 mostra o resultado dos testes para escolher a melhor maneira de substituir os blocos classificados como silêncio. A coluna bloco indica o tipo de substituição utilizada, nulo para a substituição por um bloco formado apenas por amostras nulas, ruído para a substituição por um bloco formado por ruído branco pré-gravado e silêncio para a substituição por um bloco formado por silêncio pré-gravado. A linha em negrito foi escolhida para ser incorporada ao codificador por apresentar uma diminuição de cerca de 11% da taxa média de transmissão e uma diminuição da qualidade de codificação em cerca de 5% em relação ao codificador final sem detecção de silêncio e um aumento da qualidade em cerca de 2% em relação ao codificador inicial, ambos relativos ao banco de treinamento.

Tabela 5. 13 - Resultado dos testes de substituição do bloco de silêncio.

E_{lim} (dB)	Bloco	Tr	Te
38,0	Nulo	3,252	3,227
	Ruído	3,248	3,222
	Silêncio	3,305	3,270

39,0	Nulo	3,165	3,146
	Ruído	3,198	3,167
	Silêncio	3,268	3,225
39,5	Nulo	3,123	3,109
	Ruído	3,170	3,142
	Silêncio	3,242	3,205

5.10 Comparação entre os codificadores

A tabela 5.14 resume de forma comparativa os codificadores inicial, final sem detecção de silêncio e final com detecção de silêncio e a Figura 5. 15 mostra a comparação entre os codificadores da Figura 2.3 com os da tabela 5.14. O resultado do codificador final com detecção de silêncio (CELP Final DS) é considerado bom, pois, apesar de não ter uma melhora expressiva em relação ao CELP Inicial, a melhora na taxa média de transmissão é bastante representativa.

Tabela 5. 14 - Comparação entre os Codificadores.

Codificador	Taxa média	Tr	Te
CELP Inicial	7,60 kbps	3,182	3,133
CELP Final	7,40 kbps	3,414	3,367
CELP Final DS	6,60 kbps	3,242	3,205

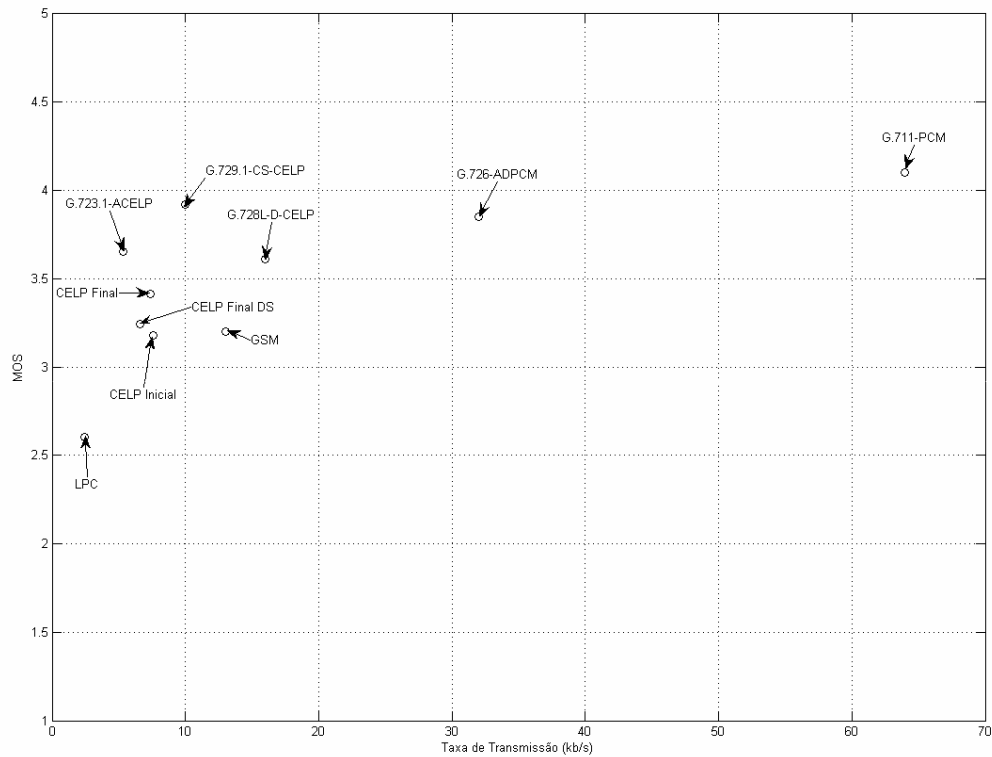


Figura 5.15 - Comparação entre os codificadores da Figura 2.3 com os da Tabela 5.14.

5.11 Conclusão

Vimos então no capítulo 5 os resultados obtidos pela introdução do pré-processamento e do pós-processamento e pela reavaliação da quantização dos coeficientes DLSF, da quantização dos ganhos dos dicionários adaptativo e fixo, dos pesos da interpolação dos coeficientes LPC e dos coeficientes K_a e K_f do duplo ciclo.

Foi verificado que cada um desses casos gera um aumento na qualidade do sinal reconstituído ao final do sistema. Os pré- e pós- processamentos geram uma boa melhoria na qualidade do sinal, sendo definitivamente adicionados ao codificador sendo utilizado no estudo das outras alterações.

A inclusão da detecção de silêncio teve um resultado considerado bom, pois aumentou a qualidade de codificação em relação ao CELP Inicial, mesmo que pouco significativamente, e a

diminuição da taxa média de transmissão em quase 11% é bastante significativo no contexto de codificação de voz.

Capítulo 6

Conclusão

6.1 Contribuições do trabalho

Este projeto apresentou reavaliações de alguns processos e inclusões de outros ao sistema CELP de codificação de voz. As reavaliações realizadas neste trabalho foram feitas por serem consideradas, e provadas, importantes.

O capítulo 2 informou de maneira pouco detalhada como é feita a codificação de voz através de exemplos de diversas técnicas, com ênfase no codificador CELP, por ser o objeto de estudo deste trabalho.

O capítulo 3 explica o que é um banco de voz, detalhando o banco de treinamento e o banco de teste, ambos utilizados neste trabalho, assim como foram obtidos.

O capítulo 4 informa a situação em que se encontrava o codificador CELP antes de serem implementadas as modificações feitas neste trabalho. Sendo intitulado de codificador inicial, este codificador teve suas características principais detalhadas, tais como taxa de transmissão, qualidade de codificação e complexidade computacional.

No capítulo 5, encontram-se as alterações no CELP Inicial realizadas neste trabalho de maneira detalhada, cujos resultados foram todos satisfatórios, pois melhoraram ora a qualidade de transmissão ora a complexidade computacional ora a taxa média de transmissão. As requantizações dos coeficientes DLSF e dos ganhos dos dicionários adaptativo e fixo foram feitas de forma bastante gradativa a fim de serem realizadas da melhor maneira possível. A reavaliação do duplo ciclo foi feita de maneira também cuidadosa, pois a complexidade computacional é algo importante, pois quanto mais rápida é feita uma codificação, mais rápido podem ser feitos testes para buscar melhorias no sistema. Os pesos da interpolação foram reavaliados e a melhora na qualidade de codificação foi significativa.

A inclusão do bloco de detecção de silêncio era um dos focos deste trabalho, por não fazer parte do sistema do CELP Inicial e ser um bloco bastante importante pelo ponto de vista da taxa de transmissão que é um parâmetro muito importante no contexto de Telecomunicações. O desempenho do CELP Final DS foi considerado bom por ter diminuído significativamente a taxa média de transmissão sem diminuir muito a qualidade de codificação, em relação ao CELP Final.

