

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE ENGENHARIA
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA ELÉTRICA

FLÁVIO LUIS WISNEVSKI.

**CODIFICADOR G729a ORIENTADO À AVALIAÇÃO DA QUALIDADE
PERCEPTUAL DO SINAL DE VOZ**

Porto Alegre
Agosto - 2011

FLÁVIO LUIS WISNEVSKI

**CODIFICADOR G729a ORIENTADO À AVALIAÇÃO DA QUALIDADE
PERCEPTUAL DO SINAL DE VOZ**

**DISSERTAÇÃO APRESENTADA COMO
REQUISITO PARCIAL PARA OBTENÇÃO
DO GRAU DE MESTRE, PELO
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA DA
PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO GRANDE DO SUL.**

Orientador: Prof. Dr. Rubem Dutra Ribeiro Fagundes

Porto Alegre
Agosto - 2011

**CODIFICADOR G729a ORIENTADO À AVALIAÇÃO DA QUALIDADE
PERCEPTUAL DO SINAL DE VOZ**

FLÁVIO LUIS WISNEVSKI

**DISSERTAÇÃO APRESENTADA COMO
REQUISITO PARCIAL PARA OBTENÇÃO
DO GRAU DE MESTRE, PELO
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA DA
PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO GRANDE DO SUL.**

Porto Alegre, 25 de Agosto de 2011.

Prof. Dr. Rubem Dutra Ribeiro Fagundes
Orientador

Prof. Dra. Letícia Maria Bolzani Pöhls
Coordenadora do Programa de Pós-Graduação em Engenharia Elétrica PUCRS

Banca Examinadora:

Prof. Dr. Miguel Arjona Ramírez – EPUSP/USP

Prof. Dra. Letícia Maria Bolzani Pöhls – PUCRS

Prof. Dr. Rubem Dutra Ribeiro Fagundes – PUCRS

À minha esposa Lisiane,
meu amor e meu carinho.

AGRADECIMENTOS

Agradeço à Deus pelas oportunidades de vida que tem me proporcionado.

À minha amada Lisiane, por estar sempre comigo, confiando na nossa trajetória em construção de uma família.

Agradeço, também, à minha mãe, por me lembrar como a vida é difícil quando tomamos decisões erradas.

Ao professor Rubem, pela sua paciência em ensinar sem criticar, e pela sua sabedoria em me mostrar que sempre existe uma saída no final.

Aos amigos Lúcio e Lucas pelo seu auxílio durante a pesquisa e desenvolvimento deste trabalho.

Ao professor João Ernandes, pelos ensinamentos de vida acadêmica e profissional.

Enfim, meu obrigado a todos os que estiveram junto comigo nesta jornada.

“Um professor afeta a eternidade; ele nunca sabe onde a sua influência termina.”

Henry Adans

RESUMO

Esta dissertação apresenta um modelo de codificador da voz que avalia a qualidade perceptual, utilizando-se um codificador paramétrico definido pela Recomendação do ITU-T, o G729a, conhecido tecnicamente como *Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP)*.

Atualmente, a codificação de voz é avaliada por parâmetros subjetivos, em que não há uma análise para melhoria perceptual que altere os parâmetros dinamicamente. O objetivo deste trabalho é a melhoria no desempenho do sinal codificado na fonte, considerando as limitações de trabalhar com as variáveis escalares do codec.

Foi realizada uma análise sobre os sinais de voz para dar subsídios ao leitor, possibilitando o entendimento teórico e uma descrição sobre o codificador G729a, enfatizando os estágios de codificação dos parâmetros escalares, objeto deste estudo.

Programou-se no codificador G729a, o algoritmo PESQ, o qual realiza a análise perceptual do sinal de voz. Também foi avaliada recomendação P.862 do ITU-T que descreve o algoritmo de avaliação da qualidade perceptual.

Para validar a metodologia proposta, foram realizados experimentos em sinais de voz do banco de dados Timit, em que se estudou o comportamento do sinal de voz e a melhoria perceptual devido à alteração dos parâmetros escalares de codificação do sinal.

No experimento realizado, foi realizada a modificação no codificador CS-ACELP com a utilização de uma análise perceptual ponderando a avaliação do algoritmo do PESQ para decisão de atuação sobre a codificação, o que resultou em sinais de maior qualidade para os usuários, garantindo a estabilidade de 72,41% e melhorando 50,38% dos quadros avaliados pelo PESQ. Houve alteração no ganho de pitch, sem a modificação da essência do codec.

Palavras-chaves: Codificação de voz, predição linear, análise perceptual, G729a, PESQ.

ABSTRACT

This thesis presents a model that evaluates the perceptual quality of the encoding voice, using a parametric encoder defined by the ITU-T Recommendation G729a, technically known as Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP).

Currently the voice coding is evaluated by subjective parameters, where there is a perceptual analysis to improve the parameters that change dynamically. The objective of this study is improvement in the performance of the coded signal source, considering the limitations of working with the codec scalar variables.

An analysis of the speech signals to improve input to the reader, allowing for better understanding and a theoretical description of the encoder G729a, emphasizing the stages of encoding scalar parameters, object of this study.

The PESQ algorithm has been inserted into the coder, which by these means performs a perceptually based analysis of the speech signal. In so doing, we have also analyzed thoroughly ITU-T Recommendation P.862 describing an algorithm for objectively evaluating perceptual speech quality.

To validate the proposed methodology, tests were performed for comparison with speech signals from the Timit database for studying the improvement in coded speech quality after the perceptual improvement algorithm applied to the scalar parameters in coding.

In the experiment, was performed in modified CS-ACELP coder using a perceptual analysis considering the evaluation of the PESQ algorithm for decision to operate on the encoding, resulting in higher quality signals to users, ensuring the stability of 72.41% and 50.38% of the frames better evaluated by PESQ. There was change in the pitch gain, without changing the essence of the codec.

Key words: Speech coding, linear prediction, perceptual analysis, G729a, PESQ.

SUMÁRIO

1. Introdução	15
1.1. Objetivo	15
1.2. Motivação	16
1.3. Estrutura da Dissertação	16
2. Análise do Aparelho Fonador.....	18
2.1. Processo Fisiológico de Produção de Voz.....	18
2.1.1. Trato Vocal.....	19
2.1.2. Sistema de produção de voz e as suas formantes	20
2.2. Modelo de Produção do Sinal de Voz	23
2.2.1. Gerador de excitação	23
2.2.2. Modelagem do Trato vocal.....	24
2.2.3. Radiação	25
2.2.4. Modelo completo de produção de voz.....	26
3. Codificação de voz em DSP, Processamento Digital de Sinais	28
3.1. Sinais e comunicação.....	28
3.1.1. Classificação do sinal para formação de modelos	29
3.1.2. Amostragem do sinal	29
3.1.3. Quantização do sinal.....	29
3.1.4. Codificação do sinal	29
3.1.5. Análise Homomórfica ou Análise Cepstral	30
3.2. Tipos de codificadores.....	31
3.2.1. Codificadores de forma de onda.....	31
3.2.2. Codificadores paramétricos	32
3.2.3. Codificadores híbridos.....	32
3.3. Codificação Preditiva Linear	33
3.3.1. Análise da Codificação Preditiva Linear	33
3.3.2. <i>Line Spectrum Frequency</i> e <i>Line Spectrum Pairs</i>	36
3.3.3. Janelamento	37
3.4. Quantização Vetorial	39
3.4.1. Entendendo quantização vetorial.....	39
4. Análise da qualidade do sinal de voz	42
4.1. Medidas subjetivas de qualidade	43
4.1.1. <i>Mean Opinion Score</i> (MOS).....	43
4.2. Medidas objetivas de qualidade.....	44
4.2.1. <i>Perceptual Speech Quality Measure</i> (PSQM).....	45
4.2.2. <i>Perceptual Analysis Measurement System</i> (PAMS)	46
4.2.3. <i>Perceptual Evaluation of Speech Quality</i> (PESQ)	46
5. Codificadores ITU-T	48
5.1. Codificadores ITU-T utilizados em telefonia.....	48
5.1.1. Recomendação ITU-T G.711	49
5.1.2. Recomendação ITU-T G.726	50
5.1.3. Recomendação ITU-T G.728	51
5.1.4. Recomendação ITU-T G.723.1	52
5.1.5. Recomendação ITU-T G.729	53
5.2. CS-ACELP: Conjugate Structure Code-Excited Linear Prediction (G.729a).....	54

5.2.1.	O codificador CS-ACELP - (G.729a).....	54
5.2.2.	Codificador	55
5.2.3.	Decodificador	57
6.	Proposta.....	58
7.	Estudo de Caso.....	62
7.1.	Plataforma de experimento (G729a com PESQ).....	62
7.2.	Análise dos parâmetros de avaliação.....	63
7.2.1.	Limites de frames do PESQ no bloco do codec	63
7.2.2.	Formatação dos arquivos de testes	64
7.2.3.	Sinais analisados para alteração da codificação	67
7.2.4.	Ajuste da avaliação Perceptual (alterações significativas).....	68
7.3.	Padrão de avaliação perceptual proposto.....	73
8.	Resultados Obtidos.....	75
8.1.	Testes realizados.....	75
8.1.1.	Análise dos parâmetros escalares	75
8.1.2.	Avaliação Perceptual – Oradores masculino (Arquivo 01).....	77
8.1.3.	Avaliação Perceptual – Oradores feminino (Arquivo 01).....	80
8.1.4.	Avaliação Perceptual – Oradores masculino (Arquivo 02).....	83
8.1.5.	Avaliação Perceptual – Oradores feminino (Arquivo 02).....	86
9.	Conclusões.....	89
9.1.	Sugestões para trabalhos futuros	91
10.	Referências Bibliográficas	93
11.	Anexo	98
11.1.	Arquivos de testes do Banco de dados Timit	98
12.	Apêndices.....	100
12.1.	Apêndice A - Resultado do Arquivo1 de Oradores masculinos.....	100
12.2.	Apêndice B - Resultado do Arquivo1 de Oradores femininos	105
12.3.	Apêndice C - Resultado do Arquivo2 de Oradores masculinos	111
12.4.	Apêndice D - Resultado do Arquivo2 de Oradores femininos.....	119

LISTA DE FIGURAS

Figura 2-1 – Órgãos de produção da fala – Sistema do Trato Vocal.	19
Figura 2-2 – Segmento sonoro de voz com segmentos surdos de voz – sinal irradiado	20
Figura 2-3 – Ptch do sinal de voz amostrado.....	21
Figura 2-4 – Exemplo de historiograma da frequência fundamental	22
Figura 2-5 –Diagrama em blocos do modelo para produção de voz.	23
Figura 2-6 – Diagrama em blocos do sinal de excitação para sons sonoros	23
Figura 2-7 – (a) Exemplo de dois tubos e três tubos.	24
Figura 2-8 – Modelo completo para produção de voz.....	26
Figura 2-9 – Modelo simplificado para produção de voz.....	27
Figura 3-1 - Elementos de um sistema de comunicação.	28
Figura 3-2 - Classificação do sinal de voz em sons sonoros e surdos.....	29
Figura 3-3 – Partes básica de um conversor analógico-digital (A/D)	30
Figura 3-4 – AnáliseCepstral do sinal de voz.....	31
Figura 3-5 – Processo de geração de voz humana num codificador LPC	32
Figura 3-6 – Qualidade x Taxa de transmissão dos codificadores	33
Figura 3-7 a) Modelo de produção da fala b) Modelo do trato vocal.....	34
Figura 3-8 – Diagrama de blocos do processo de análise (a) e de síntese (b).....	36
Figura 3-9 – Análise das janelas de Hanning e Hamming.	38
Figura 3-10 – Processo de janelamento	38
Figura 3-11 Exemplo de quantização em 1 dimensão.....	39
Figura 3-12 – Exemplo de quantização em 2 dimensões	40
Figura 3-13 – Exemplo de espaço vetorial com centróides de vetores de códigos	41
Figura 4-1 – Comportamento da avaliação MOS para dos diferentes tipos de codificadores..	44
Figura 4-2 – Modelo de avaliação objetiva, utilizando o PSQM.	45
Figura 5-1 – Sistema de codificação de voz.	48
Figura 5-2 – Diagrama de blocos da técnica PCM.....	49
Figura 5-3 – Diagrama de blocos do processo de codificação ADPCM (VARY e MARTIN, 2006).....	50
Figura 5-4 – Diagrama de blocos do processo de codificação LD-CELP (FURUI, 2001)	51
Figura 5-5 – Diagrama de blocos do processo de codificação LD-CELP	52
Figura 5-6 – Diagrama de blocos do processo de codificação CS-ACELP	55
Figura 5-7 – Diagrama de blocos do processo de decodificação CS-ACELP	57

Figura 6-1 – Relação do MOS x Taxa de Transferência dos diversos codificadores.....	58
Figura 6-2 - Diagrama de avaliação perceptual do codec no sinal de origem.....	59
Figura 6-3 - Diagrama de blocos simplificado de análise do codec	61
Figura 7-1 Arquivo01 Oradores masculinos – Sinal de voz.....	64
Figura 7-2 Arquivo01 Oradores masculinos – Análise de frequência de pitch.....	65
Figura 7-3 Arquivo02 Oradores masculinos – Sinal de voz.....	65
Figura 7-4 Arquivo02 Oradores masculinos – Análise de frequência de pitch.....	65
Figura 7-5 Arquivo01 Oradores femininos – Sinal de voz.....	66
Figura 7-6 Arquivo01 Oradores femininos – Análise de frequência de pitch.....	66
Figura 7-7 Arquivo02 Oradores femininos – Sinal de voz.....	66
Figura 7-8 Arquivo02 Oradores femininos – Análise de frequência de pitch.....	67
Figura 7-9 Inicialização dos três primeiros <i>frames</i> e ação da avaliação zero.....	69
Figura 7-10 Avaliação dos ganhos no arquivo 01 de Oradores Femininos.....	70
Figura 7-11 Avaliação dos ganhos no arquivo 01 de Oradores Masculinos	71
Figura 7-12 Avaliação dos ganhos no arquivo 02 de Oradores Femininos.....	71
Figura 7-13 Avaliação dos ganhos no arquivo 02 de Oradores Masculinos	72
Figura 8-1 Gráfico da aplicação do ganho no arquivo 01 de Oradores Masculinos	77
Figura 8-2 Arquivo1 Oradores masculinos, sinal codificado e decodificado com atuação do PESQ	79
Figura 8-3 Arquivo1 Oradores masculinos, análise da frequência de pitch.....	79
Figura 8-4 Gráfico da aplicação do ganho no arquivo 01 de Oradores Femininos	80
Figura 8-5 Arquivo1 Oradores femininos, sinal codificado e decodificado com atuação do PESQ	82
Figura 8-6 Arquivo1 Oradores femininos, análise da frequência de pitch.....	82
Figura 8-7 Gráfico da aplicação do ganho no arquivo 02 de Oradores Masculinos	83
Figura 8-8 Arquivo2 Oradores masculinos, sinal codificado e decodificado com atuação do PESQ	85
Figura 8-9 Arquivo2 Oradores masculinos, análise da frequência de pitch.....	85
Figura 8-10 Gráfico da aplicação do ganho no arquivo 02 de Oradores Femininos.....	86
Figura 8-11 Arquivo2 Oradores femininos, sinal codificado e decodificado com atuação do PESQ	88
Figura 8-12 Arquivo2 Oradores femininos, análise da frequência de pitch.....	88

LISTA DE TABELAS

Tabela 4-1 – Escala de classificação do MOS.....	43
Tabela 4-2 – Comparação de eficiência do Modelo PSQM & PESQ.....	47
Tabela 4-3 – Valores de referência do MOS para os codificadores <i>Standards</i> utilizados em telefonia.....	47
Tabela 5-1 – Comparação de <i>codecs</i> ITU-T para telefonia.....	48
Tabela 5-2 – Parâmetros codificados pelo G729a (ITU-T G729).....	54
Tabela 7-1 – Avaliação do PESQ para adição de ganho nos sinais de pitch e codificação	68
Tabela 7-2 – Ganhos avaliados para os arquivos de teste	70
Tabela 7-3 – Ajuste do controle de ganho para os arquivos de teste.....	73
Tabela 8-1 – Testes somente com o PESQ inserido no codificador.....	76
Tabela 8-2 – Resultado do programa executado no arquivo1 de Oradores Masculinos	77
Tabela 8-3 – Avaliação dos parâmetros extraídos do Arquivo1 de Oradores Masculinos.....	78
Tabela 8-4 – Resultado do programa executado no arquivo1 de Oradores Femininos.....	80
Tabela 8-5 – Avaliação dos parâmetros extraídos do Arquivo1 de Oradores Femininos	81
Tabela 8-6 – Resultado do programa executado no arquivo2 de Oradores Masculinos	83
Tabela 8-7 – Avaliação dos parâmetros extraídos do Arquivo2 de Oradores Masculinos.....	84
Tabela 8-8 – Resultado do programa executado no arquivo2 de Oradores Femininos.....	86
Tabela 8-9 – Avaliação dos parâmetros extraídos do Arquivo2 de Oradores Femininos	87
Tabela 9-1 – Análise do ganho de pitch sobre a proposta de avaliação	90

LISTA DE SIGLAS

CS-ACELP	- Conjugate Structure Algebraic Code Excited Linear Prediction
PESQ	- Perceptual Evaluation of Speech Quality
MOS	- Mean Opinion Score
ITU-T	- Telecommunication Standardization of International Telecommunication Union
PCM	- Pulse Code Modulation
G.729	- Recomendação ITU-T código de voz à 8kbps para CS-ACELP
G.729.A	- Recomendação ITU-T código de voz à 8kbps para CS-ACELP Anexo A
GSM	- Global System for Mobile Communications
RELp	- Residual Excited Linear Prediction
CODEC	- Dispositivo de hardware ou software que codifica e decodifica sinais.
ENCODER	- Codificador do sinal de voz amostrado
DECODER	- Decodificador do sinal de voz amostrado
LSF	- Line Spectrum Frequency
LSP	- Line Spectrum Pair
Timit	- Banco de dados de 630 oradores em 8 dialetos do inglês Americano.
Audacity	- Software de análise e manipulação de sinais de áudio.