6.2 Propostas para trabalhos futuros

Como mencionado anteriormente, este trabalho é uma continuação de um outro trabalho [4] intitulado Codificador Inicial neste trabalho. Por se tratar de um tema bastante amplo, há diversos incentivos para melhorar ainda mais este Sistema de Codificação de Voz. A seguir estão algumas sugestões de continuação deste trabalho:

- Estudo sobre uma ou mais formas mais sofisticadas de fazer a identificação de blocos de Silêncio, de forma a não interpretar de maneira errada blocos do tipo Plosivo, por exemplo, como Silêncio;
- Estudo sobre uma outra maneira de substituir o bloco já identificado como Silêncio, de maneira se aproximar mais do desempenho da codificação sem detecção de silêncio;
- Estudo sobre a interpolação dos coeficientes LSF, de forma a melhorar a codificação, sem aumentar muito ou até diminuindo a complexidade computacional;
- Estudo da possibilidade de um codificador de taxa variável, controlada externamente;

Referências Bibliográficas

- [1] J. R. Deller, J. G. Proakis, J. H. L. Hansen, “*Discrete-Time Processing of Speech Signals*”, MacMillan Coll Div, 1995.
- [2] A. M. Kondoz, “*Digital Speech: Coding for Low Bit Rate Communications Systems*”, Wiley, 1999.
- [3] P. S. R. Diniz, E. A. B. da Silva, S. L. Netto, “*Processamento Digital de Sinais: Projeto e Análise de Sistemas*”, Bookman, 2004.
- [4] B. C. Bispo, “*Otimização do codificador de voz CELP*”, Poli/UFRJ, 2005.
- [5] V. L. Latsch, “*Projeto Maritaca*”, COPPE/UFRJ.
- [6] ITU-T *Recommendation P.800: “Methods for subjective determination of transmission quality*”, 1996.
- [7] ITU-T *Recommendation P.861: “Objective quality measurement of telephoneband(300-3400 Hz) speech codecs*”, 1998.
- [8] ITU-T *Recommendation P.729: “Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)”*, 1996.
- [9] B. B. Oliveira, “*Análise e testes de um codificador CELP*”, Poli/UFRJ, 2001.
- [10] F. C. C. B. Diniz, “*Implementação de um codificador de voz CELP em tempo real*”, Poli/UFRJ, 2003.
- [11] R. S. Maia, “*Codificação CELP e análise espectral de voz*”, COPPE/UFRJ, 2000.
- [12] *The Open Speech Repository*, http://www.voiptroubleshooter.com/open_speech/index.html.

Apêndice A

Este apêndice contém a tabela A.1, que caracteriza de forma detalhada os bancos de treinamento e teste descritos no capítulo 3.

Tabela A 1 - Caracterização detalhada dos arquivos dos bancos de treinamento e teste.

Nome	Idioma	Sexo	Energia	% silêncio	Duração	Máximo	Mínimo	Te/Tr
ch1.wav	CH	M	78,67348	43,40153	3,82475	0,502594	-0,32465	Tr
ch10.wav	CH	M	58,13249	42,97559	3,211125	0,66568	-0,38242	Tr
ch11.wav	CH	M	64,53865	42,06646	4,61175	0,516113	-0,31366	Tr
ch17.wav	CH	M	102,0848	38,11691	3,5155	0,409149	-0,32953	Tr
ch2.wav	CH	M	57,23958	39,0106	3,5375	0,4198	-0,28189	Tr
ch20.wav	CH	M	78,81723	76,58643	10,054	0,49408	-0,30533	Tr
ch5.wav	CH	M	80,7362	33,61685	4,938	0,40744	-0,25583	Tr
ch6.wav	CH	M	81,62686	29,72533	4,373375	0,453918	-0,4111	Tr
ch9.wav	CH	M	72,5686	37,82888	6,18575	0,459686	-0,28494	Tr
fr104.wav	FR	M	109,4457	81,03629	9,427875	0,507111	-0,36304	Tr
fr13.wav	FR	M	80,70157	59,27552	4,555	0,556061	-0,33633	Tr
fr14.wav	FR	M	106,3145	58,50567	6,15325	0,560242	-0,31601	Tr
fr16.wav	FR	M	40,95635	71,77641	4,235375	0,312042	-0,2663	Tr
fr18.wav	FR	M	99,32536	77,42237	8,5505	0,378601	-0,24616	Tr
fr2.wav	FR	M	81,64773	63,2519	4,395125	0,427368	-0,29111	Tr
fr20.wav	FR	M	60,39702	57,61773	4,96375	0,391571	-0,24521	Tr
fr22.wav	FR	M	96,26124	58,9049	5,873875	0,417206	-0,25406	Tr
fr27.wav	FR	M	73,31785	54,85821	4,302	0,428864	-0,33405	Tr
fr28.wav	FR	M	37,69842	66,10322	5,29475	0,321655	-0,21143	Tr
fr30.wav	FR	M	86,30397	63,73481	5,460125	0,349304	-0,2511	Tr
fr32.wav	FR	M	131,3737	55,16921	6,452875	0,343842	-0,27252	Tr
fr33.wav	FR	M	59,9777	72,14294	5,211875	0,589142	-0,27759	Tr
fr34.wav	FR	M	52,42336	65,1863	4,632875	0,293365	-0,22418	Tr
fr37.wav	FR	M	64,76558	67,89503	6,71625	0,307892	-0,30722	Tr
fr38.wav	FR	M	66,71393	62,18975	5,3385	0,516327	-0,27249	Tr
fr41.wav	FR	M	73,64793	61,93912	5,683	0,468262	-0,26682	Tr
fr42.wav	FR	M	30,87786	80,46454	4,822	0,322693	-0,21314	Tr
fr43.wav	FR	M	117,6162	57,45721	6,544	0,407196	-0,26166	Tr
fr44.wav	FR	M	62,96495	75,6233	7,405125	0,327728	-0,31699	Tr
fr46.wav	FR	M	67,15499	64,56375	5,08025	0,358948	-0,21954	Tr
fr47.wav	FR	M	72,12539	71,31214	4,908	0,288849	-0,27707	Tr
fr49.wav	FR	M	74,7453	72,66845	6,027375	0,411804	-0,23279	Tr
fr52.wav	FR	M	81,66388	49,73688	5,629625	0,416321	-0,26062	Tr
fr53.wav	FR	M	107,9305	55,81395	4,945	0,402557	-0,34119	Tr

fr56.wav	FR	M	87,27913	60,10599	5,590125	0,337128	-0,31824	Tr
fr59.wav	FR	M	82,10241	49,88227	5,7335	0,495239	-0,23377	Tr
fr6.wav	FR	M	125,3491	59,6399	3,755875	0,563019	-0,35873	Tr
fr60.wav	FR	M	42,16579	69,55797	4,80175	0,302338	-0,2117	Tr
fr61.wav	FR	M	82,04496	52,92596	4,15675	0,565247	-0,38419	Tr
fr62.wav	FR	M	84,00525	52,2111	4,4435	0,472107	-0,29117	Tr
fr63.wav	FR	M	63,05197	56,72948	4,371625	0,428558	-0,31702	Tr
fr67.wav	FR	M	92,01453	61,30916	4,730125	0,437683	-0,27899	Tr
fr7.wav	FR	M	99,47657	56,31211	5,434	0,437775	-0,29761	Tr
fr70.wav	FR	M	27,57301	56,96462	3,967375	0,358765	-0,12936	Tr
fr74.wav	FR	M	69,29192	58,52728	4,71575	0,525452	-0,26642	Tr
fr8.wav	FR	M	79,7868	61,2999	5,513875	0,543945	-0,26157	Tr
fr80.wav	FR	M	59,26936	64,88704	5,239875	0,318604	-0,20194	Tr
fr82.wav	FR	M	62,27193	55,13439	4,71575	0,328613	-0,22577	Tr
fr91.wav	FR	M	75,3418	61,00672	4,556875	0,396057	-0,26285	Tr
fr92.wav	FR	M	51,47122	54,15256	4,321125	0,290314	-0,18887	Tr
fr94.wav	FR	M	134,2095	44,54855	5,65675	0,556091	-0,28125	Tr
fr96.wav	FR	M	87,934	59,24383	4,321125	0,523712	-0,25687	Tr
in1.wav	IN	M	54,56085	57,37249	4,04375	0,434906	-0,31924	Tr
in11.wav	IN	M	41,65375	53,10211	4,180625	0,325348	-0,19412	Tr
in17.wav	IN	M	94,5762	50,90236	6,483	0,44223	-0,28226	Tr
in18.wav	IN	M	108,3138	42,32032	5,482	0,395782	-0,31705	Tr
in2.wav	IN	M	56,11237	54,32243	3,424	0,450775	-0,26114	Tr
in21.wav	IN	M	108,7245	55,84491	6,912	0,571594	-0,31229	Tr
in23.wav	IN	M	56,28483	62,09519	4,766875	0,327179	-0,22	Tr
in24.wav	IN	M	58,26977	55,66373	4,814625	0,422638	-0,2402	Tr
in25.wav	IN	M	69,50674	59,56311	5,338875	0,417389	-0,25549	Tr
in27.wav	IN	M	67,035	57,83257	3,87325	0,559814	-0,34848	Tr
in29.wav	IN	M	38,61058	65,06236	4,119125	0,474243	-0,2551	Tr
in36.wav	IN	M	48,40062	61,96347	3,87325	0,454681	-0,25824	Tr
in37.wav	IN	M	58,00191	54,48954	4,918375	0,465668	-0,3027	Tr
in4.wav	IN	M	92,85125	49,63181	4,634125	0,363373	-0,2626	Tr
in43.wav	IN	M	111,1361	60,4433	6,98175	0,462677	-0,33255	Tr
in45.wav	IN	M	55,23424	68,03129	4,586125	0,563446	-0,41907	Tr
in46.wav	IN	M	85,8797	63,17556	5,06525	0,596466	-0,43097	Tr
in47.wav	IN	M	73,69362	57,26947	5,133625	0,438904	-0,21634	Tr
in50.wav	IN	M	103,8415	55,13059	7,2555	0,485046	-0,27048	Tr
in52.wav	IN	M	75,64545	54,17153	5,464125	0,411957	-0,26767	Tr
in54.wav	IN	M	67,75121	56,28141	4,4775	0,405273	-0,19119	Tr
in55.wav	IN	M	72,41569	55,93277	3,71875	0,338379	-0,2757	Tr

in56.wav	IN	M	90,12144	50,51079	6,374875	0,391968	-0,2338	Tr
in59.wav	IN	M	53,22866	55,80855	3,870375	0,383575	-0,35071	Tr
in61.wav	IN	M	96,70515	54,3992	4,522125	0,427917	-0,29044	Tr
in62.wav	IN	M	114,4753	47,05882	5,2275	0,443024	-0,26849	Tr
in63.wav	IN	M	86,66021	58,14568	4,024375	0,491638	-0,23773	Tr
in64.wav	IN	M	144,7976	48,20648	7,467875	0,461487	-0,33789	Tr
in65.wav	IN	M	126,5564	47,8169	5,144625	0,532654	-0,3157	Tr
in68.wav	IN	M	23,53651	64,47482	3,319125	0,419952	-0,2576	Tr
in71.wav	IN	M	49,33376	61,55326	4,678875	0,363831	-0,26508	Tr
in73.wav	IN	M	45,57726	59,44584	5,955	0,282898	-0,2045	Tr
in74.wav	IN	M	41,80518	60,76499	5,5295	0,287384	-0,20093	Tr
in8.wav	IN	M	35,16099	60,45198	3,54	0,424133	-0,19882	Tr
in80.wav	IN	M	49,28175	55,02788	7,777875	0,262207	-0,19836	Tr
in9.wav	IN	M	55,29383	54,29352	3,978375	0,35437	-0,24799	Tr
uk100.wav	UK	M	93,90598	53,85682	4,419125	0,333374	-0,30627	Tr
uk103.wav	UK	M	28,06261	54,63566	3,88025	0,229401	-0,22455	Tr
uk104.wav	UK	M	33,8056	55,70325	3,9495	0,280243	-0,18152	Tr
uk107.wav	UK	M	33,14286	54,36648	3,568375	0,232269	-0,24069	Tr
uk108.wav	UK	M	44,26008	52,17874	3,603	0,276276	-0,28244	Tr
uk11.wav	UK	M	54,84284	36,36568	4,45475	0,263336	-0,21109	Tr
uk115.wav	UK	M	33,83906	58,08433	3,581	0,23645	-0,20331	Tr
uk116.wav	UK	M	44,62128	55,34243	3,97525	0,287048	-0,32632	Tr
uk117.wav	UK	M	25,69825	57,39754	3,449625	0,244171	-0,22675	Tr
uk118.wav	UK	M	53,99417	49,42929	3,843875	0,267181	-0,21634	Tr
uk120.wav	UK	M	38,79752	75,80146	5,09225	0,26059	-0,22913	Tr
uk121.wav	UK	M	67,04244	51,36731	4,36075	0,505005	-0,32977	Tr
uk130.wav	UK	M	21,63773	64,67786	3,989	0,205017	-0,14044	Tr
uk132.wav	UK	M	69,56443	45,73701	4,023	0,331573	-0,27414	Tr
uk134.wav	UK	M	54,07307	51,69708	4,062125	0,280457	-0,25546	Tr
uk136.wav	UK	M	39,62246	56,84254	4,257375	0,355682	-0,24463	Tr
uk137.wav	UK	M	38,7014	54,97911	3,7105	0,289093	-0,23532	Tr
uk139.wav	UK	M	56,30331	57,49256	4,45275	0,343323	-0,26608	Tr
uk14.wav	UK	M	28,29439	64,23588	3,7985	0,299011	-0,26041	Tr
uk146.wav	UK	M	31,62206	56,14637	4,666375	0,360168	-0,2139	Tr
uk151.wav	UK	M	48,43603	73,28959	7,258875	0,353149	-0,23004	Tr
uk152.wav	UK	M	48,4363	73,21385	7,21175	0,35321	-0,23004	Tr
uk17.wav	UK	M	33,43745	49,87264	3,7295	0,256866	-0,21234	Tr
uk2.wav	UK	M	39,23439	50,79365	3,9375	0,290741	-0,19434	Tr
uk21.wav	UK	M	62,35976	42,76884	4,48925	0,269409	-0,28482	Tr
uk23.wav	UK	M	42,19474	52,34321	4,69975	0,225281	-0,2439	Tr

uk27.wav	UK	M	77,10948	46,82048	3,331875	0,377899	-0,49152	Tr
uk28.wav	UK	M	62,49137	41,27631	3,6825	0,284454	-0,30679	Tr
uk29.wav	UK	M	60,14252	44,55087	4,48925	0,2966	-0,271	Tr
uk3.wav	UK	M	32,44716	52,87389	4,425625	0,273224	-0,2262	Tr
uk30.wav	UK	M	45,07498	52,13849	3,6825	0,318939	-0,30454	Tr
uk31.wav	UK	M	39,09574	68,91051	6,269	0,258972	-0,3017	Tr
uk36.wav	UK	M	43,91464	67,24707	4,9965	0,385773	-0,20609	Tr
uk37.wav	UK	M	35,5313	70,16079	5,98625	0,262939	-0,20508	Tr
uk38.wav	UK	M	34,60346	60,73527	4,80775	0,241302	-0,1962	Tr
uk39.wav	UK	M	35,48281	57,15214	4,61925	0,291351	-0,17819	Tr
uk4.wav	UK	M	58,23029	57,36236	3,905	0,328156	-0,25476	Tr
uk40.wav	UK	M	61,83384	78,7677	8,8615	0,344757	-0,30661	Tr
uk43.wav	UK	M	34,15045	53,58467	5,225375	0,313049	-0,15009	Tr
uk47.wav	UK	M	24,10911	63,95767	4,346625	0,249542	-0,15707	Tr
uk50.wav	UK	M	33,47272	82,61707	7,722375	0,276215	-0,25461	Tr
uk51.wav	UK	M	40,94309	61,77868	5,892	0,356537	-0,22195	Tr
uk55.wav	UK	M	27,28562	64,89353	4,807875	0,434387	-0,21634	Tr
uk58.wav	UK	M	54,99398	58,28355	4,289375	0,429565	-0,28204	Tr
uk64.wav	UK	M	35,82948	60,29412	4,08	0,256653	-0,28928	Tr
uk65.wav	UK	M	26,27164	58,0953	3,61475	0,221161	-0,18213	Tr
uk68.wav	UK	M	56,57082	50,05163	4,11575	0,307129	-0,27325	Tr
uk69.wav	UK	M	56,74677	53,49573	4,18725	0,35202	-0,23923	Tr
uk72.wav	UK	M	27,67892	58,10845	4,95625	0,204163	-0,23245	Tr
uk76.wav	UK	M	35,3356	62,19322	4,405625	0,327393	-0,24109	Tr
uk78.wav	UK	M	31,41531	64,2496	4,326875	0,258026	-0,22452	Tr
uk79.wav	UK	M	31,4342	57,41837	4,563	0,215179	-0,2294	Tr
uk8.wav	UK	M	42,70298	50,8643	3,77475	0,326294	-0,24625	Tr
uk81.wav	UK	M	104,6897	50,75337	4,72875	0,546143	-0,65964	Tr
uk87.wav	UK	M	65,27339	49,16895	3,782875	0,415405	-0,39182	Tr
uk89.wav	UK	M	79,38692	45,99023	3,479	0,289856	-0,36945	Tr
uk91.wav	UK	M	110,8759	46,49826	3,914125	0,452423	-0,46722	Tr
uk92.wav	UK	M	80,22644	47,98155	3,25125	0,442078	-0,47852	Tr
uk93.wav	UK	M	89,73862	42,59567	3,75625	0,365295	-0,3924	Tr
uk94.wav	UK	M	68,77334	51,94435	3,50375	0,349274	-0,45783	Tr
uk99.wav	UK	M	83,56197	52,41747	3,472125	0,560394	-0,47348	Tr
us1.wav	US	F	50,70848	41,02394	3,022625	0,224915	-0,17963	Tr
us10.wav	US	F	25,25763	62,72091	3,89025	0,247345	-0,17752	Tr
us107.wav	US	F	22,36043	64,31921	3,35825	0,195251	-0,17847	Tr
us112.wav	US	F	19,89811	53,94722	3,225375	0,206177	-0,18051	Tr
us113.wav	US	F	15,3499	58,09456	3,16725	0,177368	-0,14801	Tr