1. Introdução

O sistema de telecomunicações abrange uma gama de sinais e serviços e os sinais constituem um ingrediente básico de nossa vida. Uma forma comum de comunicação humana é o uso de sinais de fala, seja conversação frente a frente ou por canal telefônico. Neste trabalho sobre codificação do sinal de voz utilizado em telefonia e redes de pacotes, a análise de qualidade perceptual do sinal codificado na fonte é o principal fator de estudo. Embora o estudo do sinal de voz e a avaliação perceptual sejam processos amplamente difundidos, a pesquisa sobre esses tópicos estão relacionadas com a abordagem do codificador utilizado, isto é, um algoritmo de avaliação perceptual, tendo como vantagem a análise do sinal de voz, com alterações dinâmicas em tempo real para melhoria em tempo real.

1.1. Objetivo

A análise de codificação de fala avalia alguns requisitos para busca de um desempenho satisfatório. Um codificador de voz pode ser útil por reduzir a taxa de transmissão apesar de aumentar a distorção, diminuindo o desempenho mesmo sendo mais eficiente na sua função. O procedimento experimental deste trabalho consiste em avaliar a qualidade perceptual do sinal codificado com os experimentos e alterações realizadas no *encoder* do codificador CS-ACELP (*Conjugate Structure Algebraic Code Excited Linear Prediction*).

Os trabalhos de avaliação de qualidade que foram pesquisados durante este projeto, enfatizavam o processo de codificar/transmitir/decodificar o sinal e avaliar externamente o sinal original de entrada no *encoder* com o sinal de saída do *decoder*. A qualidade estava associada às alterações que o sinal sofria quando submetido às ações do meio físico¹. Neste trabalho foi avaliada a estrutura do codificador com suas alterações propostas para melhoria perceptual, excluindo-se as variações do meio de transmissão.

¹ Entende-se por ações do meio físico toda oscilação que ocorre no sinal de voz: degradação, perda, variação e atraso.

1.2. Motivação

Avaliar as características do sinal de voz amostrado no bloco do codificador, através da implementação do algoritmo do PESQ (*Perceptual Evaluation of Speech Quality*) no código fonte do codificador G729a possibilitando, assim, a avaliação perceptual do sinal, independente das variações do meio de transmissão entre os blocos *encoder* e *decoder*.

Implantar um modelo de treinamento para os padrões do algoritmo PESQ sobre o bloco do codificador, fazendo com que exista coerência no valor dos resultados do MOS (*Mean Opinion Score*), tendo como base as especificações da recomendação P.862 do ITU-T, a qual descreve o tamanho da amostra em relação ao tempo de envio dos pacotes.

Propor alterações na estrutura do bloco *encoder* do codificador CS-ACELP para obter uma melhoria do sinal codificado analisado pelo algoritmo PESQ.

Avaliar a codificação do sinal de voz e a sua qualidade perceptual baseado nos parâmetros restritos ao codificador, ou seja, o objetivo é melhorar qualidade perceptual nos frames codificados.

1.3. Estrutura da Dissertação

Esta dissertação de mestrado foi estruturada em três grandes partes compostas por capítulos conforme abaixo:

Parte I – Fundamentos Teóricos

Capítulos: 2 ,3,4 e 5

Esses capítulos abordam a fisiologia do aparelho fonador, o modelo de produção de voz, princípios de codificação do sinal de voz, modelo de quantização e análise para mensurar a qualidade do sinal de voz. São descritos os conceitos relacionados à arquitetura dos codificadores, introduz a codificação de sinais e aborda o *codec* G729a e os parâmetros específicos para análise desta pesquisa.

Parte II – Metodologia

Capítulos: 6 e 7

Esses capítulos detalham a proposta de avaliação dos parâmetros escalares do G729a, bem como a implementação do algoritmo PESQ, enfatizando as particularidades e limitações para análise dos *frames* de voz. Assim, com a metodologia aplicada pretende-se obter uma qualidade perceptual através do estudo do comportamento dos parâmetros escalares, que atuam sobre o sinal de voz codificado.

Parte III – Resultados e Conclusões

Capítulos: 8 e 9

Esses capítulos identificam os objetivos de melhoria perceptual sobre as alterações passíveis de modificação da estrutura do codec. Também são apresentadas as especificações para obtenção dos resultados, descrevendo possíveis desenvolvimentos de trabalhos futuros sobre a proposta da técnica apresentada nesta dissertação.

2. Análise do Aparelho Fonador

O processo de produção da fala utilizado para a comunicação entre os interlocutores dá-se através das ondas acústicas que são formadas e emitidas pelo sistema vocal. Este processo tem como fonte a variação de pressão gerando, assim, um sinal variante no tempo em que altera suas características quando o sistema vocal tem sua dimensão e forma alterada.

2.1. Processo Fisiológico de Produção de Voz

O sistema de produção de voz é formado pelos órgãos da fala: pulmões, laringe, traquéia, faringe, cavidade nasal e cavidade oral. Na figura 2.1 podemos verificar os órgãos que integram o sistema vocal (FURUI, 2001). Esta divisão é feita em três grandes grupos: pulmões, laringe e trato vocal.

O sinal de voz é originado pelo fluxo de ar dos pulmões, esse fluxo é percorrido pela traquéia até a laringe.

A laringe é formada por quatro cartilagens, possui as cordas vocais na parte superior, um par de estruturas elásticas de tendão e músculos. A movimentação destes músculos da laringe faz com que as cordas vocais possam variar de comprimento e espessura, tendo diferentes configurações. Durante esse processo de respiração, a glote está normalmente aberta. Quando a glote se fecha ocorre obstrução do fluxo de ar que vem dos pulmões. Neste instante é que ocorre a vibração das cordas vocais. Quando a glote está aberta, a passagem de ar e as cordas vocais não vibram.

Durante esse processo, o ar originado nos pulmões tendo sua variação de pressão na laringe, chega até a faringe e a cavidade oral, o que forma o trato vocal, o qual inicia na glote e vai até a irradiação dos lábios. Existe uma cavidade auxiliar, a cavidade nasal, a qual é acoplada ao trato vocal para a produção de sons nasais.

O aparelho fonador humano é composto pelos órgãos responsáveis pela geração dos sons da voz, em que cada órgão efetua seu trabalho de forma dinâmica e simultânea durante o processo da fala.

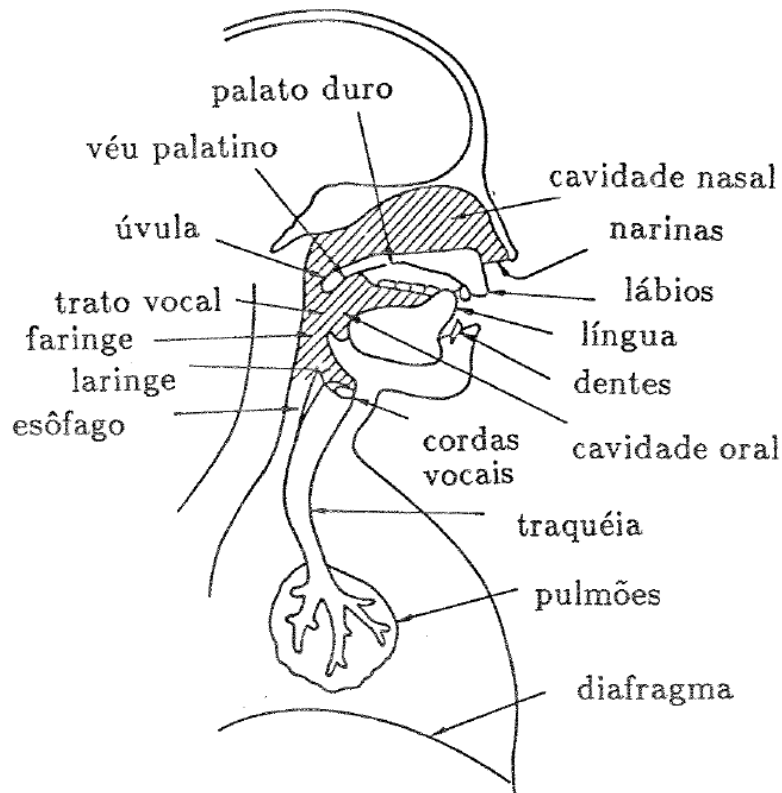


Figura 2-1 – Órgãos de produção da fala – Sistema do Trato Vocal.

2.1.1. Trato Vocal

A descrição do trato vocal (O'SHAUGHNESSY, 1999) relata que os pulmões são responsáveis por fornecer o fluxo de ar e pressão para o processo de fala. As cordas vocais geralmente modulam o fluxo de ar para criar um som, mas é o trato vocal o componente mais importante na produção da fala.

Uma passagem tubular composta de tecidos musculares e ósseos, o trato vocal, fornece os meios para produzir os sons diversos que caracterizam a linguagem falada.

O trato vocal tem duas funções:

- pode modificar a distribuição espectral de energia das ondas sonoras da glote;
- pode contribuir para a geração de som obstruído (fricativos).

Diferentes sons são distinguidos principalmente por suas características: periodicidade (fala ou silêncio), forma espectral (frequências com seus níveis de energia) e a duração. As cordas vocais especificam a característica sonora e duração de um som, são resultados das ações sincronizadas, mas a divisão principal do sinal de voz é realizada pelo trato vocal via filtragem espectral.

2.1.2. Sistema de produção de voz e as suas formantes

O sistema de produção de voz é formado por um conjunto de frequências de ressonância e anti-ressonância e estas dependem do formato do trato vocal. O trato vocal é um tubo acústico com área e secção transversal não uniforme e variável com o tempo, em que as frequências de ressonância do tubo são classificadas como formantes.

O movimento dos órgãos de produção da fala, quando irradiados para o espaço livre, tanto pelos lábios quanto pelas narinas, alteram a forma do tubo acústico, conseqüentemente alteram a resposta em frequência. Cada forma do trato vocal é representada por um conjunto de formantes. Diferentes sons irradiados variam a frequência de ressonância e as suas formantes. Abaixo, na figura 2.2, o sinal de voz e, na figura 2.3, as formantes deste sinal distribuídas no tempo.

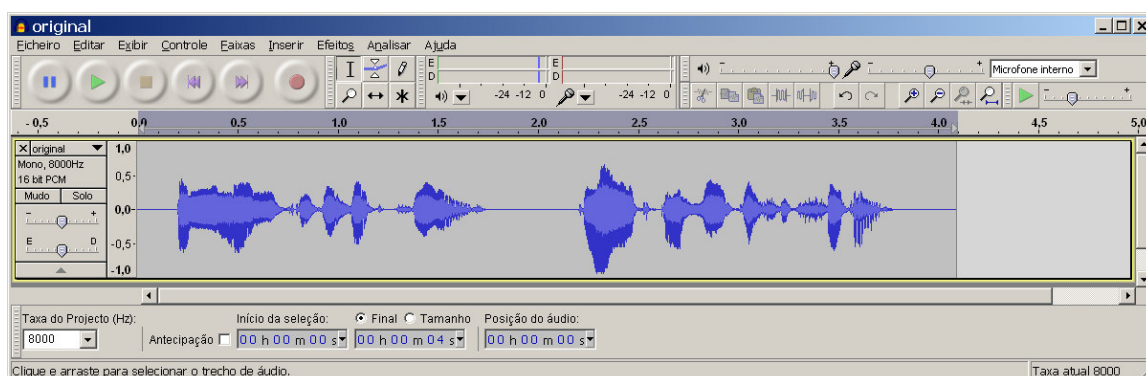


Figura 2-2 – Segmento sonoro de voz com segmentos surdos de voz – sinal irradiado

Abaixo a análise do espectro de frequência de *pitch* realizada no sinal da figura 2.2. Os valores mais escuros (em vermelho) representam as formantes do sinal de voz, sendo a primeira formante a frequência fundamental ou período de *pitch*.

Na figura 2.3, o arquivo formatado possui sentenças com o sinal de voz elaborado por seis oradores do gênero masculino. Observa-se na primeira formante que a oscilação do sinal é associada com a frequência de *pitch* do orador.

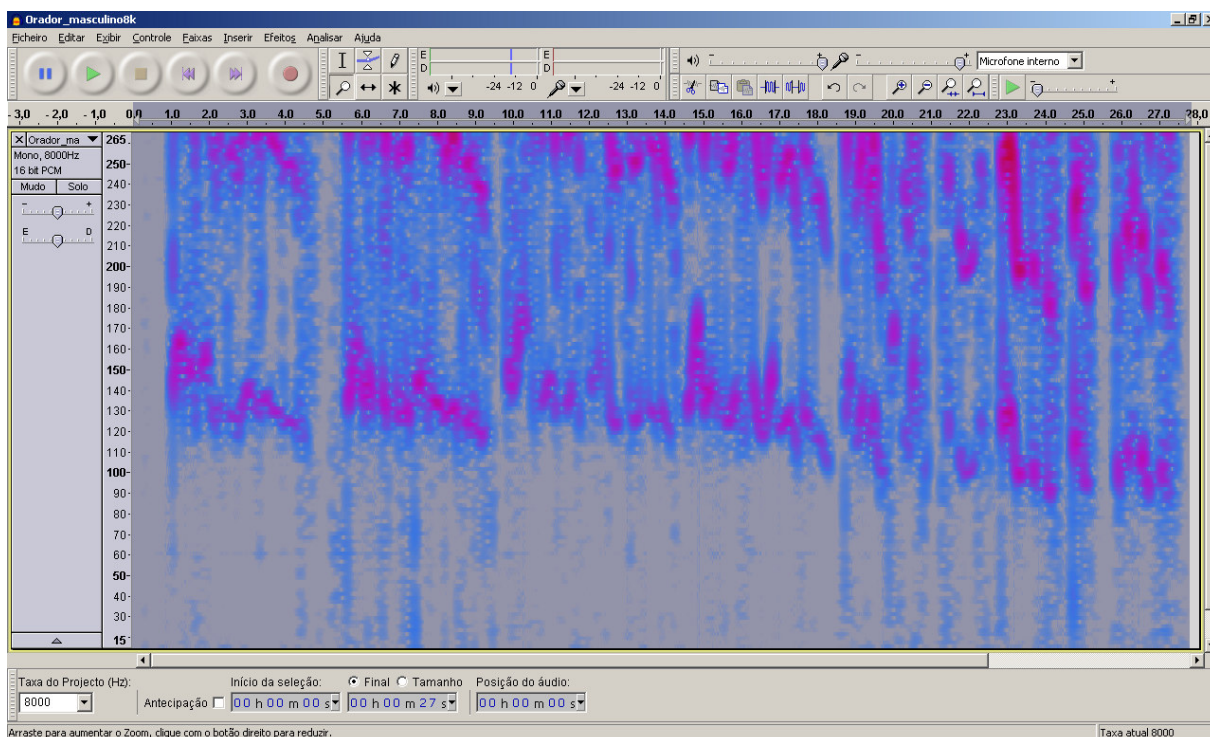


Figura 2-3 – Ptch do sinal de voz amostrado

O sistema vocal, conforme excitação do trato vocal, classifica-se em três categorias: sonoros, fricativos e explosivos.

Sons Sonoros (vocálicos)

Ocorre quando a pressão de ar aumenta nos pulmões e força a passagem do ar pela glote, fazendo com que as cordas vocais que estão tensionadas, entrem em vibração. Isso porque a passagem de fluxo de ar, quase periódico, gera a excitação do trato vocal, os chamados pulsos glotais, os sons vocais são recorrentes em intervalos espaçados igualmente. A frequência fundamental ou frequência de *pitch* é a taxa de vibração das cordas vocais, a qual depende da pressão do ar na traqueia e da variação de espessura e comprimento das cordas vocais. A variação das cordas vocais, o tom, são os valores entre 50Hz e 500Hz. Um exemplo é a sonoridade da vogal “a”.

Sons Fricativos (não vocálicos):

Esses sons ocorrem quando existe o estreitamento criado pelos órgãos de produção de fala em algum ponto do trato vocal, e o ar dos pulmões tem velocidade suficiente para produzir uma turbulência. A localização da compressão do ar no trato vocal é que produz o

som fricativo. Um exemplo é a produção do som da vogal “f” (compressão do trato vocal com o a articulação dos lábios).

Sons Plosivos:

Esse tipo de som é resultante do fechamento completo de algum ponto do trato vocal, em que o ar originado pelos pulmões é interrompido, fazendo uma pressão atrás da obstrução. Quando o trato vocal é desobstruído, resulta em um abrupto relaxamento de pressão, gerando um som com baixa energia. Exemplos de som explosivo são as letra “p” e “b” .

O sistema de produção de voz é formado por frequências de ressonância e anti-ressonância, dependendo do formato do trato vocal. Os diversos sons produzidos no processo de fala são consequências da utilização das três formas de excitação do trato vocal.

As formantes, quando identificadas, podem detectar o gênero do orador através das frequências fundamentais dos sinais avaliados, em que o sinal produzido por mulheres e crianças é mais elevado do que o sinal produzido por homens.