us121.wav	US	M	86,862	63,19411	5,886625	0,529877	-0,30188	Tr
us125.wav	US	M	75,98693	51,03181	4,585375	0,529144	-0,33685	Tr
us126.wav	US	M	83,84756	47,33983	4,64725	0,35672	-0,20252	Tr
us130.wav	US	M	116,4861	41,95945	3,717875	0,560822	-0,36618	Tr
us142.wav	US	M	63,17685	65,52494	3,69325	0,389526	-0,23206	Tr
us15.wav	US	F	31,88276	56,12836	4,09775	0,198578	-0,19858	Tr
us16.wav	US	F	37,77057	54,17607	4,09775	0,214752	-0,16629	Tr
us160.wav	US	M	99,11845	42,71882	3,885875	0,377014	-0,23029	Tr
us168.wav	US	M	231,6552	47,48937	4,085125	0,565918	-0,39145	Tr
us171.wav	US	M	148,7477	52,626	3,534375	0,563049	-0,44367	Tr
us186.wav	US	M	111,6507	44,38887	3,199	0,505646	-0,34076	Tr
us192.wav	US	M	99,23868	56,40473	4,751375	0,373627	-0,27646	Tr
us194.wav	US	M	224,7713	43,22784	4,21025	0,56311	-0,30594	Tr
us195.wav	US	M	390,9798	46,16923	5,5015	0,633636	-0,5419	Tr
us198.wav	US	M	207,9134	53,68472	4,098	0,555115	-0,5224	Tr
us200.wav	US	M	207,1317	52,6125	4,82775	0,562164	-0,4422	Tr
us21.wav	US	F	34,85485	50,8744	3,538125	0,218475	-0,14398	Tr
us211.wav	US	M	110,8976	66,8335	6,643375	0,529602	-0,36902	Tr
us215.wav	US	M	145,8365	49,55188	5,32775	0,549652	-0,42371	Tr
us217.wav	US	M	116,4662	50,03488	5,196375	0,604553	-0,50986	Tr
us22.wav	US	F	31,80043	49,86344	3,56975	0,200623	-0,18918	Tr
us224.wav	US	M	118,3021	52,79897	5,227375	0,56958	-0,52713	Tr
us226.wav	US	M	131,0901	53,84788	5,57125	0,419983	-0,36456	Tr
us229.wav	US	M	140,1529	59,59165	5,57125	0,526276	-0,42172	Tr
us23.wav	US	F	22,05091	51,7031	3,790875	0,174805	-0,11673	Tr
us25.wav	US	F	25,32396	59,54713	3,19075	0,257019	-0,16043	Tr
us26.wav	US	F	24,05673	58,52231	3,75925	0,221588	-0,18018	Tr
us28.wav	US	F	32,32884	53,70076	3,538125	0,236511	-0,19696	Tr
us29.wav	US	F	33,29369	56,46656	3,5065	0,214691	-0,18362	Tr
us35.wav	US	M	47,93615	43,34531	4,66025	0,279388	-0,17041	Tr
us36.wav	US	M	71,99709	43,05918	4,551875	0,434387	-0,19794	Tr
us39.wav	US	M	52,81161	43,70656	4,118375	0,347778	-0,17157	Tr
us4.wav	US	F	39,15323	51,47615	3,3025	0,274109	-0,17346	Tr
us40.wav	US	M	43,09421	76,52709	9,591375	0,331604	-0,14429	Tr
us42.wav	US	M	60,24817	41,88591	4,774875	0,291534	-0,22263	Tr
us46.wav	US	M	46,94479	53,28973	4,128375	0,38208	-0,22452	Tr
us48.wav	US	M	44,05577	50,26274	4,377	0,494446	-0,29965	Tr
us52.wav	US	F	53,27108	54,16964	3,5075	0,280029	-0,22928	Tr
us60.wav	US	F	35,5999	62,99567	4,44475	0,255859	-0,20129	Tr
us68.wav	US	F	32,4572	61,91084	3,036625	0,240784	-0,13882	Tr

us78.wav	US	F	71,65113	54,58448	3,261	0,323242	-0,26706	Tr
us79.wav	US	F	75,02496	47,08447	3,440625	0,329865	-0,20764	Tr
us83.wav	US	F	42,10471	61,91363	3,424125	0,319031	-0,185	Tr
us86.wav	US	F	59,20935	53,42859	3,21925	0,297241	-0,28873	Tr
us89.wav	US	F	72,5762	55,64166	3,306875	0,323944	-0,21033	Tr
us90.wav	US	F	68,48071	66,61881	3,4825	0,331787	-0,32083	Tr
ch12.wav	CH	M	74,89284	40,33874	3,86725	0,431732	-0,32977	Te
ch13.wav	CH	M	74,15994	45,37642	4,3635	0,451202	-0,3519	Te
ch14.wav	CH	M	70,42666	50,35342	5,004625	0,478241	-0,28253	Te
ch15.wav	CH	M	200,7416	36,85963	5,48025	0,524353	-0,46924	Te
ch16.wav	CH	M	156,3375	40,97606	4,344	0,419159	-0,35477	Te
ch18.wav	CH	M	69,52003	37,54667	3,515625	0,472107	-0,36612	Te
ch19.wav	CH	M	95,26791	41,22349	3,78425	0,706146	-0,33078	Te
ch3.wav	CH	M	54,96475	37,82754	3,80675	0,436584	-0,2981	Te
ch4.wav	CH	M	58,45976	44,4564	4,183875	0,43927	-0,23425	Te
ch7.wav	CH	M	41,28452	42,39445	3,585375	0,608795	-0,36359	Te
ch8.wav	CH	M	112,0873	36,80671	4,94475	0,481567	-0,36423	Te
fr1.wav	FR	M	71,35131	59,36812	6,872375	0,345276	-0,24442	Te
fr10.wav	FR	M	123,4002	57,60254	5,0345	0,378937	-0,28256	Te
fr100.wav	FR	M	76,3575	58,505	4,47825	0,55899	-0,34927	Te
fr101.wav	FR	M	93,3838	65,27282	6,128125	0,437286	-0,35062	Te
fr102.wav	FR	M	121,2997	67,04762	5,578125	0,474121	-0,43216	Te
fr103.wav	FR	M	97,9747	64,73316	5,4995	0,478882	-0,341	Te
fr11.wav	FR	M	52,54071	65,79877	4,9545	0,519897	-0,22427	Te
fr12.wav	FR	M	73,66376	65,64608	4,874625	0,551056	-0,22754	Te
fr15.wav	FR	M	93,40567	63,37673	4,9545	0,487579	-0,38519	Te
fr17.wav	FR	M	49,92732	61,35836	4,9545	0,387756	-0,34811	Te
fr19.wav	FR	M	89,72218	55,74517	7,032	0,46048	-0,28775	Te
fr21.wav	FR	M	52,27446	66,00154	5,212	0,580292	-0,22226	Te
fr23.wav	FR	M	129,1084	61,01268	6,949375	0,450562	-0,45685	Te
fr24.wav	FR	M	31,6917	75,60232	4,550125	0,306091	-0,25256	Te
fr25.wav	FR	M	78,48199	65,56476	5,46025	0,422729	-0,29044	Te
fr26.wav	FR	M	76,163	64,33689	5,12925	0,399628	-0,24146	Te
fr29.wav	FR	M	58,94346	67,03143	5,460125	0,471252	-0,35516	Te
fr3.wav	FR	M	57,07008	61,22905	4,475	0,487274	-0,27237	Te
fr31.wav	FR	M	52,2351	66,68924	4,798375	0,363373	-0,23657	Te
fr35.wav	FR	M	69,89082	52,12711	5,29475	0,330597	-0,3309	Te
fr36.wav	FR	M	43,69964	69,98151	6,287375	0,481995	-0,20126	Te
fr39.wav	FR	M	93,86137	58,96147	5,596875	0,565155	-0,2919	Te
fr4.wav	FR	M	98,45093	51,19333	5,274125	0,385376	-0,25366	Te

fr40.wav	FR	M	88,69167	56,16103	5,769125	0,482758	-0,23633	Te
fr45.wav	FR	M	126,5567	61,93798	6,19975	0,634338	-0,42233	Te
fr48.wav	FR	M	52,63299	76,49575	6,457875	0,328125	-0,20053	Te
fr5.wav	FR	M	53,71319	61,22905	4,475	0,408112	-0,25552	Te
fr50.wav	FR	M	39,99456	74,25656	5,683	0,352173	-0,1731	Te
fr51.wav	FR	M	87,73537	67,71708	5,434375	0,415741	-0,28748	Te
fr54.wav	FR	M	74,29228	60,19834	5,01675	0,515564	-0,23511	Te
fr55.wav	FR	M	88,79494	54,09612	4,658375	0,386993	-0,26453	Te
fr57.wav	FR	M	81,64833	66,69229	4,228375	0,410736	-0,28265	Te
fr58.wav	FR	M	103,8628	52,53033	4,873375	0,428802	-0,29684	Te
fr64.wav	FR	M	100,2212	53,88825	4,15675	0,472839	-0,29636	Te
fr65.wav	FR	M	35,1491	61,96234	4,228375	0,378601	-0,25385	Te
fr66.wav	FR	M	81,45133	62,32558	4,3	0,362122	-0,27716	Te
fr68.wav	FR	M	75,66062	64,53488	3,44	0,386719	-0,28857	Te
fr69.wav	FR	M	58,34264	70,17685	6,212875	0,450104	-0,18546	Te
fr71.wav	FR	M	53,3617	66,06753	4,117	0,560577	-0,20071	Te
fr72.wav	FR	M	58,32974	58,50513	4,3415	0,330688	-0,17944	Te
fr73.wav	FR	M	42,78723	59,56911	4,566125	0,345825	-0,20728	Te
fr75.wav	FR	M	93,30391	51,33594	6,662	0,365051	-0,23001	Te
fr76.wav	FR	M	47,04748	68,44489	5,46425	0,33432	-0,22171	Te
fr77.wav	FR	M	46,30174	67,31109	3,892375	0,351868	-0,20947	Te
fr78.wav	FR	M	50,22287	64,3226	4,042125	0,344543	-0,22229	Te
fr79.wav	FR	M	47,32757	65,73542	4,715875	0,343384	-0,26022	Te
fr81.wav	FR	M	50,07539	60,11809	4,19175	0,298157	-0,18765	Te
fr83.wav	FR	M	48,52681	56,25055	4,266625	0,293427	-0,18573	Te
fr84.wav	FR	M	50,11953	63,03979	4,790625	0,302002	-0,17462	Te
fr85.wav	FR	M	39,44446	64,44008	3,8175	0,350677	-0,20715	Te
fr86.wav	FR	M	53,89581	73,41794	8,60825	0,368805	-0,21903	Te
fr87.wav	FR	M	41,65175	68,53674	5,10675	0,361725	-0,17188	Te
fr88.wav	FR	M	79,92579	47,44226	5,18525	0,356812	-0,22076	Te
fr89.wav	FR	M	76,55968	50,42223	4,0855	0,368317	-0,26523	Te
fr9.wav	FR	M	130,8597	51,24672	6,712625	0,5466	-0,46445	Te
fr90.wav	FR	M	55,00076	55,42867	4,871125	0,431335	-0,2435	Te
fr93.wav	FR	M	71,40948	50,91258	4,164	0,407532	-0,20584	Te
fr95.wav	FR	M	144,1446	51,61065	5,73525	0,556549	-0,28253	Te
fr97.wav	FR	M	132,9975	51,31137	5,028125	0,500336	-0,34305	Te
fr98.wav	FR	M	70,66972	66,88504	4,006875	0,537689	-0,51831	Te
fr99.wav	FR	M	82,80718	62,92027	4,164	0,503662	-0,31491	Te
in10.wav	IN	M	34,98603	56,4048	4,113125	0,324097	-0,17465	Te
in12.wav	IN	M	65,25437	54,62644	4,686375	0,400085	-0,27121	Te

in13.wav	IN	M	67,98744	60,25574	4,28175	0,344452	-0,31644	Te
in14.wav	IN	M	59,44473	55,36674	4,04575	0,402466	-0,27197	Te
in15.wav	IN	M	45,59238	57,71832	4,98975	0,390839	-0,23843	Te
in16.wav	IN	M	69,43001	56,17895	5,090875	0,402161	-0,22009	Te
in19.wav	IN	M	59,41614	63,59546	4,528625	0,389435	-0,30621	Te
in20.wav	IN	M	64,31403	55,02418	5,815625	0,365417	-0,24957	Te
in22.wav	IN	M	69,29101	50,34714	5,00525	0,534149	-0,33334	Te
in26.wav	IN	M	70,23449	58,92334	6,483	0,297852	-0,22037	Te
in28.wav	IN	M	50,0372	66,45754	4,3035	0,411987	-0,27451	Te
in3.wav	IN	M	143,2457	44,25008	7,231625	0,507141	-0,3486	Te
in30.wav	IN	M	47,89021	66,24524	3,381375	0,509918	-0,45709	Te
in31.wav	IN	M	51,7178	54,83908	4,3035	0,473206	-0,23169	Te
in32.wav	IN	M	32,29168	67,77364	3,68875	0,472931	-0,21429	Te
in33.wav	IN	M	80,4796	54,48954	4,918375	0,472626	-0,24857	Te
in34.wav	IN	M	45,73484	56,92884	4,6725	0,34787	-0,26685	Te
in35.wav	IN	M	98,95283	59,99314	4,733875	0,5159	-0,3494	Te
in38.wav	IN	M	62,92939	67,33179	5,28725	0,552307	-0,42642	Te
in39.wav	IN	M	34,96431	66,04849	4,057625	0,467468	-0,27264	Te
in40.wav	IN	M	68,10257	60,79027	6,0865	0,36142	-0,26697	Te
in41.wav	IN	M	66,30295	50,8298	4,918375	0,280884	-0,18335	Te
in42.wav	IN	M	97,3503	48,97138	5,717625	0,399017	-0,30341	Te
in44.wav	IN	M	85,40016	53,00263	5,8865	0,471375	-0,38178	Te
in48.wav	IN	M	98,32657	56,88419	6,434125	0,466644	-0,24371	Te
in49.wav	IN	M	72,81561	56,56622	5,338875	0,423035	-0,26273	Te
in5.wav	IN	M	57,11987	53,57495	3,80775	0,501709	-0,24051	Te
in51.wav	IN	M	70,995	49,09085	3,870375	0,404083	-0,26413	Te
in53.wav	IN	M	82,29669	54,82965	5,653875	0,587006	-0,33664	Te
in57.wav	IN	M	50,44091	55,73257	4,62925	0,551849	-0,25131	Te
in58.wav	IN	M	85,35087	51,1909	5,274375	0,451569	-0,27506	Te
in6.wav	IN	M	82,33519	46,37996	5,519625	0,439484	-0,33203	Te
in60.wav	IN	M	68,29709	52,85831	4,729625	0,443146	-0,34305	Te
in66.wav	IN	M	86,09379	49,06226	7,0115	0,375	-0,22614	Te
in67.wav	IN	M	122,9679	50,94126	5,849875	0,378906	-0,26389	Te
in69.wav	IN	M	71,72897	57,33372	5,651125	0,345734	-0,28037	Te
in7.wav	IN	M	47,73633	52,18196	5,135875	0,326752	-0,24179	Te
in70.wav	IN	M	56,60676	63,54286	4,375	0,486389	-0,31879	Te
in72.wav	IN	M	49,24099	63,50436	5,165	0,296356	-0,19858	Te
in75.wav	IN	M	75,52837	50,74228	7,29175	0,285522	-0,19824	Te
in76.wav	IN	M	58,66307	57,09488	5,955	0,30899	-0,24744	Te
in77.wav	IN	M	39,41571	56,92159	5,165	0,272583	-0,21793	Te