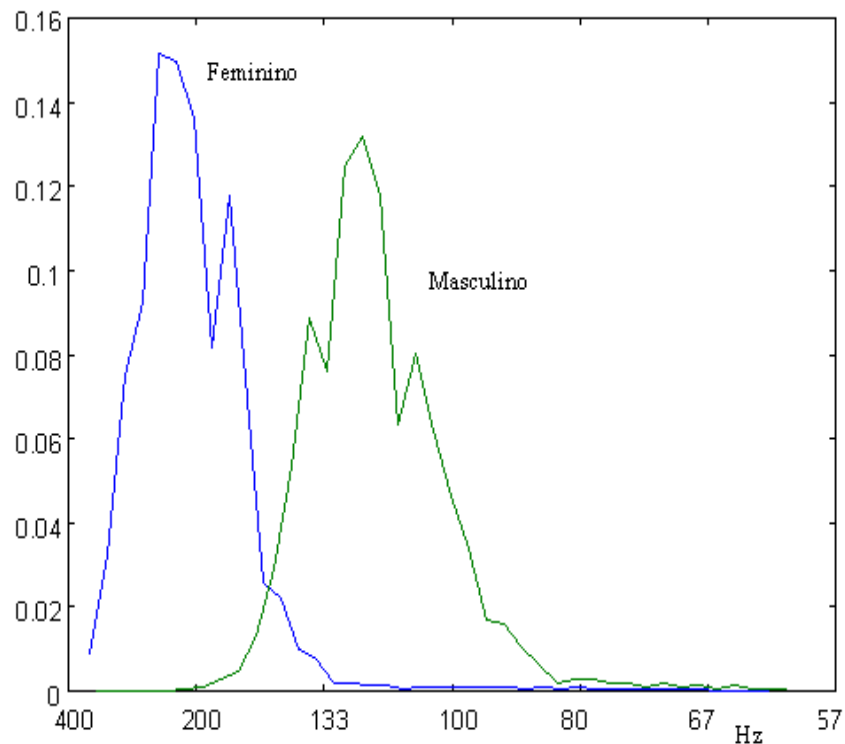


Figura 2-4 – Exemplo de historiograma da frequência fundamental

2.2. Modelo de Produção do Sinal de Voz

Para a produção de voz existe um modelo representado pelas fontes de excitação e pelo trato vocal que possuem uma independência. O sistema de produção de voz pode ser representado por um modelo linear invariante no tempo e um gerador de excitação, segundo (RABINER e SCHAFER, 1978).

O sistema modela as ressonâncias do trato vocal e os efeitos da radiação dos lábios. Essa representação pode ser modelada por tubos acústicos ou pelos filtros digitais.

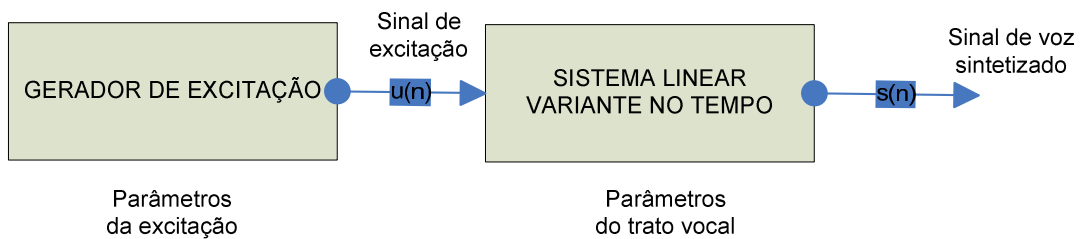


Figura 2-5 –Diagrama em blocos do modelo para produção de voz.

2.2.1. Gerador de excitação

Quando excitamos o trato vocal com um trem de impulsos quase periódico (*pitch*), os sons sonoros são gerados. Esse modelo de geração de excitação pode ser representado conforme figura abaixo.

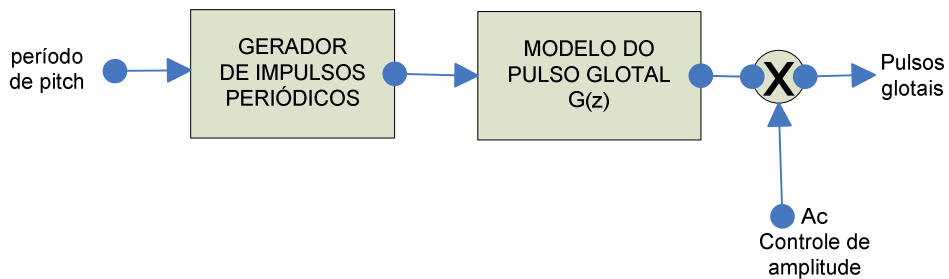


Figura 2-6 – Diagrama em blocos do sinal de excitação para sons sonoros

Para representar a frequência fundamental do sinal, um gerador de impulso produz o trem de impulsos unitários, que excitam o sistema linear com uma resposta impulsiva que é a forma de onda dos pulsos glotais, representada pela formulação abaixo:

$$g(n) = \begin{cases} \frac{1}{2} \times \left[1 - \cos\left(\frac{\pi \times n}{N_1}\right) \right] & \text{se } 0 \leq n \leq N_1 \end{cases}$$

$$g(n) = \begin{cases} \cos\left(\frac{\pi \times (n - N_1)}{2 \times N_2}\right) & \text{se } N_1 \leq n \leq N_1 + N_2 \\ 0 & \text{caso contrário} \end{cases} \quad (2.1)$$

$$g(n) = 0 \quad \text{caso contrário}$$

A função $g(n)$ tem comprimento finito e sua transformada Z apresenta apenas zeros.

A representação para $g(n)$, utilizando um modelo de dois pólos, pode ser formulada para representar $G(Z)$, tendo a seguinte formulação:

$$G(Z) = \frac{1}{(1 - e^{c \times T} \times Z^{-1})^2} \quad (2.2)$$

No domínio da frequência, o pulso glotal introduz um efeito passa-baixas, onde a intensidade do pulso glotal é controlada por um controle de ganho.

2.2.2. Modelagem do Trato vocal

O trato vocal pode ser modelado como uma associação de tubos conectados em cascata, tendo variação na área da secção transversal. Dessa forma, a frequência de ressonância em cada tubo corresponde a uma formante do espectro do sinal de voz.

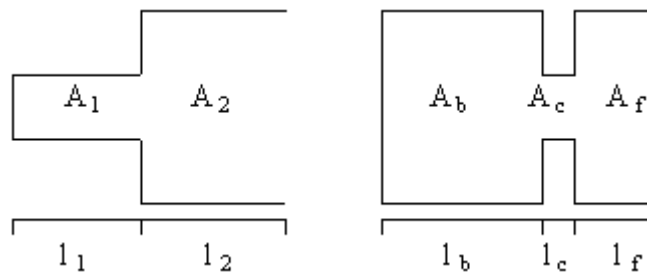


Figura 2-7 – (a) Exemplo de dois tubos e três tubos.

Na figura acima os modelos segundo (O'SHAUGHNESSY, 1999) representam vogais e consoantes. O modelo de dois tubos com secções A_1 e A_2 representam uma aproximação das vogais, por exemplo a letra "a". O tubo estreito (secção A_1) representa a abertura da faringe e o tubo maior (secção A_2) representa a cavidade oral, considerando que do ponto de vista técnico, o sistema de voz é um único tubo acústico entre a glote e a boca, medindo 17cm, para representar a vogal "a" em que temos dois tubos. Iremos considerar comprimentos iguais ($l_1 = l_2$) medindo 8,5cm, sendo as formantes múltiplos de 1kHz e devido ao acoplamento não se aproximam uma das outras por 200Hz, temos que $F_1 \approx 900\text{Hz}$ e $F_2 \approx 1100\text{Hz}$ e $F_3 \approx 2900\text{Hz}$ e $F_4 \approx 3100\text{Hz}$. Já o modelo de três tubos com

secções A_b , A_c e A_f representa uma aproximação das consoantes, sendo um modelo de tubo estreito com constrição do trato vocal, em que a parte traseira (secção A_b) e o tubo do meio (secção A_c) são ressonadores de meio comprimento de onda, e o tubo dianteiro (secção A_f) é um ressonador de quarto de onda, com ressonâncias $\frac{c \times i}{2 \times l_b}$, $\frac{c \times i}{2 \times l_c}$, $\frac{c \times (2 \times i - 1)}{4 \times l_f}$, para $i=1,2,3,\dots$ onde c (velocidade do som) e l_b, l_c, l_f são os comprimentos dos tubos, gerando ressonâncias em múltiplos de 5.333Hz para constrições de 3cm em períodos de fala com duração típica, podendo ser desconsiderado em aplicações que utilizam sinais de voz de 4 ou 5kHz de largura de banda.

A função de transferência do trato vocal pode ser modelada pela formulação:

$$V(Z) = \frac{G}{\prod_{i=1}^N (1 - p_i \times Z^{-1})} \quad (2.3)$$

Na equação acima, estamos desprezando os efeitos gerados pela radiação dos lábios. Temos o ganho G associado à amplitude do sinal de voz e os pólos p_i , com i sendo o índice de cada pólo, variando de 1 até N , esses pólos fazem a modelagem da frequência de ressonância do trato vocal.

O modelo apresentado é uma boa representação do trato vocal, ele modela apenas as frequências de ressonâncias, que abrange a maioria dos sons vocálicos, pois para considerarmos a produção de sons fricativos e nasais, precisamos representar as frequências de anti-ressonâncias. Para isso seria necessário que, na formulação de transferência do trato vocal ($V(Z)$), tivéssemos pólos e zeros. Entretanto, uma forma de conseguirmos o efeito dos zeros, é aumentando o número de pólos da função.

Desta forma, o trato vocal é representado por um sistema linear, um sistema estável, formado somente por pólos, onde todos os pólos de $V(Z)$ estão dentro do raio de círculo unitário.

2.2.3. Radiação

No sistema de produção de fala, temos a atuação dos lábios e das narinas (no caso das vogais nasais como “avião”), que irradiam para o espaço livre os sons articulados pelo trato

vocal, ocorrendo uma difração das ondas sonoras. Esse efeito de radiação pode ser modelado como um filtro passa-altas (representa um ganho de 6dB por oitava).

$$R(Z) = R_o(1 - Z^{-1}) \quad (2.4)$$

2.2.4. Modelo completo de produção de voz

O sistema vocal para a produção da voz pode ser representado no modelo completo abaixo:

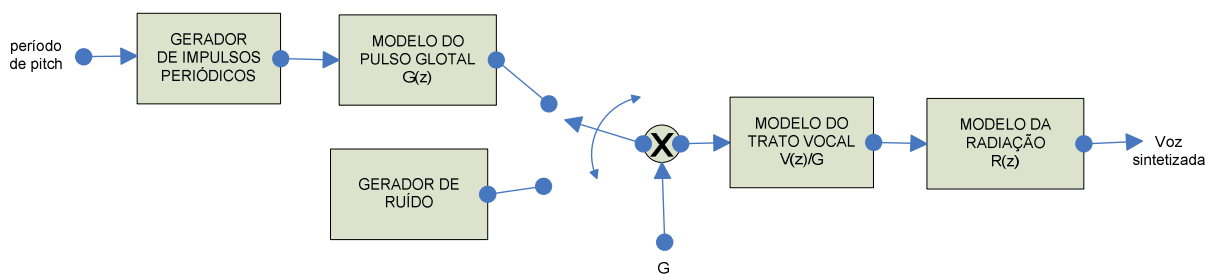


Figura 2-8 – Modelo completo para produção de voz.

Pela representação acima (RABINER e SCHAFER, 1978), podemos definir um modelo combinando as funções de transferência do pulso glotal do trato vocal e da radiação, em que temos:

$$H(Z) = G(Z) \times V(Z) \times R(Z) \quad (2.5)$$

Com o equacionamento acima, temos um modelo simples de para a produção de voz, que pode ser representado por um modelo apenas com pólos, onde a função $H(Z)$ é descrita como:

$$H(Z) = \frac{G}{1 - \sum_{k=1}^P a_k \times Z^{-k}} \quad (2.6)$$

Para representar o sinal de voz variante no tempo, são atualizados em intervalos regulares os coeficientes de $H(Z)$ e o sinal de excitação, tendo como resposta a produção de voz sintetizada.

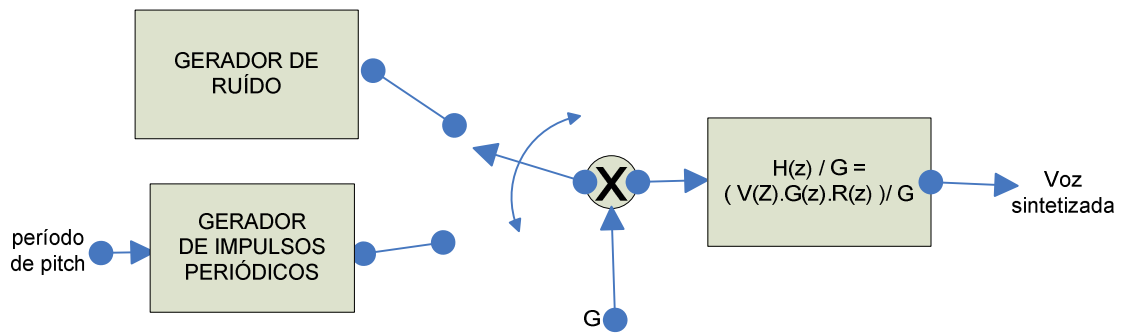


Figura 2-9 – Modelo simplificado para produção de voz.

O modelo apresentado na figura 2-9 (RABINER e SCHAFER, 1978) tem uma boa representação para sons que possuem variações lentas, a exemplo das vogais. Entretanto, esse modelo não consegue representar fielmente os sons transitórios, a exemplo dos sons plosivos. Já os sons fricativos sonoros, cuja excitação é gerada pela combinação de ruídos e pulsos periódicos, não são possíveis representar a partir desse modelo, pois uma forma de excitação (ruído ou pulso periódico) exclui a outra. O som originado nesse processo consegue obter uma voz sintetizada com boa qualidade, mas com perda de naturalidade.

3. Codificação de voz em DSP, Processamento Digital de Sinais

Um sinal é definido como uma função de uma variável, no caso dos sinais de voz. A fala é um sinal unidimensional. A amplitude varia com o tempo, dependendo da palavra falada e o locutor que fala essa palavra. Esse sinal é um veículo de informações sobre a natureza de um fenômeno físico. (HAYKIN e VEEN, 1999).

3.1. Sinais e comunicação

Existem três elementos fundamentais em todos os sistemas de comunicação: o transmissor, o canal e o receptor. Cada um desses elementos possui um sistema de sinais próprios associados, a figura 3.1 mostra essa relação.

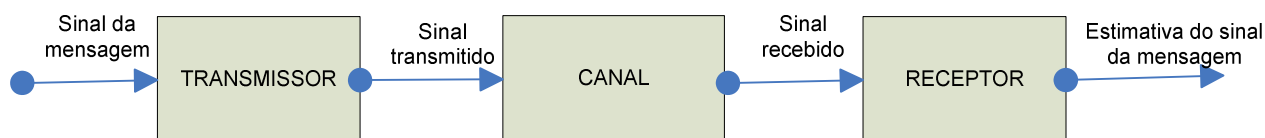


Figura 3-1 - Elementos de um sistema de comunicação.

O transmissor converte ou codifica o sinal da mensagem produzida por uma fonte de informação em um formato apropriado para ser transmitido pelo canal. O canal é o meio pelo qual o sinal se propaga. À medida em que o sinal é transmitido, ele sofre distorções devido as características físicas deste canal, tendo também a contribuição para degradação do sinal transmitido através de ruídos e interferências originados de outras fontes. O receptor é responsável por receber o sinal codificado, o qual pode ter sido corrompido (alterando a informação codificada), sendo este responsável pelo processamento do sinal convertendo (decodificando) o formato recebido pelo canal na informação da mensagem, em que o objetivo é estimar o sinal original.

3.1.1. Classificação do sinal para formação de modelos

A produção da fala é originada através de uma onda sonora gerada pela vibração das cordas vocais, que é uma consequência da propagação de ar emitida pelos pulmões, passando por todo trato vocal e irradiada pelos lábios. Esse processo tem como resultado um sinal de voz com sons sonoros e sons surdos.

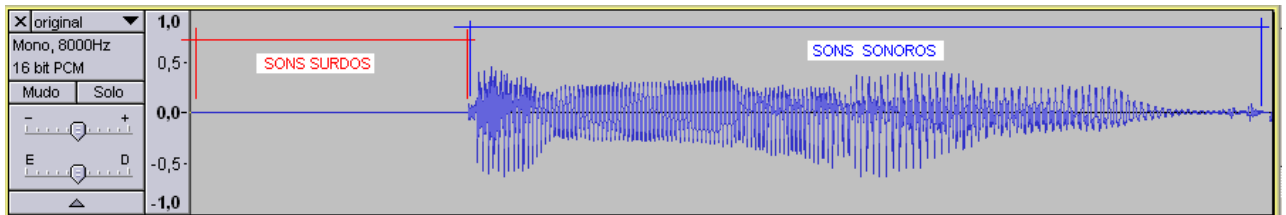


Figura 3-2 - Classificação do sinal de voz em sons sonoros e surdos

3.1.2. Amostragem do sinal

O processo de amostragem é importante para a transformação do sinal analógico contínuo em amplitude e no tempo, em um sinal discreto no tempo e em amplitude.

3.1.3. Quantização do sinal

Segundo (EMBREE e KIMBLE, 1991), "quantização é o domínio da amplitude de um sinal analógico contínuo amostrado em um certo domínio de tempo", ou seja, o processo de quantização é, basicamente, a medida discreta da intensidade do sinal. Essa discretização da amplitude é usualmente definida em termos de número de bits. Uma conversão de 8 bits, por exemplo, proporciona a representação de 256 níveis de quantização.

3.1.4. Codificação do sinal

O processo de codificação faz a representação do sinal amostrado entre o domínio contínuo e o domínio discreto. A otimização de representação desses sinais, a melhoria da codificação, está associada à quantização do sinal na origem do processo no domínio discreto.

Os sinais digitais são representados pela amostragem, quantização e codificação. O último estágio será tratado na abordagem da melhoria de qualidade perceptual. A conversão analógico/digital é um processo de três passos (PROAKIS & MANOLAKIS, 2007), conforme figura 3.3.

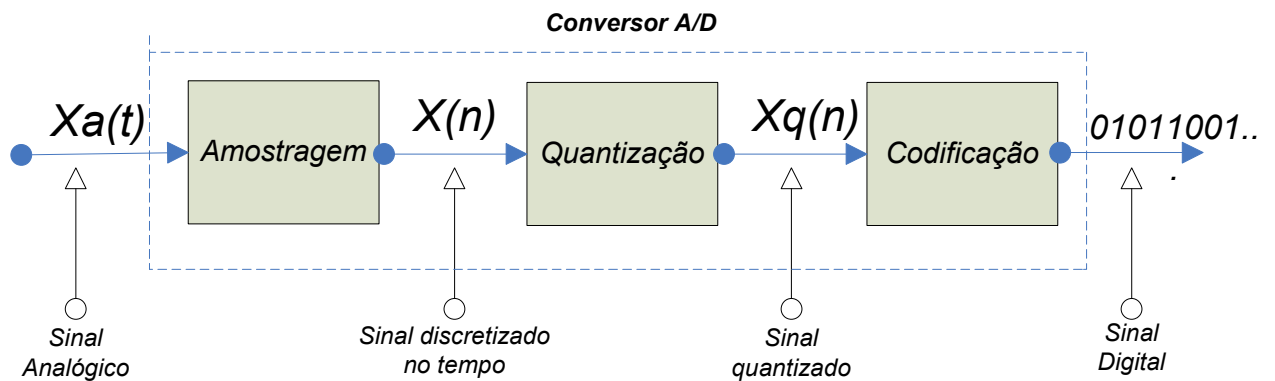


Figura 3-3 – Partes básica de um conversor analógico-digital (A/D)

3.1.5. Análise Homomórfica ou Análise Cepstral

Os sinais de voz são formados pelo o sinal de excitação e pela resposta impulsiva do trato vocal, conforme vimos no capítulo anterior. Neste contexto, existe uma técnica que é a Análise Homomórfica ou Análise Cepstral, muito útil para desconvoluir os dois sinais. A partir do modelo matemático para a produção de voz temos:

$$s(t) = e(t) \times v(t) \quad \Leftrightarrow \text{convolução do sinal} \quad (3.1)$$

$$S(W) = E(W) \cdot V(W) \quad \Leftrightarrow \text{sinal no domínio da frequência} \quad (3.2)$$

Para realizar a análise homomórfica é aplicada a função logarítmica para separar o sinal:

$$\log S(W) = \log(E(W) \cdot V(W)) \quad (3.3)$$

$$\log S(W) = \log E(W) + \log V(W) \quad (3.4)$$

Aplicando a transformada inversa nesse sinal, tem-se o *cepstrum* ou coeficientes *cepstrais* do sinal de voz.

$$F^{-1}[\log S(W)] = F^{-1}[\log E(W)] + F^{-1}[\log V(W)] \quad (3.5)$$

Com essa manipulação algébrica pode-se obter o sinal de excitação e a resposta impulsiva separadamente.

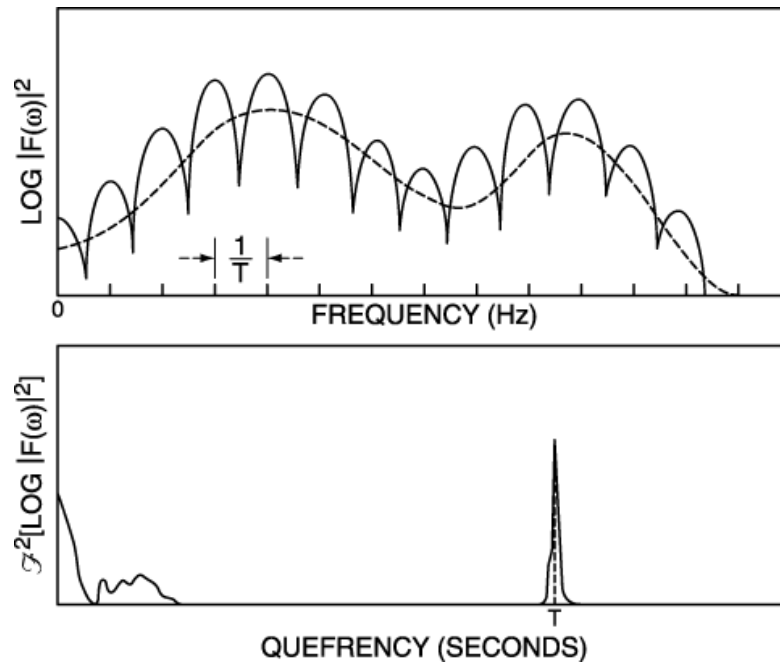


Figura 3-4 – Análise Cepstral do sinal de voz

3.2. Tipos de codificadores

Um paradigma para os codificadores de voz é alcançar a melhor qualidade com a menor taxa de bits possível. Neste processo, há três classificações: codificadores de forma de onda, codificadores paramétricos e codificadores híbridos. Essas classificações diferem na forma de como a informação é transmitida: os codificadores de forma de onda encaminham o sinal de voz e suas variações, os codificadores paramétricos encaminham parâmetros extraídos do sinal de voz original da mensagem e os codificadores híbridos realizam uma combinação dos dois casos citados anteriormente.

3.2.1. Codificadores de forma de onda

São codificadores de baixa complexidade, com pouco atraso. Esses codificadores reproduzem o mais exato possível a forma de onda analógica, em que o tratamento da mensagem é fielmente reproduzido incluindo, até mesmo, ruído de fundo. Uma característica desse codificador é a alta qualidade o sinal que eleva também a largura de banda para

transmissão. Um exemplo desse tipo de codificação é a recomendação ITU-T G.711 (PCM) que utiliza a taxa de transmissão de 64kbps (ITU G711, 1988).

3.2.2. Codificadores paramétricos

Esses codificadores avaliam o sinal de voz, não reproduzem a forma de onda original, constroem um conjunto de parâmetros que são enviados ao decodificador. A codificação de predição linear é utilizada para obter os parâmetros do filtro utilizado, isso causa um aumento na complexidade computacional, atraso no tempo de processamento, reduz a qualidade do sinal decodificado e, em troca, reduz a largura de banda para transmissão. Portanto, não são sinais expressivos para o uso nos sistemas de telefonia.

Nos codificadores paramétricos temos um gerador de pulsos (sons vocálicos), um gerador de ruído branco (sons não vocálicos). Esses dois parâmetros combinados formam a excitação $e(t)$, a excitação gerada passa por um filtro de pólos que representa o trato vocal $v(t)$ e tem como resultado a saída do sinal de voz gerado $s(n)$.

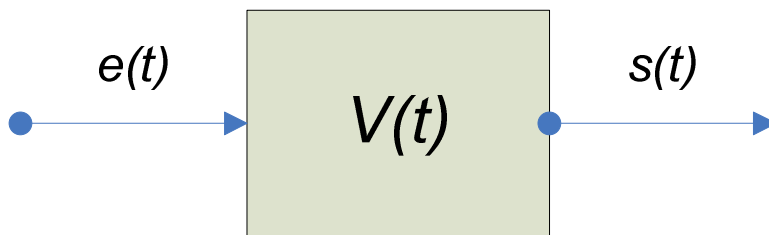


Figura 3-5 – Processo de geração de voz humana num codificador LPC

3.2.3. Codificadores híbridos

São codificadores que utilizam a técnica de análise por síntese baseados na predição linear. Esses codificadores utilizam as vantagens dos codificadores de forma de onda (reproduzir o sinal com alta qualidade) analisando as características espectrais e temporais do sinal, e as vantagens dos codificadores paramétricos (enviar sinal com baixa taxa transmissão) que conseguem extrair os parâmetros do sinal, o que resulta na máxima dos codificadores de voz, “fazer mais por menos”.

Será analisado, no capítulo seguinte, a técnica de codificação CELP (*Code-Excited Linear Prediction*), sobre o estudo do codificador G729a, o qual é a base o tema de análise desta dissertação.

Abaixo uma análise realizada por (GOMES, 2008), descreve os tipos de codificadores e a qualidade obtida em relação à taxa de transmissão. Podemos observar que os codificadores híbridos, por reunirem a característica dos codificadores de forma de onda e paramétricos, são os que possuem melhor qualidade com baixas taxas de transmissão (valores aceitáveis para o sistema de telefonia, otimizando o sinal de voz entre 2kbps e 16kbps) .

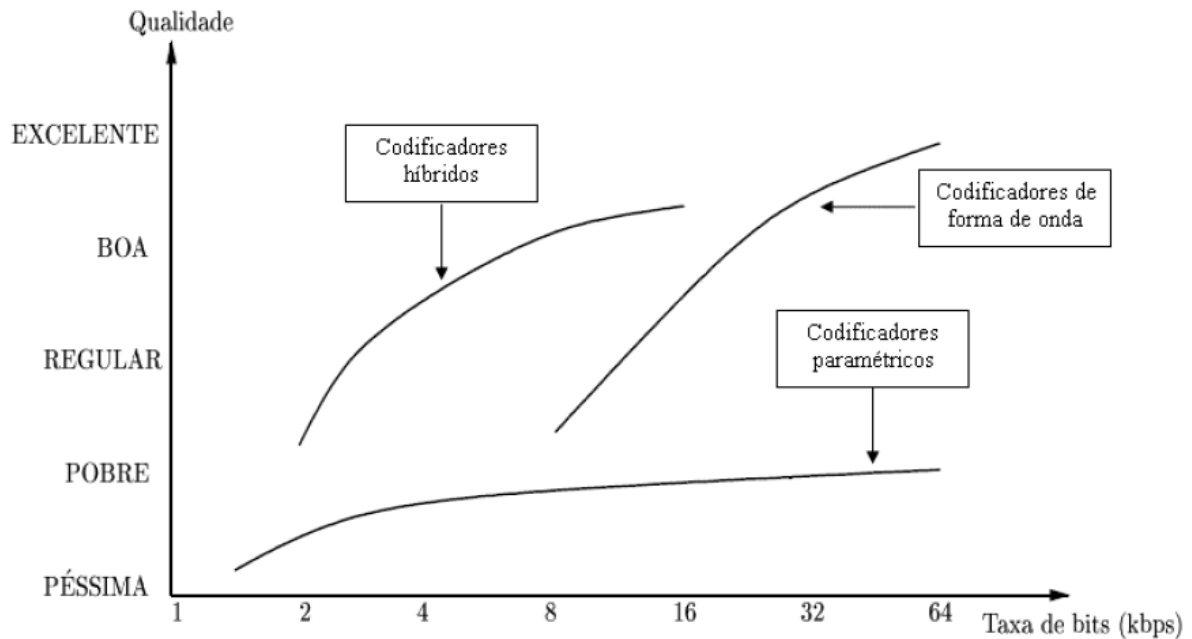


Figura 3-6 – Qualidade x Taxa de transmissão dos codificadores

3.3. Codificação Preditiva Linear

A predição linear é um arranjo muito importante na codificação do sinal de voz e sua formulação está relacionada à produção da fala. É uma técnica comum para a codificação de baixa taxa de bits e também uma importante ferramenta na análise do sinal de voz.

3.3.1. Análise da Codificação Preditiva Linear

A codificação linear é uma síntese para estimativa da frequência fundamental, realizando uma análise sobre as funções do trato vocal, representando as formantes, em que se deseja estimar o sinal amostrado baseado numa combinação linear utilizando filtros digitais.

Foi verificado, no capítulo anterior, o modelo de produção de fala e o modelo do trato vocal, conforme figura abaixo:

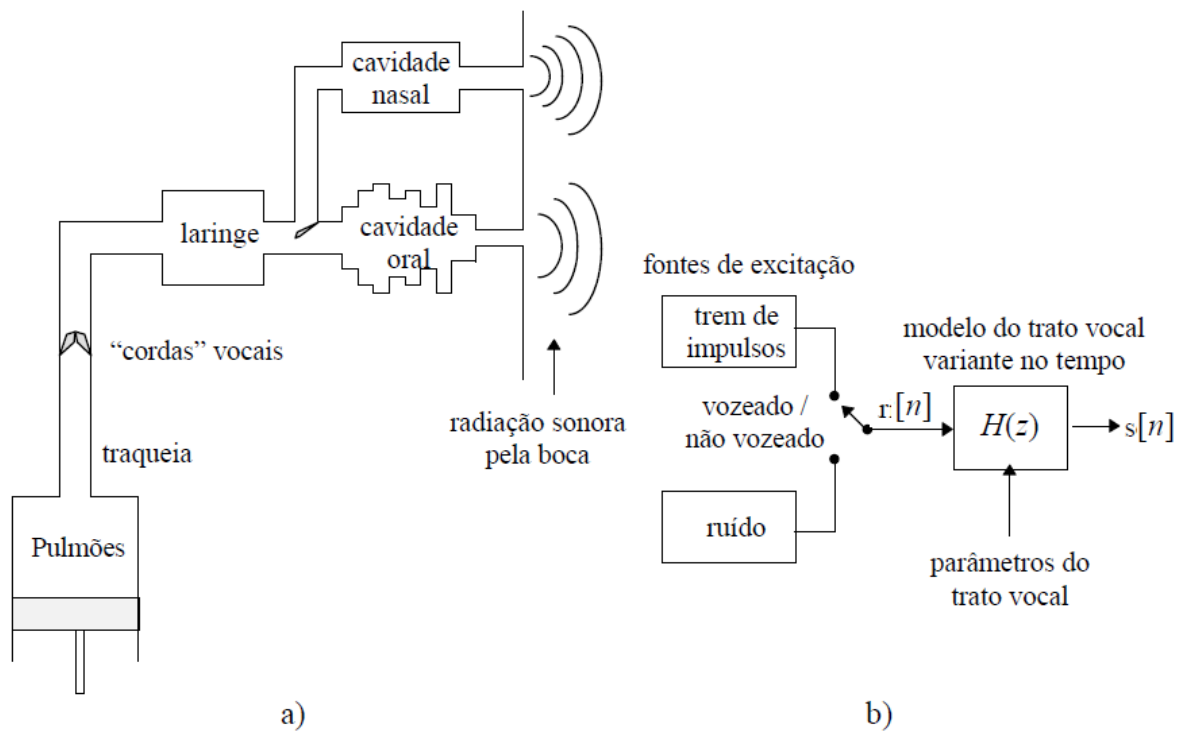


Figura 3-7 a) Modelo de produção da fala b) Modelo do trato vocal

Sabemos que o trato vocal é um modelo auto-regressivo (AR), formado por um filtro só de pólos descrito pela equação:

$$s(n) = b_0 \times u(n) - \sum_{k=1}^m c_k \times s(n-k) \quad (3.6)$$

Em que $s(n)$ corresponde ao sinal de voz sintetizado e $u(n)$ ao sinal de excitação que passa pelo filtro $H(Z)$. Nesta análise, existe a modelagem os parâmetros reais c_k que são desconhecidos. Para modelar esses parâmetros reais, tenta-se obter uma predição, ou seja, uma estimativa de $s(n)$ através da equação:

$$s(n)_{estimado} = \sum_{k=1}^p a_k \times s(n-k) \quad (3.7)$$

Esse equacionamento modela um filtro não-recursivo (FIR), que é um preditor linear de ordem p , em que os parâmetros a_k representam os coeficientes do preditor.

Para este procedimento é considerado que não há uma solução computacionalmente praticável que modele o filtro de forma totalmente fiel ao sinal original de voz, sendo assim iremos utilizar uma quantidade p de amostras passadas de voz para predizer o sinal com o menor erro possível.

O sinal residual $r(n)$ ou sinal de erro de predição é o resultado da diferença entre o sinal de voz e a sua aproximação, $r(n) = s(n) - s(n)_{estimado}$.

Para que o sinal de voz estimado tenha uma boa aproximação do sinal de voz original, devem-se obter valores ótimos para os coeficientes a_k do preditor dentro do intervalo de tempo $n_1 \leq n \leq n_2$, esse valor é representado por:

$$\{a_k\}_{otimo} = ArgMin_{a_k} \sum_{n=n_1}^{n_2} r^2(n) \quad (3.8)$$

Estabelecendo que o sinal de voz é obtido pelo $s(n)_{estimado}$ descrito da equação (3.7), e assumindo que $p=m$, resulta que os coeficientes a_k são uma boa estimativa dos coeficientes c_k descrito na equação (3.6).