in78.wav	IN	M	43,75332	52,73286	5,651125	0,308563	-0,20337	Te
in79.wav	IN	M	46,80832	59,65653	6,805625	0,334625	-0,27084	Te
uk1.wav	UK	M	51,58768	59,47808	5,043875	0,273102	-0,24185	Te
uk10.wav	UK	M	32,0108	51,69628	3,481875	0,241608	-0,17346	Te
uk101.wav	UK	M	132,9603	61,38835	4,9195	0,756531	-0,44583	Te
uk102.wav	UK	M	47,54293	46,40962	3,53375	0,332001	-0,30011	Te
uk105.wav	UK	M	39,9793	46,58818	3,9495	0,243469	-0,19324	Te
uk106.wav	UK	M	40,1035	61,61532	3,603	0,244232	-0,2142	Te
uk109.wav	UK	M	51,49407	46,08324	3,949375	0,302307	-0,22049	Te
uk110.wav	UK	M	35,32397	70,46401	4,711625	0,276123	-0,20273	Te
uk111.wav	UK	M	45,74118	53,85235	3,41675	0,304962	-0,2681	Te
uk112.wav	UK	M	39,90339	50,92559	3,41675	0,258545	-0,21854	Te
uk113.wav	UK	M	29,54857	53,68199	3,613875	0,227356	-0,22604	Te
uk114.wav	UK	M	23,89509	56,64185	3,778125	0,260559	-0,18448	Te
uk119.wav	UK	M	57,32125	53,48126	3,51525	0,384003	-0,2916	Te
uk12.wav	UK	M	33,95714	45,10363	3,902125	0,284058	-0,24774	Te
uk122.wav	UK	M	31,79684	55,65479	3,988875	0,352264	-0,42639	Te
uk123.wav	UK	M	36,49321	60,23934	3,7185	0,294373	-0,25119	Te
uk124.wav	UK	M	42,23347	53,55435	3,92125	0,287476	-0,33084	Te
uk125.wav	UK	M	49,55664	47,84566	3,8875	0,25235	-0,28531	Te
uk126.wav	UK	M	36,99216	54,32298	3,7185	0,269104	-0,2496	Te
uk127.wav	UK	M	46,64327	48,07026	3,786125	0,326324	-0,26517	Te
uk128.wav	UK	M	39,56609	60,80734	3,650875	0,321991	-0,27121	Te
uk129.wav	UK	M	37,12095	59,1645	3,75225	0,2966	-0,23651	Te
uk13.wav	UK	M	41,09856	49,272	4,627375	0,255096	-0,22427	Te
uk131.wav	UK	M	35,4948	71,4664	5,429125	0,439087	-0,28232	Te
uk133.wav	UK	M	20,30801	56,32676	3,124625	0,338379	-0,18808	Te
uk135.wav	UK	M	33,72391	50,71393	4,062	0,241486	-0,15836	Te
uk138.wav	UK	M	51,20879	46,93834	4,687	0,355865	-0,19424	Te
uk140.wav	UK	M	45,68529	78,26308	6,87425	0,381012	-0,2348	Te
uk141.wav	UK	M	40,95208	61,96199	5,939125	0,356567	-0,22195	Te
uk142.wav	UK	M	40,59612	61,24067	4,572125	0,29126	-0,16794	Te
uk143.wav	UK	M	25,67895	61,28863	3,818	0,420898	-0,23502	Te
uk144.wav	UK	M	28,49412	61,87674	4,525125	0,312897	-0,16773	Te
uk145.wav	UK	M	27,29049	64,89353	4,807875	0,434387	-0,2164	Te
uk147.wav	UK	M	53,81276	62,9763	4,477875	0,425232	-0,28955	Te
uk148.wav	UK	M	54,99576	60,29646	4,477875	0,429596	-0,28204	Te
uk149.wav	UK	M	54,31831	58,55521	4,7135	0,36557	-0,26538	Te
uk15.wav	UK	M	28,45815	57,91748	3,384125	0,253143	-0,2327	Te
uk150.wav	UK	M	36,39562	64,89353	4,807875	0,409729	-0,24518	Te

uk16.wav	UK	M	17,18217	50,2327	3,38425	0,181915	-0,19125	Te
uk18.wav	UK	M	31,78776	56,5371	4,351125	0,222443	-0,22263	Te
uk19.wav	UK	M	35,91231	48,80011	3,7295	0,274109	-0,30801	Te
uk20.wav	UK	M	58,75695	45,46685	3,695	0,299652	-0,25034	Te
uk22.wav	UK	M	57,0976	51,41538	4,278875	0,330383	-0,29739	Te
uk24.wav	UK	M	31,1743	51,05227	3,6825	0,262817	-0,20737	Te
uk25.wav	UK	M	49,64716	35,83031	3,963125	0,282318	-0,21124	Te
uk26.wav	UK	M	35,64841	42,61865	3,331875	0,222595	-0,18454	Te
uk32.wav	UK	M	24,63475	62,74333	4,43075	0,273651	-0,18594	Te
uk33.wav	UK	M	31,37169	61,28763	5,09075	0,253693	-0,1864	Te
uk34.wav	UK	M	30,05041	61,71649	4,6665	0,487091	-0,24558	Te
uk35.wav	UK	M	30,41912	65,95599	4,76075	0,267212	-0,17304	Te
uk41.wav	UK	M	34,99807	68,88128	6,242625	0,289215	-0,18839	Te
uk42.wav	UK	M	18,2045	67,37992	4,393	0,27771	-0,21188	Te
uk44.wav	UK	M	26,29139	62,02025	4,901625	0,37326	-0,15414	Te
uk45.wav	UK	M	33,75441	65,95783	5,549	0,278198	-0,18045	Te
uk46.wav	UK	M	33,07513	65,30093	4,716625	0,288879	-0,15814	Te
uk48.wav	UK	M	42,17553	60,11179	5,4565	0,281097	-0,21671	Te
uk49.wav	UK	M	46,36074	57,94819	4,762875	0,416138	-0,18958	Te
uk5.wav	UK	M	26,21102	59,09642	4,230375	0,251617	-0,18988	Te
uk52.wav	UK	M	40,61601	63,64497	4,619375	0,291321	-0,16788	Te
uk53.wav	UK	M	25,68495	63,89075	4,10075	0,420898	-0,23505	Te
uk54.wav	UK	M	28,49311	61,8388	4,430875	0,312836	-0,16773	Te
uk56.wav	UK	M	31,62437	57,8218	4,807875	0,360168	-0,21393	Te
uk57.wav	UK	M	53,8161	64,7166	4,6665	0,425262	-0,28949	Te
uk59.wav	UK	M	54,31587	57,58511	4,61925	0,36557	-0,26529	Te
uk6.wav	UK	M	54,34031	55,16908	4,13275	0,333771	-0,34699	Te
uk60.wav	UK	M	36,39429	63,86113	4,666375	0,40979	-0,24518	Te
uk61.wav	UK	M	47,26692	52,70332	4,402	0,284973	-0,22882	Te
uk62.wav	UK	M	35,08482	53,97176	4,187375	0,287445	-0,22586	Te
uk63.wav	UK	M	42,39272	55,88289	4,11575	0,245911	-0,19971	Te
uk66.wav	UK	M	45,06257	52,06281	4,18725	0,354736	-0,41312	Te
uk67.wav	UK	M	40,0527	56,37808	4,044125	0,331665	-0,37189	Te
uk7.wav	UK	M	45,67238	51,88917	3,97	0,409058	-0,24307	Te
uk70.wav	UK	M	52,26675	56,86275	4,08	0,27417	-0,28012	Te
uk71.wav	UK	M	35,5595	52,17055	4,523625	0,237701	-0,17264	Te
uk73.wav	UK	M	27,0247	66,00881	4,48425	0,269257	-0,27756	Te
uk74.wav	UK	M	40,54412	54,61072	4,24825	0,249329	-0,24805	Te
uk75.wav	UK	M	38,24448	55,16249	4,1695	0,269287	-0,26559	Te
uk77.wav	UK	M	52,448	59,59575	4,799	0,415894	-0,33759	Te

uk80.wav	UK	M	26,90386	64,94339	4,681	0,217468	-0,17053	Te
uk82.wav	UK	M	82,89557	35,29998	3,51275	0,411041	-0,33621	Te
uk83.wav	UK	M	54,03211	42,54096	3,479	0,322021	-0,37067	Te
uk84.wav	UK	M	70,5362	42,51147	3,951875	0,298676	-0,27252	Te
uk85.wav	UK	M	81,22467	38,61749	3,88425	0,345795	-0,33765	Te
uk86.wav	UK	M	64,02903	43,42309	3,5465	0,301392	-0,32367	Te
uk88.wav	UK	M	85,12666	47,786	3,8505	0,411102	-0,32184	Te
uk9.wav	UK	M	42,77921	43,41697	3,547	0,284943	-0,22797	Te
uk90.wav	UK	M	54,68989	55,87059	4,18825	0,325562	-0,29474	Te
uk95.wav	UK	M	54,03956	51,72344	3,093375	0,351837	-0,31491	Te
uk96.wav	UK	M	52,80732	52,27378	3,78775	0,333038	-0,28128	Te
uk97.wav	UK	M	58,32972	51,5193	3,377375	0,316803	-0,28317	Te
uk98.wav	UK	M	78,78329	55,43966	3,535375	0,436279	-0,39084	Te
us100.wav	US	F	45,65146	66,03745	4,118875	0,29364	-0,24851	Te
us101.wav	US	F	28,79279	62,83435	4,042375	0,221802	-0,17191	Te
us102.wav	US	F	27,13518	48,72701	3,078375	0,229797	-0,15781	Te
us103.wav	US	F	27,95587	61,74428	3,887	0,231628	-0,19531	Te
us104.wav	US	F	17,85225	60,53441	3,17175	0,174438	-0,14859	Te
us105.wav	US	F	22,24259	61,25574	3,265	0,203033	-0,12378	Te
us106.wav	US	F	37,40144	54,54078	3,887	0,240936	-0,19598	Te
us108.wav	US	F	43,73202	50,86976	3,4205	0,234894	-0,16815	Te
us109.wav	US	F	35,32844	51,92506	3,3895	0,237579	-0,1889	Te
us11.wav	US	F	29,27374	70,12258	5,476125	0,21228	-0,12515	Te
us110.wav	US	F	30,20246	69,06599	4,63325	0,241821	-0,1806	Te
us111.wav	US	F	34,57191	53,79026	3,457875	0,208618	-0,1983	Te
us114.wav	US	F	27,25424	50,3937	3,254375	0,229645	-0,17984	Te
us115.wav	US	F	29,21432	52,85193	3,254375	0,243103	-0,15665	Te
us116.wav	US	F	21,50634	61,25188	3,16725	0,191345	-0,14819	Te
us117.wav	US	F	31,48536	55,99708	3,42875	0,232697	-0,20337	Te
us118.wav	US	F	20,18262	55,32103	3,109125	0,195587	-0,20758	Te
us119.wav	US	F	29,39684	54,54324	3,080125	0,245697	-0,14603	Te
us12.wav	US	F	38,40838	52,237	4,135	0,227325	-0,20334	Te
us120.wav	US	F	24,23319	69,29427	4,3005	0,21936	-0,15024	Te
us122.wav	US	M	70,59221	52,33889	4,5855	0,40744	-0,25726	Te
us123.wav	US	M	79,8393	40,34531	4,213625	0,321594	-0,28046	Te
us124.wav	US	M	67,65177	39,29365	4,2755	0,366394	-0,17539	Te
us127.wav	US	M	84,15179	44,61717	4,213625	0,411987	-0,28745	Te
us128.wav	US	M	61,8568	49,72195	4,5855	0,419891	-0,2222	Te
us129.wav	US	M	56,10648	63,39829	3,470125	0,483826	-0,24765	Te
us13.wav	US	F	29,14852	56,19669	3,986	0,210571	-0,16989	Te