O sinal de erro de predição no domínio tempo é descrito por:

$$r(n) = s(n) - s(n)_{estimado} = s(n) - \sum_{k=1}^p a_k \times s(n-k) \quad (3.9)$$

Aplicando a transformada Z na expressão, tem-se:

$$R(z) = A(z) \times S(z), \quad (3.10)$$

Nesta expressão $R(z)$ e $S(z)$ são respectivamente a transformada Z do sinal residual e a transformada Z do sinal de voz, em que $A(z)$ é um filtro de análise, formado somente por zeros, descrito por:

$$A(z) = 1 - \sum_{k=1}^p a_k \times Z^{-k} \quad (3.11)$$

Temos o filtro inverso de $A(z)$, que é um filtro de síntese, formado somente por pólos, que representa o comportamento espectral do sinal de voz:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^p a_k \times Z^{-k}} \quad (3.12)$$

Realizando uma substituição algébrica da equação (3.10) e (3.12), resulta em

$$S(z) = H(z) \times R(z) \quad (3.13)$$

Aplicando inversa da transformada Z na equação acima ou utilizando a equação (3.9), obtém-se o modelo, no domínio tempo, o modelo de síntese, representado pela expressão:

$$s(n) = \sum_{k=1}^p a_k \times s(n-k) + r(n) \quad (3.14)$$

Com esse desenvolvimento, podemos representar o processo de análise do sinal de voz, equações (3.9) ou (3.10), e o processo de síntese do sinal de voz equações (3.13) ou

(3.14) através dos diagramas de blocos “a” e “b” descritos (RABINER e SCHAFER, 1978), e representados na figura 3.8.

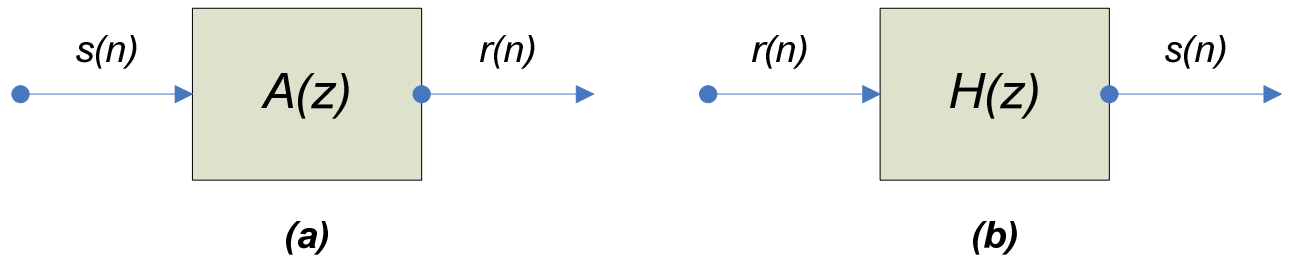


Figura 3-8 – Diagrama de blocos do processo de análise (a) e de síntese (b)

Os filtros de análise $A(z)$ e os filtros de síntese $H(z)$ são modelados pela relação de precisão espectral, complexidade computacional e quantidade de bits à transmitir. Neste caso, cada formante do espectro do sinal de voz é constituída por um par de pólos que estão separadas, em média, a cada 1kHz de banda. Para o sinal amostrado à taxa de 8kHz temos uma banda inferior a 4 kHz, em que o filtro de síntese utilizado é de ordem 10 geralmente, que seriam os 8 pólos do espaçamento espectral de 4kHz mais 2 pólos para evitar anti ressonância, aproximando possíveis zeros.

3.3.2. *Line Spectrum Frequency e Line Spectrum Pairs*

Para análise de predição linear são gerados os coeficientes de predição, coeficientes LPC. Muito sensíveis à transmissão do sinal de voz e ao processo de quantização, uma alternativa para a codificação dos coeficientes de predição é representar esses em coeficientes LSF (*Line Spectrum Frequency*) e LSP (*Line Spectrum Pairs*).

Segundo (O’SHAUGHNESSY, 1999), a representação LSF produz uma qualidade de voz melhor, pois provem a estabilidade dos coeficientes manipulados e possui propriedades adequadas, uma vez que os coeficientes de reflexão são muito sensíveis à quantização do sinal. Os coeficientes LSP são representados por dois polinômios $P(z)$ (simétrico) e $Q(z)$ (anti-simétrico), descritos abaixo:

$$P(z) = A(z) + z^{-(p+1)} \times A(z^{-1}) \quad (3.15)$$

$$Q(z) = A(z) - z^{-(p+1)} \times A(z^{-1}) \quad (3.16)$$

Como consequência desta manipulação algébrica, a representação LSF faz com que $P(z)$ e $Q(z)$ tenham as seguintes propriedades:

- todas as raízes dos polinômios estão sobre o raio de círculo unitário;
- as raízes dos polinômios $P(z)$ e $Q(z)$ estão entrelaçadas;

Esses dois modelos se relacionam com os coeficientes LPC representados pelo polinômio $A(z)$:

$$A(z) = \frac{P(z) + Q(z)}{2} \quad (3.17)$$

A definição acima é originada pelos cálculos de conversão dos coeficientes LPCs para os coeficientes LSFs, essa transformação foi realizada em 1975 por Itakura. Dessa análise, é definido que as raízes dos polinômios $P(z)$ e $Q(z)$ correspondem às frequências que são os coeficientes LFS, esses são extraídos do filtro de análise $A(z)$ de ordem p . Nesta verificação o polinômio $P(z)$ corresponde à análise do trato vocal com o a glote fechada, em que o coeficiente de reflexão $K_{p+1} = 1$ e o polinômio $Q(z)$ correspondem à análise do trato vocal com o a glote aberta, coeficiente de reflexão $K_{p+1} = -1$.

3.3.3. Janelamento

No sinal de voz amostrado, utilizado nos codificadores, temos uma gravação finita. Utilizamos a técnica de janelamento para aumentar as características do sinal amostrado, a operação algébrica é a multiplicação do sinal de voz $s(n)$ pelo sinal $W(n)$ (janela escolhida). A aplicação de uma janela sobre o sinal amostrado é para definir a duração do tempo de observação do sinal, para reduzir a perda espectral e separar sinais com frequências muito próximas e com amplitudes demasiadamente distantes.

Abaixo a avaliação realizada (FURUI, 2001) para janelamento dos sinais de voz.

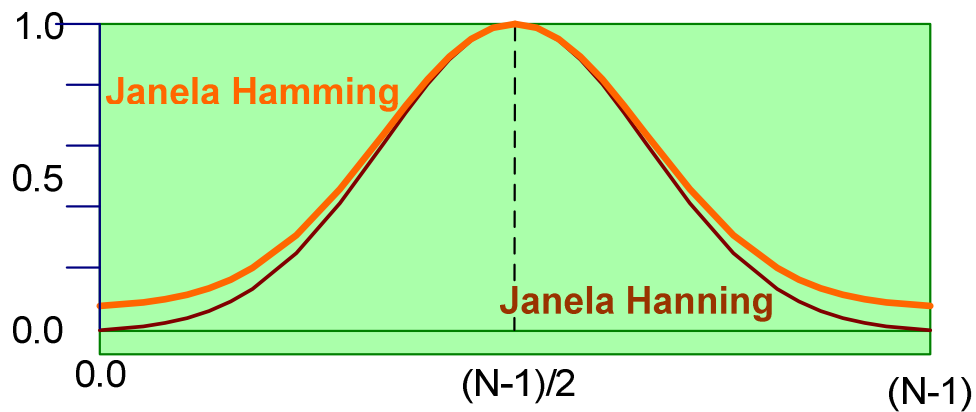


Figura 3-9 – Análise das janelas de Hanning e Hamming.

Para os casos específicos de sinais de voz, devido às características do sinal e a resposta em frequência que se deseja obter, será utilizada uma janela de Hamming, pois essa produz uma melhor resolução em frequência e é mais utilizada para o processamento de fala.

Abaixo, a função geradora do sinal para a janela de Hamming é dada por:

$$W(n) = 0,54 - 0,46 \times \cos\left(2\pi \times \left(\frac{n}{N-1}\right)\right); \quad \text{para } 0 \leq n \leq N-1 \quad (3.22)$$

· 0 para os demais casos

Para janelar o sinal de voz de forma a minimizar o erro introduzido pelas amostras dos extremos da janela, utiliza-se a sobreposição (*overlap*) dos intervalos dos frames de $s(n)$, onde amostramos o sinal atual e uma parcela de comprimento “L” do sinal anterior que serão consideradas no frame atual.

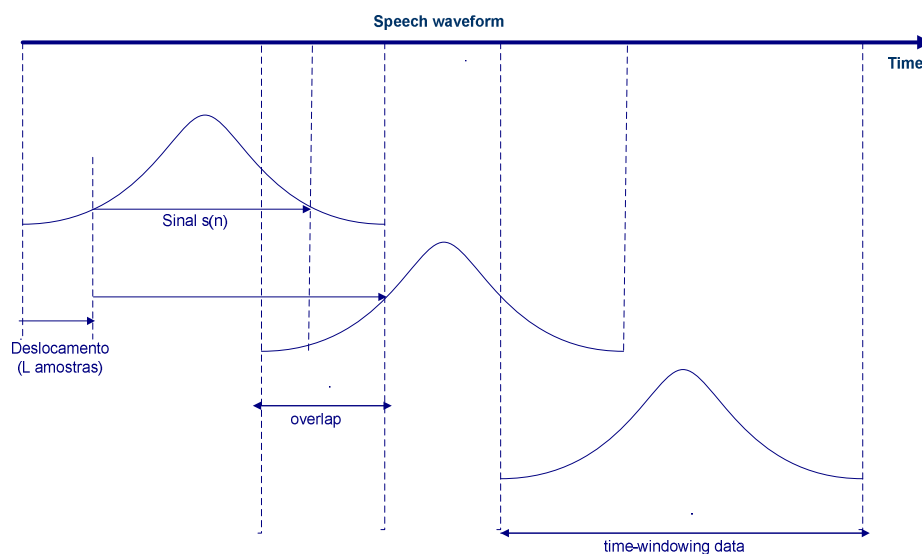


Figura 3-10 – Processo de janelamento

3.4. Quantização Vetorial

Quantização Vetorial (*VQ – Vector Quantization*) é o método de compressão de dados, em que ocorrem perdas com relação ao sinal original.

3.4.1. Entendendo quantização vetorial

O processo busca codificar um vetor de coeficientes (neste caso os coeficientes do filtro LPC) através da aproximação que será realizada com algum vetor de referência previamente calculado. Sendo assim, podemos considerar que a VQ é uma aproximação do vetor do coeficiente LPC com um dos vetores analisados da tabela de vetores códigos denominada de *codebook*.

Um exemplo de aproximação com a utilização de 1 bits é a figura abaixo:

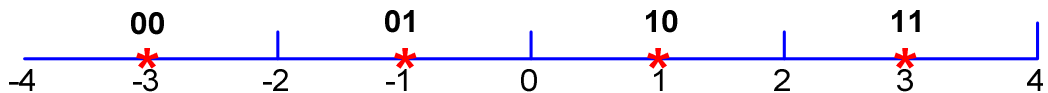


Figura 3-11 Exemplo de quantização em 1 dimensão

Na figura acima, cada número entre “-2” e “0” é aproximado por “-1”. Similarmente, cada número entre “+2” e “+4” é aproximado por “+3”. Dessa forma, temos uma quantização vetorial de uma dimensão (1-dimensional) e de com taxa de dois bits.

Um exemplo de quantização vetorial de 2 dimensões é demonstrado na figura 3.12. Existem 16 regiões, em cada região há um único ponto vermelho representado por 4 bits.

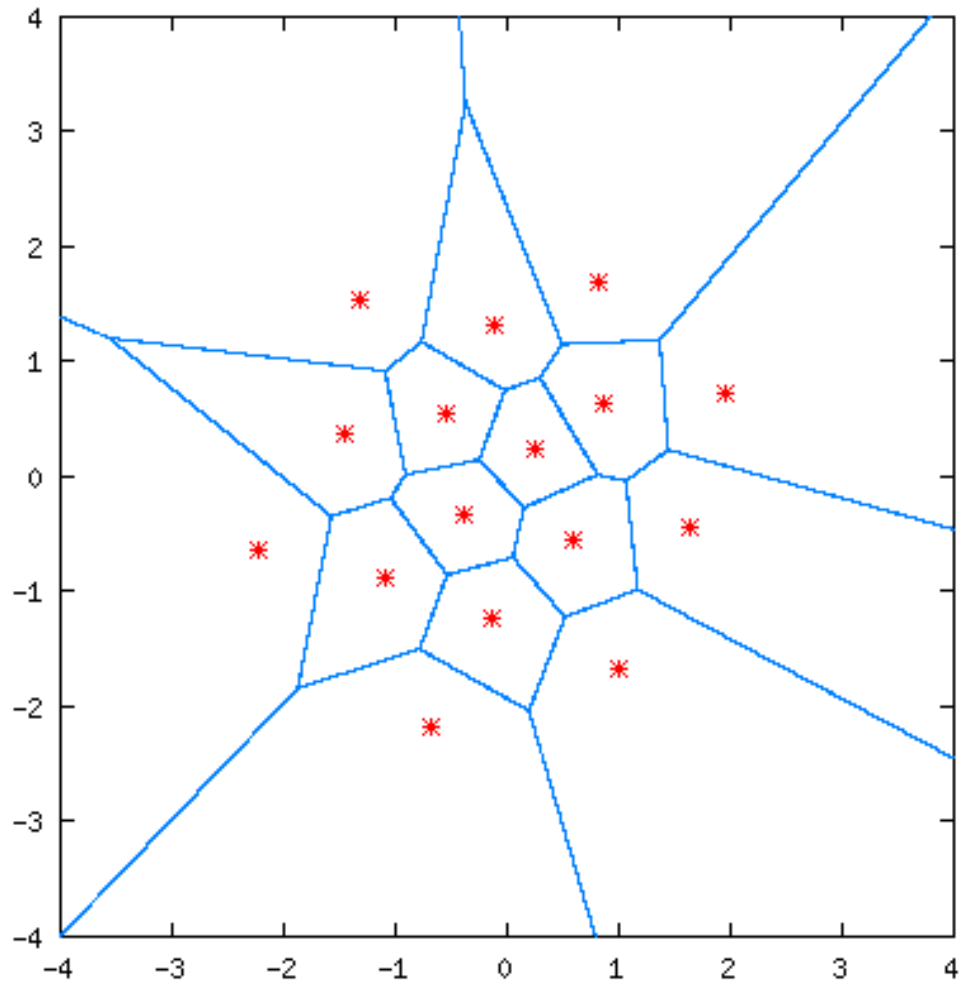


Figura 3-12 – Exemplo de quantização em 2 dimensões

Nota-se nos exemplos ilustrados nas figuras 3.11 e 3.12 que pontos de cor vermelha são os vetores de códigos ou *codevectors* e as áreas definidas pelos traçados da cor azul são as regiões de codificação. O conjunto de todos os vetores de códigos é o *codebook*.

Para gerar um conjunto de vetores de códigos é realizado um treinamento onde todos os vetores são amostras do sinal de voz obtidas de usuários distintos. Os sinais de voz (palavras ou sentenças) são avaliados no espaço n-dimensional, tendo como objetivo é dividir esse espaço em regiões com maior concentração de vetores.

Após o cálculo para delimitar as regiões de codificação, foi realizada a medida de distância entre um vetor código e os demais vetores do espaço vetorial. Esse vetor utilizado para demarcar o espaço o qual foi descrito como de centróide, é armazenado no *codebook*. A figura 3.13 ilustra o espaço vetorial com os centróides definidos pela cor vermelha e os demais vetores de códigos definidos pela cor verde.

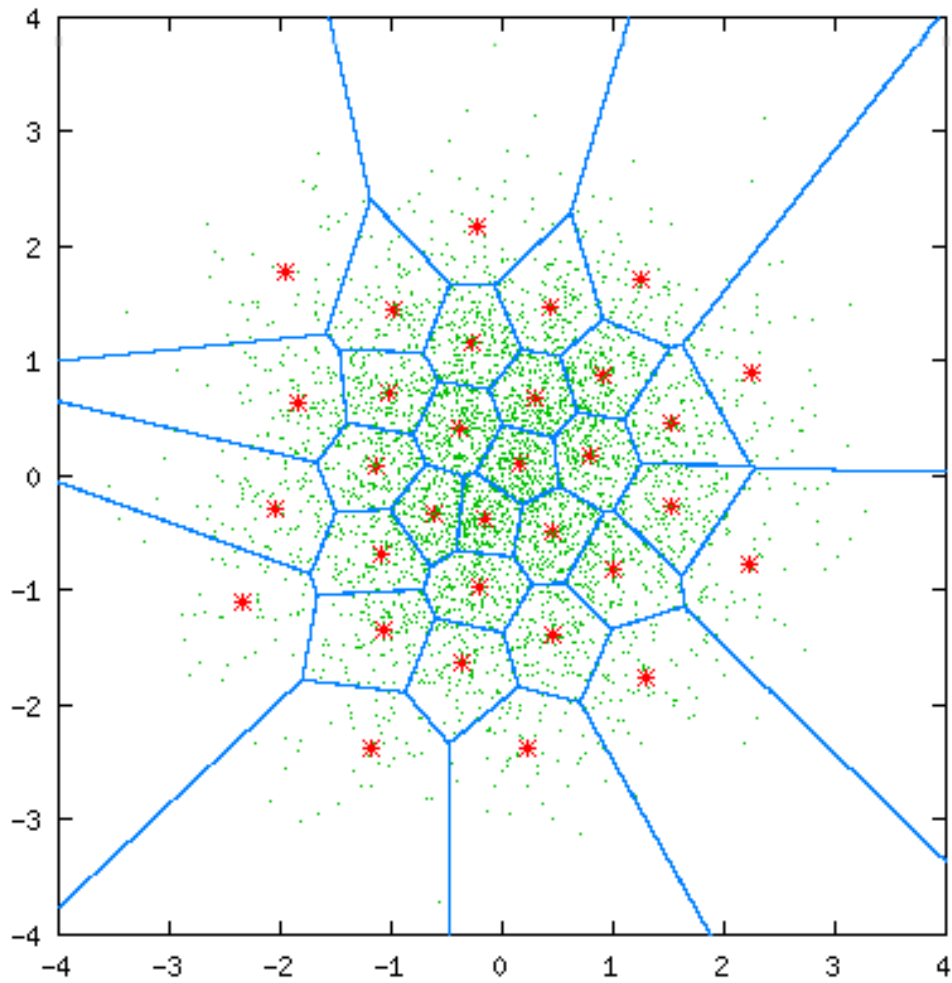


Figura 3-13 – Exemplo de espaço vetorial com centróides de vetores de códigos

Os centróides são os vetores códigos escolhidos, parâmetros armazenados no *codebook*, os quais possuem parâmetros importantes para realizar a quantização vetorial.

4. Análise da qualidade do sinal de voz

Os codificadores de voz buscam ofertar uma melhoria de qualidade em detrimento da taxa de transmissão. Entretanto, existem outros fatores que não são intrínsecos dos codificadores de voz, que são os ofensores externos (*delay*, *jitter*, eco, perda de pacotes e erro no canal de transmissão).

Para minimizar os problemas de qualidade dos codificadores, diversas técnicas são realizadas. Nesse contexto, avaliaremos, principalmente, os fatores intrínsecos aos codificadores de voz, não descrevendo a contribuição, degradação ou melhoria, relacionada aos fatores externos do codificador.

Uma das principais formas de avaliação do sinal de voz, em codificadores de forma de onda, é a relação sinal ruído, medida *SNR*, descrita por:

$$SNR = 10 \times \log_{10} \left[\frac{\sum_{n=0}^{M-1} s(n)^2}{\sum_{n=0}^{M-1} (s(n) - \bar{s}(n))^2} \right] \quad (4.1)$$

A medida SNR tem limitações, pesa todos os erros domínio do tempo de forma igual. Uma medida SNR alta, com resultados indesejáveis, pode ser obtido se o trecho da fala apresenta alta concentração de segmentos de voz (segmentos de alta energia), uma vez que o ruído tem um maior efeito na percepção de segmentos de baixa energia, tais como sons surdos ou fricativos. Uma medida de melhoria da qualidade pode ser obtida se SNR for medida em intervalos de tempo curtos e os resultados em média. Essa medida é chamada relação baseado sinal-ruído segmentada(SNRseg) uma é expressa por:

$$SNR_{seg} = \left[SNR(j) \right] \text{ onde } j = \text{intervalo de tempo da análise SNR} \quad (4.2)$$

Essas análises são ineficientes para as técnicas de codificação paramétrica, pois se faz necessário a avaliação perceptual do sinal de voz. Para análise perceptual, foram criados testes objetivos e subjetivos.

4.1. Medidas subjetivas de qualidade

Inicialmente, os testes em codificadores de voz utilizados para telefonia eram realizados pelo método de avaliação subjetiva, técnica para a medição através da percepção do ouvido humano. Esse tipo de medida necessita seguir algumas diretrizes, tais como:

- analisar um número de ouvintes suficiente para assegurar um resultado estatístico confiável;
- garantir que todos os ouvintes tenham uma percepção auditiva normal;
- garantir que todos os ouvintes efetuem corretamente as respostas dos testes mensurados e tabelados;
- garantir um material abrangente e diversificado. Neste caso, o corpo de dados para os testes, um banco de arquivos de voz;
- garantir que o codificador foi testado em todas as condições;
- escolher adequadamente as condições em que serão realizados os testes.

4.1.1. *Mean Opinion Score (MOS)*

Os testes efetuados segundo a norma especificada na recomendação (ITU P800, 1996) analisam a avaliação perceptual subjetiva e é denominado MOS (*Mean Opinion Score*). Nessa recomendação, os ouvintes utilizam uma escala para medir a qualidade do sinal de voz, o qual é classificado conforme tabela 4.1:

Pontuação	Qualidade da fala	Esforço necessário para a compreensão do significado
5	Excelente	Relaxamento completo; nenhum esforço é necessário
4	Boa	atenção necessária; não é preciso muito esforço
3	Regular	um certo esforço é necessário
2	Pobre	muito esforço é necessário
1	Péssima	Ininteligível, apesar de qualquer esforço empregado

Tabela 4-1 – Escala de classificação do MOS

Com base no processo de avaliação subjetiva², os codificadores são analisados tendo como referência a pontuação MOS. A maioria das medidas de qualidade de voz são baseadas em uma escala de classificação absoluta (ACR – Absolute Category Rating), essa escala é utilizada para verificação do MOS.

Abaixo a figura 4.1 (Gomes, A. G., 2008) apud (Jayant e Noll, 1984) mostra a relação taxa de bits em detrimento à qualidade do sinal de voz amostrado para diferentes tipos de codificadores.

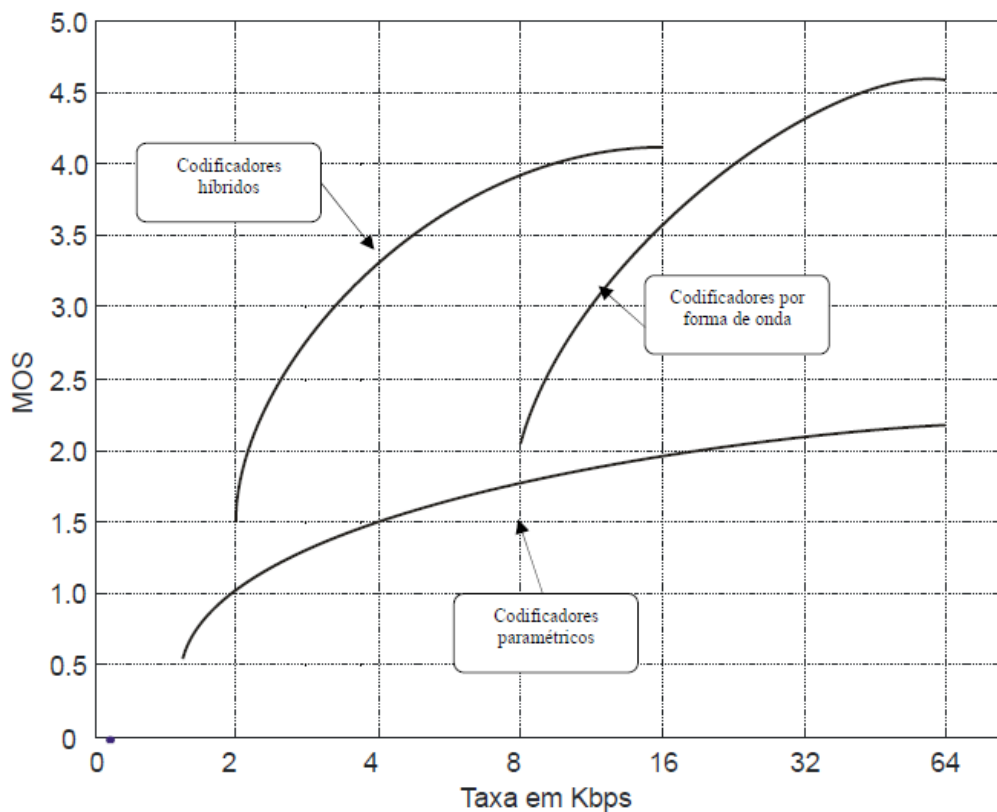


Figura 4-1 – Comportamento da avaliação MOS para dos diferentes tipos de codificadores

4.2. Medidas objetivas de qualidade

Conforme processo apresentado anteriormente, o método subjetivo de avaliação de qualidade demanda tempo. Assim, esse processo é criterioso com os parâmetros e as condições para realização dos testes conforme especificação da recomendação do ITU-T. Contudo, é necessário outro método para análise da qualidade de voz em telefonia em

² Métodos para determinação subjetiva da qualidade de transmissão são descritos na Recomendação ITU-T P.800, essa recomendação descreve os outros dois métodos: *Degradation Category Rating* (DCR) e *Comparison Category Rating* (CCR).

tempo real. Esse tipo de medida é a avaliação objetiva a qual permite a avaliação em tempo real, podendo realizar o controle da qualidade do sinal de voz.

4.2.1. *Perceptual Speech Quality Measure (PSQM)*

O PSQM é um método de medição da qualidade de voz baseado na recomendação do ITU-T (ITU P861,1998). O PSQM (*Objective quality measurement of telephone-band, 300-3400Hz, speech codecs*) estima a qualidade de voz nos codificadores. Esse algoritmo efetua a medida de qualidade por um modelo psicoacústico, o qual tenta reproduzir a qualidade do som percebida pelo ouvido humano.

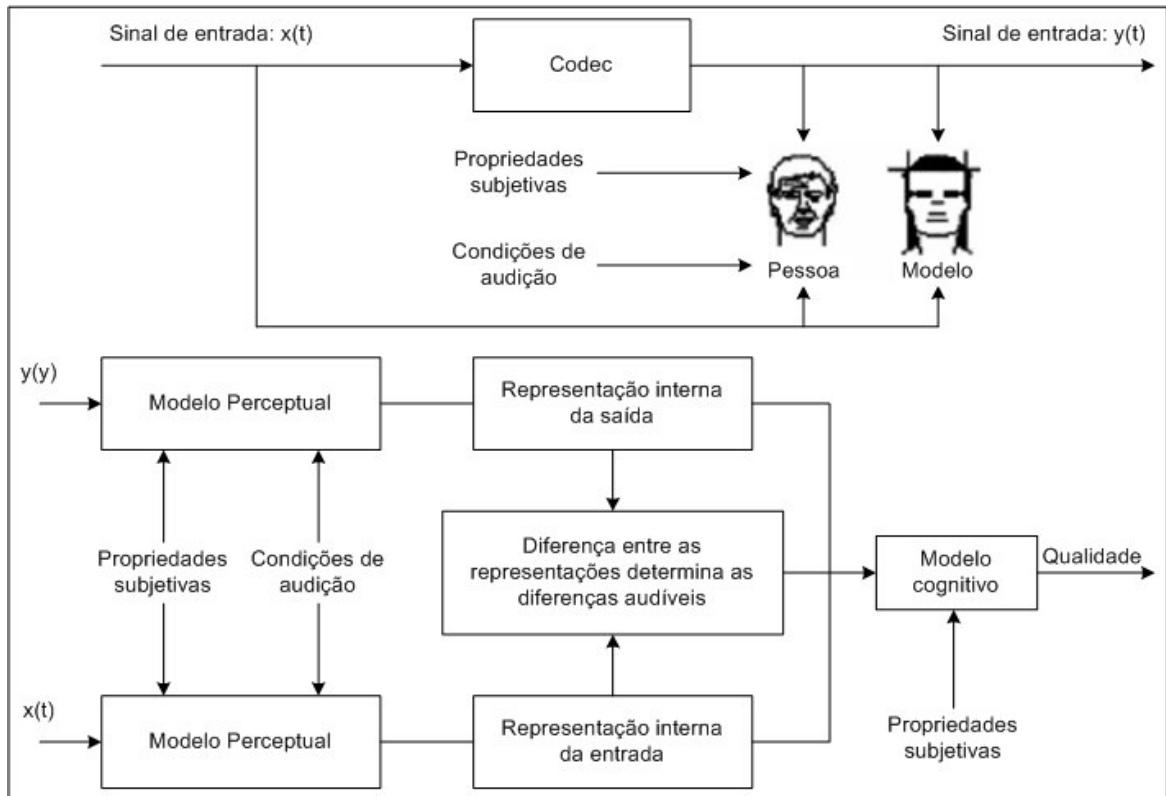


Figura 4-2 – Modelo de avaliação objetiva, utilizando o PSQM.

Na figura 4.2, o modelo analisado tem como base a análise dos sinais de entrada e saída do codificador:

- o sinal é convertido onde ocorre o mapeamento no tempo e em frequência através da implementação da FFT utilizando uma janela de Hamming;

- é realizada uma alteração de escala de frequência, utilizando uma escala psicoacústica chamada escala de Bark ³, que realiza uma análise subjetiva do áudio, cobrindo 24 bandas audíveis em Hertz;

- também é realizada uma alteração da escala de amplitude do sinal, em que efetua ajuste da potência sonora através de filtros de ponderação, cujo objetivo é o ajuste do som percebido pelo ouvido humano.

4.2.2. *Perceptual Analysis Measurement System (PAMS)*

Esse algoritmo foi desenvolvido em 1998, e também efetua uma análise no tempo e em na frequência. O PAMS é um algoritmo complementar ao PSQM, ele resolveu um problema que ocorria no PSQM, a realização de sincronismo no tempo.

4.2.3. *Perceptual Evaluation of Speech Quality (PESQ)*

A análise para este estudo utilizou o algoritmo do PESQ, que é a recomendação do ITU-T (ITU P862, 2001). Foi desenvolvida para utilização de redes de pacotes, é uma evolução dos algoritmos PSQM e PAMS, e apresenta uma medida de qualidade relacionada diretamente com a escala MOS (ITU P800, 1996).

O PESQ possui os mesmos parâmetros e processos do PSQM, entretanto, algumas modificações foram consideradas neste modelo:

- equalização dos ganhos dos sinais de entrada (sinal original) e saída (sinal degradado) do codificador, os dois sinais são analisados no domínio tempo e no domínio frequência, para que ao final da análise tenham o mesmo nível de potência;

- os sinais são filtrados para que tenham a mesma característica, e são alinhados no tempo, para definir o intervalo de análise;

- realiza-se uma conversão para o domínio frequência utilizando uma janela de Hamming com 50% de sobreposição em quadros de 32ms.

³ É uma escala psicoacústica proposta por Eberhard Zwicker em 1961. Ela foi nomeada após Heinrich Barkhausen ter proposto a primeira medição subjetiva de intensidade sonora. A escala varia de 1 à 24, corresponde a 24 bandas críticas de audição. As frequências base da escala Bark de audiometria estão no range de 20Hz à 15500Hz.

Segundo (MAGRO, 2005), em uma análise comparativa entre os algoritmos de medida de qualidade, o PESQ é o modelo de avaliação objetiva, que possui maior exatidão quando realizado testes cujo resultado é correlacionado com a medida subjetiva MOS, conforme tabela 4.2 .

Tipo	Coefficiente Correlação	PESQ	PSQM
Rede Móvel	média	0,962	0,924
	piores caso	0,905	0,843
Rede Fixa	média	0,942	0,881
	piores caso	0,902	0,657
VoIP	média	0,918	0,674
	piores caso	0,810	0,260

Tabela 4-2 – Comparação de eficiência do Modelo PSQM & PESQ.

Dentre os diversos codificadores de voz analisados em telecomunicações, abaixo segue um quadro comparativo com valores orientativos, segundo (KONDOZ, 2003). Essas medidas de qualidade foram pontuadas através de testes subjetivos pontuais, usando material de ensaio de literaturas variadas. Esses valores são úteis para um parâmetro entretanto não deve ser tomado como uma indicação definitiva de desempenho do codec.

Recomendação	Ano	Algoritmo	Taxa (kbits/seg)	MOS
G.711	1992	PCM	64	4.3
G.726	1991	VBR-ADPCM	16/24/32/40	.-x-
G.728	1994	LD-CELP	16	4
G.723.1	1995	A/MP-MLQ CELP	5.3/6.3	.-x-
G.729	1995	CS-ACELP	8	4

Tabela 4-3 – Valores de referência do MOS para os codificadores *Standards* utilizados em telefonia.

5. Codificadores ITU-T

Neste capítulo, é realizada uma análise dos codificadores utilizados comumente em redes de telefonia. Em geral os sistemas de codificação de voz possuem o seguinte diagrama, segundo (CHU, 2003).

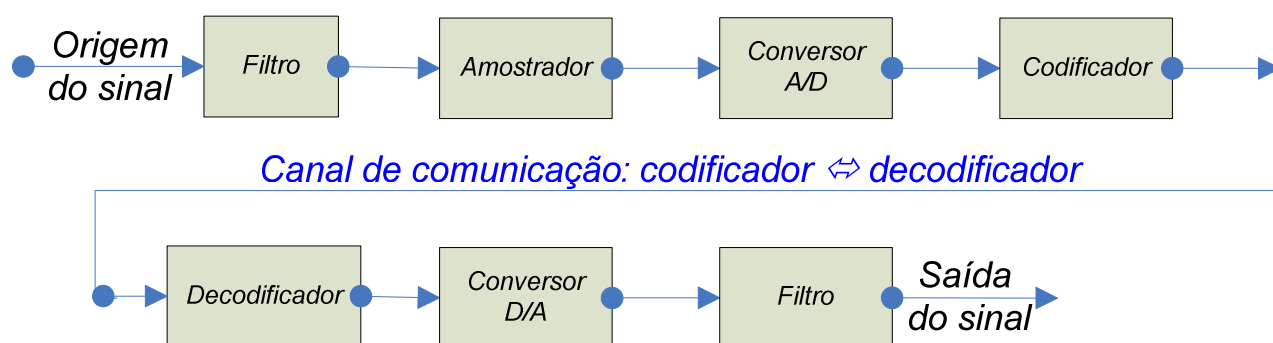


Figura 5-1 – Sistema de codificação de voz.

5.1. Codificadores ITU-T utilizados em telefonia

Abaixo segue uma breve descrição dos codificadores padronizados pelo ITU-T. Os mais relevantes descritos segundo (CHU, 2003), utilizados em redes de pacotes, estrutura atual das operadoras de telecomunicações em substituição às redes comutadas por circuito.