us131.wav	US	M	87,36118	49,68812	4,709375	0,486237	-0,24991	Te
us132.wav	US	M	82,90759	53,79336	5,391	0,498962	-0,22138	Te
us133.wav	US	M	141,7303	48,57878	4,322875	0,538635	-0,30112	Te
us134.wav	US	M	108,824	43,4497	2,853875	0,548431	-0,29837	Te
us135.wav	US	M	60,15572	39,41756	3,3995	0,37384	-0,22113	Te
us136.wav	US	M	53,4485	43,44906	4,280875	0,314453	-0,18231	Te
us137.wav	US	M	83,05897	44,84933	2,854	0,589844	-0,50827	Te
us138.wav	US	M	49,43211	45,97218	3,567375	0,423309	-0,15945	Te
us139.wav	US	M	57,2832	44,02719	3,86125	0,39682	-0,19901	Te
us14.wav	US	F	37,85716	49,74757	4,0605	0,232269	-0,18756	Te
us140.wav	US	M	75,28497	41,52045	4,238875	0,339722	-0,21539	Te
us141.wav	US	M	68,42317	43,37699	3,273625	0,391388	-0,18893	Te
us143.wav	US	M	146,7111	48,93495	3,474	0,57019	-0,43961	Te
us144.wav	US	M	189,1873	37,02238	4,105625	0,541016	-0,46079	Te
us145.wav	US	M	98,21802	40,49663	3,654625	0,532593	-0,33548	Te
us146.wav	US	M	124,4147	38,72365	3,92525	0,480469	-0,23752	Te
us147.wav	US	M	98,53219	50,8293	3,383875	0,521881	-0,33267	Te
us148.wav	US	M	109,7978	39,74269	3,92525	0,42038	-0,2731	Te
us149.wav	US	M	146,1557	44,92058	3,38375	0,556122	-0,42133	Te
us150.wav	US	M	114,7673	40,71017	4,4215	0,476837	-0,24771	Te
us151.wav	US	M	187,9578	56,46126	4,286125	0,542847	-0,47205	Te
us152.wav	US	M	138,4069	63,15893	5,09825	0,546326	-0,43256	Te
us153.wav	US	M	146,8397	69,10142	7,612	0,534302	-0,43167	Te
us154.wav	US	M	173,4451	58,56893	5,4295	0,514923	-0,32434	Te
us155.wav	US	M	118,5736	50,40236	4,364875	0,508362	-0,32898	Te
us156.wav	US	M	54,33846	56,10243	3,88575	0,376129	-0,20981	Te
us157.wav	US	M	95,21767	43,23351	3,885875	0,466614	-0,29727	Te
us158.wav	US	M	132,5109	37,57261	4,0455	0,477722	-0,29886	Te
us159.wav	US	M	157,8311	50,40236	4,364875	0,576843	-0,37247	Te
us161.wav	US	M	59,98502	64,47498	4,31175	0,560883	-0,21347	Te
us162.wav	US	M	64,14085	67,38476	4,89725	0,50296	-0,35532	Te
us163.wav	US	M	155,7592	45,38878	4,406375	0,532379	-0,31952	Te
us164.wav	US	M	102,1093	35,70618	3,30475	0,434814	-0,30145	Te
us165.wav	US	M	141,8064	37,98617	3,58025	0,516418	-0,33456	Te
us166.wav	US	M	106,6133	53,00691	4,45225	0,543671	-0,27664	Te
us167.wav	US	M	125,6643	63,0414	4,314625	0,588684	-0,51096	Te
us169.wav	US	M	133,4761	47,48937	4,085125	0,533478	-0,30637	Te
us17.wav	US	F	40,54838	64,14063	4,2095	0,262054	-0,26737	Te
us170.wav	US	M	187,6019	52,17329	3,488375	0,517334	-0,357	Te
us172.wav	US	M	186,0739	58,0976	5,094875	0,548798	-0,41055	Te

us173.wav	US	M	164,9502	57,32835	4,08175	0,544739	-0,37106	Te
us174.wav	US	M	116,0002	53,40683	4,081875	0,498413	-0,33841	Te
us175.wav	US	M	287,9097	40,39026	3,66425	0,588074	-0,42954	Te
us176.wav	US	M	154,9525	46,23527	3,849875	0,580627	-0,3125	Te
us177.wav	US	M	174,7576	47,90993	4,1745	0,592438	-0,40158	Te
us178.wav	US	M	188,8002	39,96319	3,8035	0,505951	-0,31821	Te
us179.wav	US	M	93,07834	48,99709	4,081875	0,361694	-0,2395	Te
us18.wav	US	F	44,6169	55,15222	4,09775	0,289978	-0,26559	Te
us180.wav	US	M	114,6084	48,03002	3,664375	0,406616	-0,28339	Te
us181.wav	US	M	87,26475	51,74156	3,71075	0,554871	-0,33606	Te
us182.wav	US	M	150,1151	58,46484	5,473375	0,517548	-0,27502	Te
us183.wav	US	M	219,6237	39,20642	4,234	0,490906	-0,34323	Te
us184.wav	US	M	126,2548	41,09589	4,234	0,5	-0,36639	Te
us185.wav	US	M	105,3818	43,01344	3,99875	0,389221	-0,24738	Te
us187.wav	US	M	103,485	53,60444	4,328	0,474426	-0,24497	Te
us188.wav	US	M	162,6347	49,86059	3,810625	0,456543	-0,31186	Te
us189.wav	US	M	187,3867	46,48032	3,528375	0,500885	-0,3349	Te
us19.wav	US	F	51,68017	61,12957	3,7625	0,261841	-0,22983	Te
us190.wav	US	M	241,1978	44,49113	4,04575	0,52063	-0,41559	Te
us191.wav	US	M	183,188	42,05054	4,328125	0,583405	-0,3956	Te
us193.wav	US	M	434,4335	39,58436	4,042	0,594269	-0,54935	Te
us196.wav	US	M	223,6332	45,74193	4,54725	0,590546	-0,48578	Te
us197.wav	US	M	247,4118	46,79295	3,76125	0,624573	-0,42618	Te
us199.wav	US	M	276,4884	43,17074	4,77175	0,661285	-0,49442	Te
us2.wav	US	F	29,90717	53,2811	3,190625	0,26358	-0,17987	Te
us20.wav	US	F	43,64994	62,36126	4,84275	0,294006	-0,23593	Te
us201.wav	US	M	227,5363	56,04736	4,60325	0,641144	-0,44183	Te
us202.wav	US	M	149,545	76,03855	7,5225	0,540039	-0,37915	Te
us203.wav	US	M	100,8593	41,62192	5,237625	0,461731	-0,31345	Te
us204.wav	US	M	101,8196	37,18155	4,357	0,518677	-0,41562	Te
us205.wav	US	M	109,2588	35,38296	4,635	0,517212	-0,36472	Te
us206.wav	US	M	52,21526	48,15133	3,19825	0,49884	-0,33405	Te
us207.wav	US	M	77,23323	44,80863	3,615375	0,533264	-0,35455	Te
us208.wav	US	M	170,7776	36,44891	7,4625	0,565796	-0,46954	Te
us209.wav	US	M	162,3178	44,2701	7,138	0,52475	-0,37836	Te
us210.wav	US	M	58,49141	52,90869	3,8935	0,452118	-0,35822	Te
us212.wav	US	M	113,015	53,94581	6,11725	0,501648	-0,35523	Te
us213.wav	US	M	123,0617	54,59218	5,788375	0,453186	-0,36789	Te
us214.wav	US	M	155,603	53,29995	5,591	0,41214	-0,384	Te
us216.wav	US	M	157,3779	54,65348	5,196375	0,568604	-0,43344	Te