Codificador de voz	Taxa (kbits/seg)	VAD	Redução de ruído	Atraso (ms)	Ano
G.711 (A/μ-Law PCM)	64	não	não	0	1972
G.726 (ADPCM)	16/24/32/40	não	não	0,25	1990
G.728 (LD-CELP)	16	não	não	1,25	1992
G.723.1	5.3/6.3	sim	não	67,5	1995
G.729 (CS-ACELP)	8	sim	não	25	1996

Tabela 5-1 – Comparação de *codecs* ITU-T para telefonia.

5.1.1. Recomendação ITU-T G.711

Ano de aprovação: 1972

Taxa de codificação: 64kbit/s, pois temos uma frequência de amostragem de 8000 Hz e quantizamos 8bits por amostras.

Algoritmo de codificação: PCM - *Pulse Code Modulation*

Princípio do codificador: Realiza a quantização com escala logarítmica. Desta forma, obtém a relação sinal ruído (SNR) independente da intensidade. A técnica de codificação PCM obedece ao critério de um codificador Nyquist, em que a frequência de amostragem é igual ou superior ao dobro da maior frequência presente no espectro:

$$f_{amostragem} = 2 \times f_{máxima} \quad (5.1)$$

A codificação em telefonia utiliza uma frequência de amostragem de 8kHz (8000 amostras do sinal de voz a cada segundo), as amostras são representadas com 256 níveis de quantização, o que define 8 bits por amostra. Sendo assim, o sinal de voz com a codificação PCM possui a taxa padrão 64k bits/seg.

Nesta técnica de codificação são utilizados dois algoritmos de quantização que são chamados de leis: lei a (alaw) a lei μ (μ law), pois o processo da técnica PCM é atribuir um valor discreto a amplitude do sinal amostrado. Esse processo gera ruído de quantização, sendo necessário utilizar uma das leis de quantização para minimizar o sinal de entrada no quantizador com o sinal de saída quantizado.

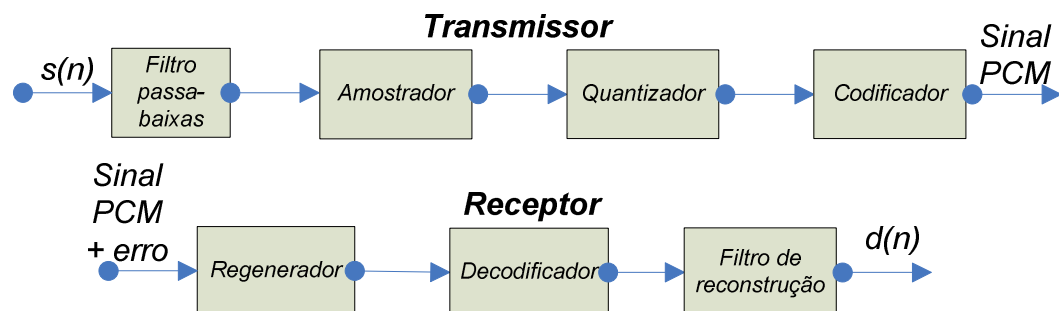


Figura 5-2 – Diagrama de blocos da técnica PCM

5.1.2. Recomendação ITU-T G.726

Ano de aprovação: 1990

Taxa de codificação: variável 16kbit/s, 24kbit/s, 32kbit/s e 40kbit/s

Algoritmo de codificação: ADPCM – *Adaptive Pulse Code Modulation*

Princípio do codificador: Este codificador foi desenvolvido com a proposta de melhorar a codificação PCM, codificar a voz com a metade da taxa utilizada no PCM, mantendo a qualidade do sinal codificado. Essa técnica de codificação possui um quantizador adaptativo, podendo fazer um ajuste no preditor linear com base nas variações do sinal à ser codificado. As diferentes taxas de codificação são relativas aos bits utilizados por amostra: 2,3,4 e 5, tendo respectivamente as taxas: 16kbit/s, 24kbit/s, 32kbit/s e 40kbit/s.

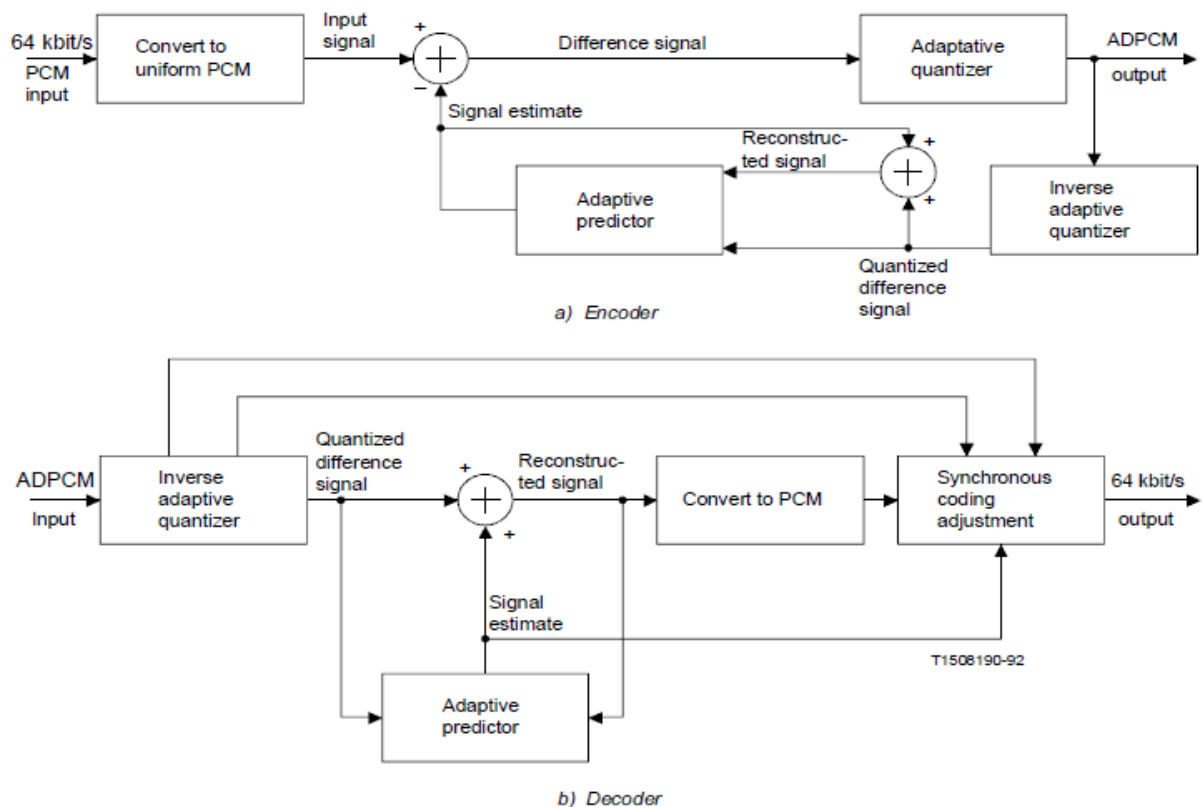


Figura 5-3 – Diagrama de blocos do processo de codificação ADPCM (ITU-T G726) apud (VARY e MARTIN, 2006)

5.1.3. Recomendação ITU-T G.728

Ano de aprovação: 1992

Taxa de codificação: 16kbit/s

Algoritmo de codificação: LD-CELP – *Low-Delay Code Excited Linear Prediction*

Princípio do codificador: Este codificador reduz o tempo da amostra processada entre codificador e o decodificador. Transmite uma única excitação e possui uma análise de predição capaz de avaliar recursivamente cinco amostras PCM, sendo essa rotina a base do algoritmo implementado para melhoria do atraso de codificação.

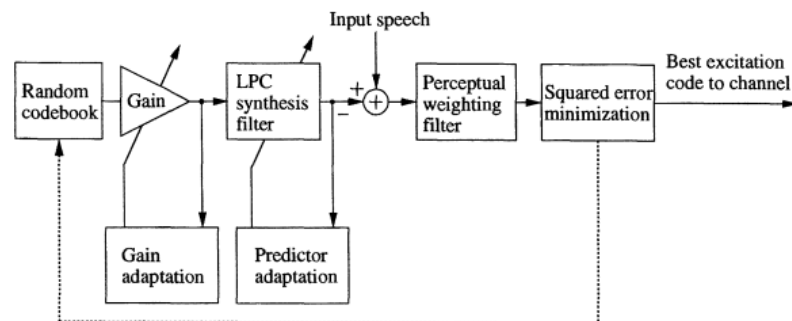


Figura 5-4 – Diagrama de blocos do processo de codificação LD-CELP (FURUI, 2001)

5.1.4. Recomendação ITU-T G.723.1

Ano de aprovação: 2006

Taxa de codificação: 5,3kbit/s e 6,3kbit/s

Algoritmo de codificação: MP-MLQ ACELP,

onde MP-MLQ-*Multi-Pulse Maximum Likelihood Quantization*

Princípio do codificador: Codifica o sinal de voz mediante codificação linear de análise por síntese. Foi projetado para videoconferência e voz sobre IP. É um codificador de taxa dupla em que codifica à taxa de 5.3 kbps utilizando a estrutura ACELP. A codificação à taxa de 6.3kbps corresponde a codificação MP-MLQ que oferece uma qualidade um pouco melhor.

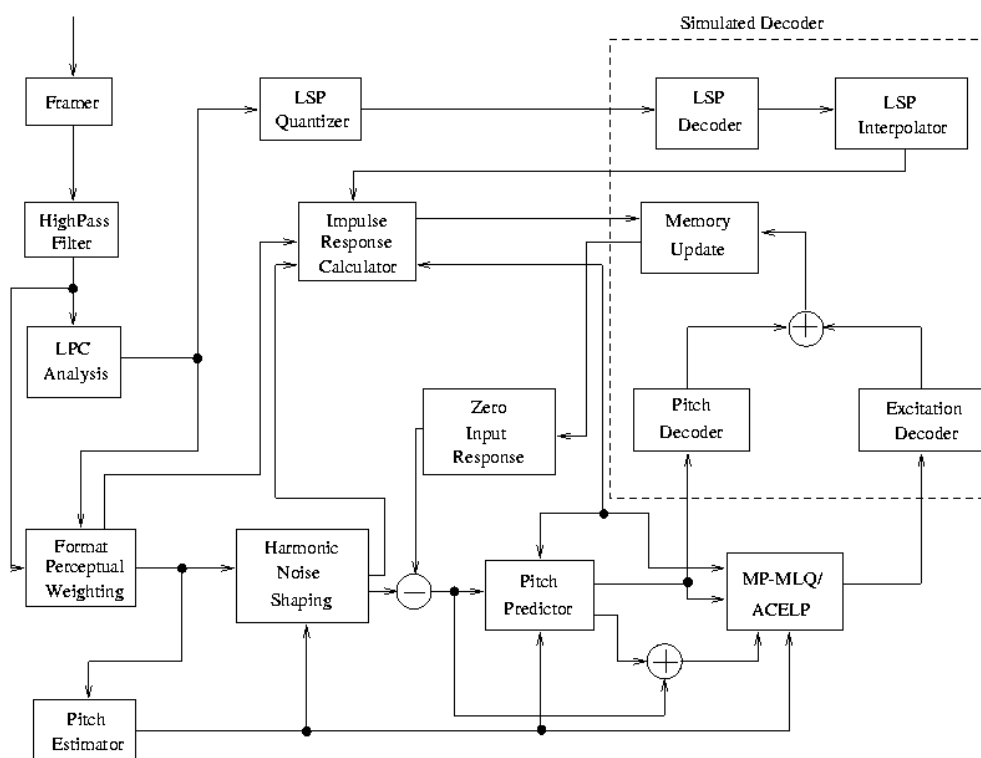


Figura 5-5 – Diagrama de blocos do processo de codificação LD-CELP

5.1.5. **Recomendação ITU-T G.729**

Ano de aprovação: 1996

Taxa de codificação: 8kbit/s

Codificação: CS-ACELP – *ConjugateStructure- Algebraic Code excited
Linear Prediction*

Princípio do codificador: É a transmissão do sinal de voz com uma baixa taxa de bits, para uso de telefonia. É muito utilizado em redes que necessitam compressão de banda. Realiza a codificação do sinal em frames de 10ms, efetua uma análise das amostras futuras de 5ms, o que resulta num atraso de 15ms para realizar o algoritmo de codificação.

No item 5.2 faremos uma descrição detalhada do bloco de codificação do sinal de voz, pois este codificador é a caso de estudo para realizar a avaliação perceptual do sinal de voz.

5.2. CS-ACELP: Conjugate Structure Code-Excited Linear Prediction (G.729a)

O codificador G.729a baseia-se na filtragem do sinal analógico especificado na recomendação G712 do ITU-T com a taxa de amostragem a 8000 amostras por segundo, codificando os sinais de áudio em frames de 10ms. Possui um atraso de 5ms, sendo seu bit rate de 8Kbps

5.2.1. O codificador CS-ACELP - (G.729a)

. O modelo de codificação é baseado algoritmo de codificação CS-ACELP. Efetua a predição linear por excitação com código algébrico. Nesse intervalo de tempo, o sinal de voz (quase estacionário) é analisado e parametrizado, comparado através de uma análise por síntese em busca da melhor excitação do sinal que corresponda ao sinal alvo. Cada frame de 10ms que é analisado, gera 80 amostras do sinal por segundo, em que são extraídos os parâmetros por frame: Coeficientes de predição linear do filtro, índices do *codebook* fixo, índices do *codebook* adaptativo e ganho. Esses parâmetros codificados e transmitidos são distribuídos de acordo com a tabela 5.2, a qual explica a alocação de bit dos 8kb/s do algoritmo CS-ACELP em um frame em 10ms:

Parameter	Codeword	Sub frame 1	Sub frame 2	Total per frame
Line spectrum pairs	L0, L1, L2, L3			18
Adaptive-codebook delay	P1, P2	8	5	13
Pitch-delay parity	P0	1		1
Fixed-codebook index	C1, C2	13	13	23
Fixed-codebook sign	S1, S2	4	4	8
Codebook gains (stage 1)	GA1, GA2	3	3	6
Codebook gains (stage 2)	GB1, GB2	4	4	8
Total				80

Tabela 5-2 – Parâmetros codificados pelo G729a (ITU-T G729).

5.2.2. Codificador

As principais etapas de codificação do sinal de voz são executadas conforme o diagrama de blocos abaixo:

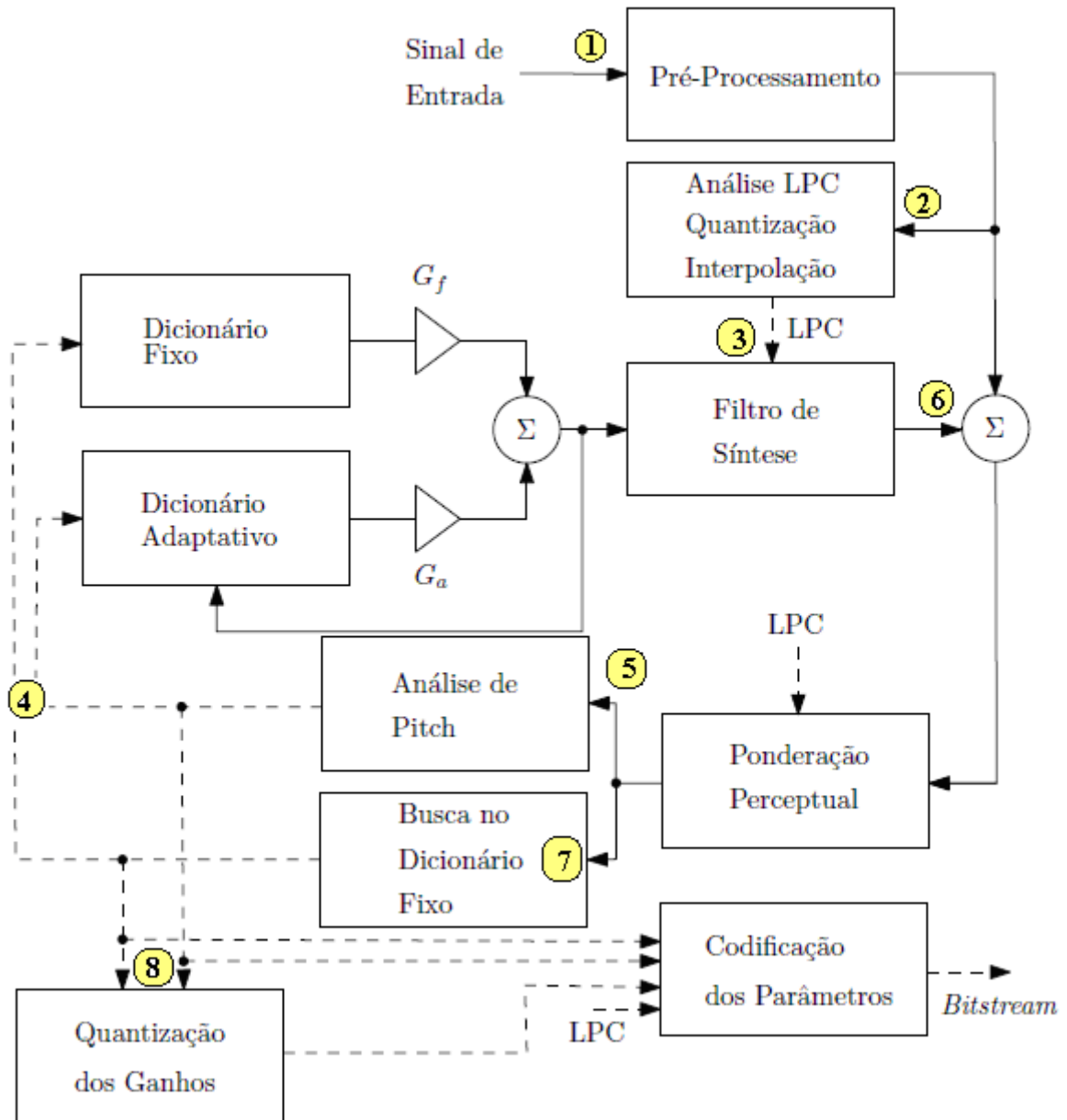


Figura 5-6 – Diagrama de blocos do processo de codificação CS-ACELP

Descrição dos estágios de codificação do CS-ACELP:

1) O sinal de entrada passa por um filtro passa alta para eliminação de ruído, é pré-processado e esse sinal é utilizado para as análises subsequentes da codificação.