us218.wav	US	M	142,863	53,55678	5,78825	0,482971	-0,32541	Te
us219.wav	US	M	151,615	55,38336	5,525125	0,603882	-0,52185	Te
us220.wav	US	M	151,6022	63,15097	6,84075	0,522034	-0,47598	Te
us221.wav	US	M	146,3852	51,90237	6,396625	0,548828	-0,40198	Te
us222.wav	US	M	147,3457	54,56463	5,571375	0,546661	-0,47134	Te
us223.wav	US	M	139,7052	53,37014	5,43375	0,500061	-0,37961	Te
us225.wav	US	M	95,64663	54,00143	5,777625	0,326721	-0,31628	Te
us227.wav	US	M	110,5877	55,47585	5,2275	0,476471	-0,38611	Te
us228.wav	US	M	124,536	53,47594	5,984	0,410217	-0,42136	Te
us230.wav	US	M	171,3109	74,48623	10,04213	0,567505	-0,43594	Te
us231.wav	US	M	119,3166	41,53362	5,441375	0,48526	-0,3299	Te
us232.wav	US	M	144,6185	50,1632	5,821	0,433655	-0,35367	Te
us233.wav	US	M	101,1335	50,77657	5,947625	0,468628	-0,28839	Te
us234.wav	US	M	107,3923	52,80172	5,568	0,41687	-0,30368	Te
us235.wav	US	M	141,1795	46,11165	5,37825	0,367828	-0,40497	Te
us236.wav	US	M	101,1742	45,70014	5,251625	0,377563	-0,31302	Te
us237.wav	US	M	82,71554	55,51897	4,93525	0,562042	-0,37741	Te
us238.wav	US	M	112,0841	46,82927	5,125	0,527802	-0,39807	Te
us239.wav	US	M	160,7121	42,14551	5,884375	0,561401	-0,40152	Te
us24.wav	US	F	31,41023	53,43595	3,443375	0,245026	-0,20651	Te
us240.wav	US	M	103,4323	59,10941	6,32725	0,463959	-0,39102	Te
us27.wav	US	F	33,12391	54,19	3,727625	0,2034	-0,1521	Te
us3.wav	US	F	33,99411	56,38189	3,050625	0,276367	-0,17566	Te
us30.wav	US	F	46,09403	59,97858	4,2015	0,290344	-0,21866	Te
us31.wav	US	M	84,93501	60,35995	5,202125	0,541504	-0,39667	Te
us32.wav	US	M	86,90824	41,21333	3,251375	0,535156	-0,33218	Te
us33.wav	US	M	76,72242	45,29726	3,576375	0,501221	-0,26755	Te
us34.wav	US	M	64,79652	48,29982	4,38925	0,413177	-0,20642	Te
us37.wav	US	M	46,47211	52,04032	5,418875	0,4422	-0,38333	Te
us38.wav	US	M	36,12691	49,74592	3,739	0,296082	-0,22455	Te
us41.wav	US	M	102,4746	48,96862	5,023625	0,539063	-0,39124	Te
us43.wav	US	M	65,91407	52,22717	4,32725	0,462677	-0,31558	Te
us44.wav	US	M	66,99941	39,65234	3,581125	0,48642	-0,2381	Te
us45.wav	US	M	41,10442	46,8253	3,9295	0,405243	-0,2113	Te
us47.wav	US	M	50,38104	55,84642	3,58125	0,398834	-0,22672	Te
us49.wav	US	M	29,22301	52,93294	3,9295	0,299286	-0,15707	Te
us5.wav	US	F	46,13606	59,65991	3,050625	0,296021	-0,24411	Te
us50.wav	US	M	45,31127	59,23798	5,57075	0,330811	-0,18405	Te
us51.wav	US	F	94,89723	52,9153	3,779625	0,292694	-0,22549	Te
us53.wav	US	F	47,55398	49,75917	3,295875	0,260895	-0,18671	Te

us54.wav	US	F	68,34537	43,50093	3,35625	0,27774	-0,18277	Te
us55.wav	US	F	56,32258	54,92396	3,3865	0,328125	-0,23346	Te
us56.wav	US	F	47,54441	63,19209	3,3865	0,302155	-0,2077	Te
us57.wav	US	F	32,00118	64,37325	3,3865	0,230103	-0,2114	Te
us58.wav	US	F	29,55814	58,46744	3,3865	0,243805	-0,16827	Te
us59.wav	US	F	39,35915	64,84822	3,084125	0,277496	-0,24576	Te
us6.wav	US	F	39,56773	59,43983	2,994625	0,277832	-0,2001	Te
us61.wav	US	F	62,3903	53,37967	3,596875	0,274872	-0,23901	Te
us62.wav	US	F	35,31745	57,0944	3,538	0,226685	-0,18369	Te
us63.wav	US	F	70,23261	46,17358	3,5085	0,320038	-0,26776	Te
us64.wav	US	F	31,44175	62,71498	3,12525	0,266968	-0,18137	Te
us65.wav	US	F	65,52845	57,09009	2,97775	0,312164	-0,23236	Te
us66.wav	US	F	52,5281	58,30414	3,5675	0,28833	-0,17349	Te
us67.wav	US	F	58,74195	62,23502	3,213625	0,3302	-0,15158	Te
us69.wav	US	F	34,60523	72,7245	3,272625	0,281891	-0,19406	Te
us7.wav	US	F	37,66256	57,66895	3,190625	0,233582	-0,22138	Te
us70.wav	US	F	54,0202	68,75865	4,334	0,407928	-0,26068	Te
us71.wav	US	F	64,67365	56,90608	3,62	0,34259	-0,18964	Te
us72.wav	US	F	80,77422	51,9337	3,62	0,61142	-0,2887	Te
us73.wav	US	F	48,77426	59,42249	4,038875	0,286285	-0,17178	Te
us74.wav	US	F	65,25642	62,76711	3,91925	0,334351	-0,20517	Te
us75.wav	US	F	88,33304	50,77937	3,1115	0,417755	-0,25723	Te
us76.wav	US	F	55,1062	59,35182	3,20125	0,302734	-0,30658	Te
us77.wav	US	F	41,55741	55,82357	2,902	0,300903	-0,17172	Te
us8.wav	US	F	29,77636	59,85647	3,2745	0,233643	-0,15839	Te
us80.wav	US	F	45,5825	62,03184	3,320875	0,314972	-0,19968	Te
us81.wav	US	F	65,64915	57,1276	3,746	0,289429	-0,20947	Te
us82.wav	US	F	59,24898	60,87581	3,4825	0,323425	-0,21164	Te
us84.wav	US	F	40,04216	65,87909	3,248375	0,294922	-0,2406	Te
us85.wav	US	F	64,60149	61,89306	3,102125	0,378784	-0,23908	Te
us87.wav	US	F	53,40512	64,48279	3,628875	0,30838	-0,23489	Te
us88.wav	US	F	67,73919	66,00467	3,424	0,346649	-0,22672	Te
us9.wav	US	F	59,39225	55,01976	3,1625	0,259094	-0,20129	Te
us91.wav	US	F	37,74297	59,76929	4,18275	0,211945	-0,19638	Te
us92.wav	US	F	33,91553	52,5927	3,3845	0,283081	-0,21832	Te
us93.wav	US	F	35,04386	51,64835	3,64	0,278015	-0,17471	Te
us94.wav	US	F	41,21808	50,44346	3,608	0,248901	-0,21536	Te
us95.wav	US	F	47,60129	47,83429	3,512125	0,230591	-0,18387	Te
us96.wav	US	F	28,59057	60,83192	3,320625	0,208466	-0,19901	Te
us97.wav	US	F	23,68336	63,15721	3,863375	0,243561	-0,12201	Te

us98.wav	US	F	41,39409	53,00318	3,735625	0,284698	-0,17535	Te
us99.wav	US	F	52,83886	47,9904	3,959125	0,256317	-0,21124	Te