2) A análise de Predição Linear é realizada uma vez a cada *frame* de 10ms calculando-se os coeficientes de filtro de Predição Linear. Esses coeficientes são convertidos para *Line Spectrum Pairs* (LSP) e quantizados, *Vector Quantization* (VQ).

3) O sinal de excitação é escolhido por meio de uma análise por síntese, procedimento de busca em que o erro entre o sinal de voz original e o reconstruído é minimizado de acordo com uma medida de distorção perceptualmente ponderada. Isso é feito através da filtragem do sinal de erro com um filtro de ponderação perceptual, cujos coeficientes são derivados da Predição Linear não quantizada do filtro.

4) Os parâmetros de excitação (parâmetros dos dicionários fixos e adaptativo) são determinados por um *subframe* de 5ms (40 amostras) cada um. Os coeficientes quantizados e não quantizados do filtro de Predição Linear são usados no segundo *subframe*, enquanto no primeiro *subframe* os coeficientes de predição linear interpolados são usados (ambos quantizados e não quantizados).

5) Um valor do *pitch* é estimado com base no sinal resultante da amostra analisada pelo filtro de ponderação, em malha aberta.

6) Os estados iniciais do filtro são atualizados pela filtragem do erro residual. Calcula-se o circuito em malha fechada para estimar o ganho e o valor do dicionário adaptativo.

7) Com os valores do dicionário adaptativo e com as duas estimativas de *pitch*, busca-se no dicionário fixo um valor para encontrar a solução ótima.

8) Os ganhos dos dicionários (fixo e adaptativo) são quantizados, o filtro é atualizado e o sinal de excitação é determinado.

5.2.3. Decodificador

Na decodificação do sinal, os parâmetros do fluxo de bits recebidos são extraídos conforme diagrama de blocos abaixo:

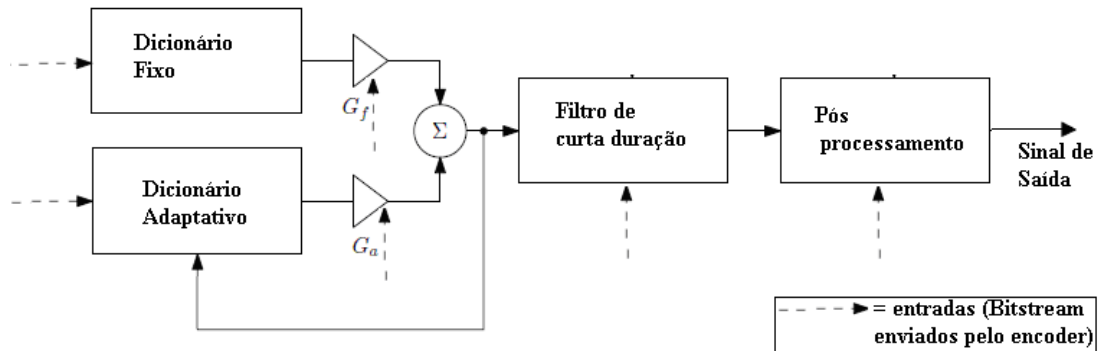


Figura 5-7 – Diagrama de blocos do processo de decodificação CS-ACELP

Descrição dos estágios de decodificação do CS-ACELP:

1) Os índices são decodificados para obter os parâmetros do codificador correspondente a um quadro de fala de 10 ms.

2) Esses parâmetros são os coeficientes LSP, e duas frações de *delay* do *pitch*, dois vetores do *codebook* fixo, e os dois conjuntos de ganhos referente ao *codebook* adaptativo e o *codebook* fixo;

3) Efetua-se a síntese de curto prazo: Os coeficientes LSP são interpolados e convertidos em coeficientes do filtro de predição linear para cada sub frame.

4) Efetua-se a síntese de longo prazo: É avaliado um subframe de 5ms, em que são executadas as seguintes etapas:

- a) A excitação é calculada adicionando os vetores do *codebook* fixo e do *codebook* adaptativo, dimensionado pelos respectivos ganhos
- b) O sinal de fala é reconstruído através da filtragem da excitação, realizada pelo filtro de síntese de predição linear.

5) A reconstrução do sinal de fala ocorre após o estágio de pós filtragem, que inclui um filtro adaptativo baseado nos filtros de síntese a curto e longo prazos, seguido por um filtro passa-alta que completa a operação de filtragem e dimensionamento da operação.

6. Proposta

Neste trabalho, foi avaliada a estrutura do codificador com suas alterações propostas para melhoria perceptual, excluindo-se as variações do meio de transmissão.

Os codificadores de voz buscam a melhoria de qualidade utilizando uma baixa taxa de bits para transmissão. Um exemplo disso é o codificador G.729a, padrão do ITU-T que consegue operar a uma taxa de 8Kbps. Geralmente, se a análise baseia-se nessa premissa (maior qualidade com menor taxa de transmissão), o processo para a garantia de qualidade e a decodificação do sinal possuem mecanismos incorporados para tratar os fenômenos inerentes ao meio de transmissão (atraso, jitter e perda de pacotes).

Segundo avaliação da figura 6.1, apresentada abaixo, que se baseia no estudo do *codec* G729a apresentado em (PREGO e NETTO, 2008), os valores de MOS estão dentro do limiar aceitável para experimentos acadêmicos e aplicações práticas implementadas no mercado.

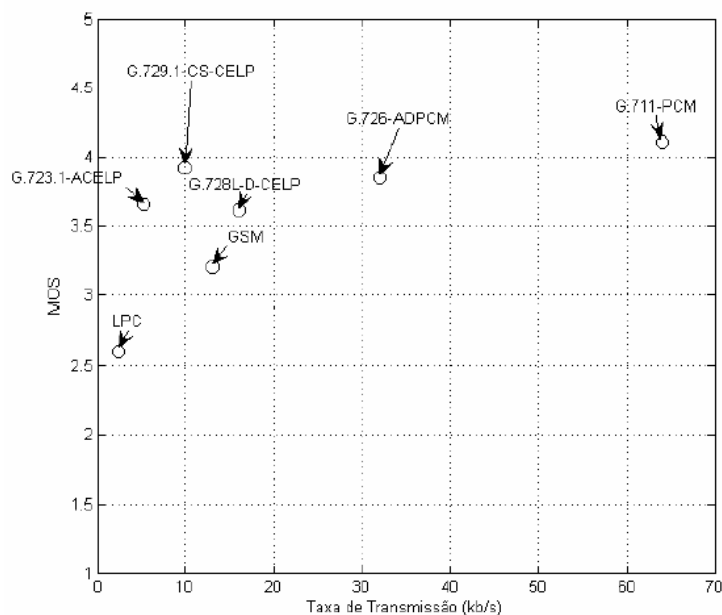


Figura 6-1 – Relação do MOS x Taxa de Transferência dos diversos codificadores

Como a análise neste estudo enfoca a melhoria perceptual no sinal de origem, optou-se pela escolha de um codificador que fosse bastante utilizado nos sistemas de telecomunicações, o *codec* G729a.

Convém mencionar que a avaliação perceptual foi realizada em trabalhos que analisavam o desempenho do codificador mediante a ponderação de qualidade sobre o meio de transmissão.

A proposta é obter uma melhora na ausência do meio de transmissão, o que é factível de ser avaliado, uma vez que as características deste codificador já foram objeto de estudo em diversos trabalhos na comunidade acadêmica e no meio corporativo. (MAGRO, 2005), (FERNANDES, 2003)

Neste estudo, foi aplicado o processo proposto, avaliação do sinal na origem do codificador com a implementação do algoritmo do PESQ no *encoder*.

Para verificação do estudo, foi implementado a recomendação do ITU-T, P.862 (PESQ), no *encoder* do codificador CS-ACELP (G729a), com a finalidade de mensurar uma ponderação do MOS, pois esta métrica é a atual referência para avaliação de qualidade do sinal de áudio percebido nos sistemas de telecomunicações.

A proposta deste trabalho realiza uma análise objetiva baseada em critérios perceptuais do sinal de voz, sendo este verificado e analisado no bloco de origem do sinal a ser transmitido.

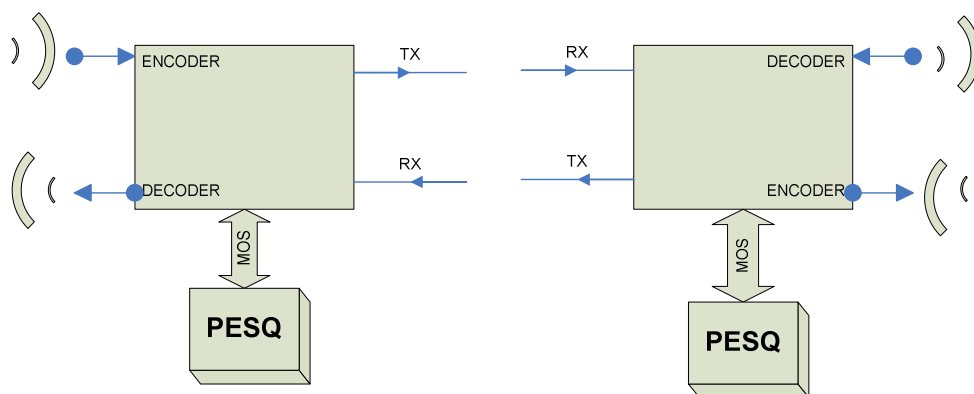


Figura 6-2 - Diagrama de avaliação perceptual do codec no sinal de origem

A produção do sinal de voz tem suas características variantes no tempo, que dependem da variação do trato vocal. Pensando do ponto de vista do orador, se conseguíssemos realizar uma análise das formantes do sinal de voz, mapeando alguns parâmetros no processo de codificação, seríamos capazes de indicar alterações dinâmicas para melhorar esse sinal de voz emitido.

Sendo constantes os avanços nos hardwares de processamento digital de sinais, entende-se que a capacidade de processamento não é mais o problema, mas a busca por uma melhoria contínua no sinal de voz é um tema ainda em discussão. Neste cenário, o tema proposto, melhoria da qualidade perceptual do sinal de voz na origem do codificador, pode ter uma expressiva contribuição na busca por “melhoria contínua”.

Na proposta apresentada, é realizada uma heurística em que o controle perceptual do sinal de voz é realizado em tempo real. Utilizaremos o codificador G729a e o algoritmo do PESQ para avaliação do sinal de voz.

A análise foi implementada no bloco do codificador de voz. Neste bloco, foi realizada a codificação, decodificação e análise perceptual do sinal de voz amostrado. Em cada instante desta análise deseja-se obter uma melhoria no sinal codificado. Essa melhoria foi avaliada com o algoritmo do PESQ interagindo no processo de codificação do sinal de voz.

A figura 6.3 abaixo mostra o diagrama do bloco de origem do codificador que efetua a análise do processo de avaliação perceptual durante a transmissão dos dados:

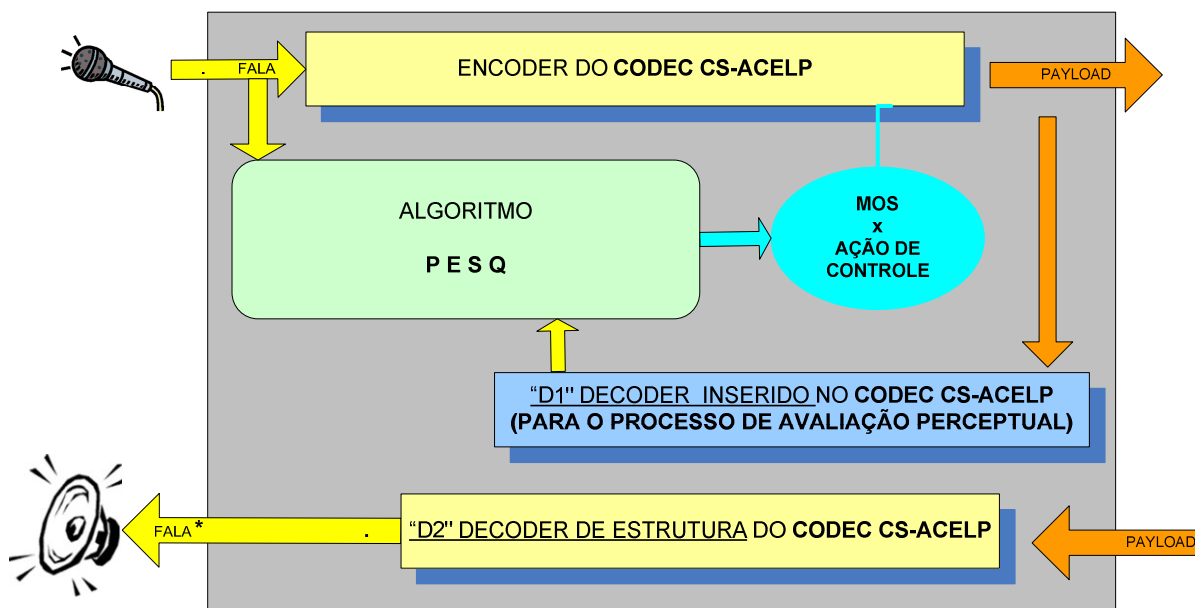


Figura 6-3 - Diagrama de blocos simplificado de análise do codec

O objeto de estudo é o codificador G729a, em que foram implementados, nesse codificador, os blocos relacionados na figura 6.3 conforme descrição abaixo:

- o bloco do *encoder* realiza a codificação do sinal, em que foi realizada a segmentação para leitura de um número específico de *frames* do sinal de entrada;
- o bloco do *decoder* foi duplicado na estrutura. O bloco “D2” é para receber o sinal de voz que necessita ser decodificado, estrutura normal do codificador (*encoder / decoder*). O bloco “D1” *decoder* inserido na estrutura do codificador é responsável pela decodificação dos *frames* codificados originados no *encoder*. Esses *frames* necessitam ser decodificados para avaliação perceptual do sinal;
- o bloco do PESQ é responsável por receber o sinal de referência da entrada do codificador e o sinal que passou pelo *encoder* e foi codificado para envio e transmissão. Esse mesmo sinal passa por um *decoder* “D1” e é decodificado para avaliação perceptual;
- a indicação de MOS, ou seja, a pontuação dos sinais avaliados incide sobre uma ação de controle no bloco do *encoder*, alterando os frames seguintes à avaliação presente.

Os blocos descritos acima, estruturados nesta formatação, compõem o estudo de caso e análise desta dissertação. O principal objetivo deste trabalho é verificar o sinal do ponto de vista perceptual, essa verificação ocorre em tempo real com ações de controle no bloco de origem de codificação, o *encoder*, para garantir e melhorar perceptual os frames codificados.

7. Estudo de Caso

Para análise do problema proposto, avaliação perceptual do sinal de voz na origem da codificação para obter melhoria perceptual, foram estudadas a codificação e a avaliação do sinal codificado. Neste trabalho utilizaram-se os programas:

- 1) ITU-T G.729a *CS-ACELP Speech Coder* *ANSI-C Source Code*
Version 1.1 *Last modified: September 1996*
Copyright (c): AT&T, France Telecom, NTT, Universite de Sherbrooke
- 2) ITU-T P862 *PESQ Algorithm Software* *ANSI-C Source Code*
Version 1.2 *Last modified: August 2002*
Copyright (c): Psytechnics Limited e OPTICOM GmbH

7.1. Plataforma de experimento (G729a com PESQ)

O processo de análise consiste no bloco do codificador G.279a (codificação/decodificação) na origem do interlocutor, sobre o qual foi inserido o algoritmo do PESQ para avaliação perceptual do sinal de voz na origem da codificação.

Nessa formatação, o processo de codificação (extrai os coeficientes LPC, o ganho e os índices dos dicionários) encaminha o fluxo de bits ao decodificador que foi duplicado no bloco do codificador G729.^{aç} O sinal de entrada, ao mesmo tempo em que é codificado na origem, tem seu valor decodificado e avaliado pelo algoritmo do PESQ.

Os arquivos PCM 16Bits, amostrado a 8KHz de entrada e de saída do bloco do codificador de origem do interlocutor, são verificados a cada intervalo de tempo. Obtêm uma

resposta do algoritmo PESQ para uma análise curta de frames, de modo que a representação seja a mais próxima do tempo real.

7.2. Análise dos parâmetros de avaliação

Durante os estudos, foram avaliados os parâmetros escalares de ganho de codificação e ganho de *pitch* do codificador G729a. Também se avaliou o algoritmo PESQ, que foram verificados parâmetros intrínsecos do algoritmo. Essa análise dos parâmetros definida no item 7.3 descreve o processo de análise dos arquivos codificados e decodificados no bloco do *encoder* para obter uma garantia de melhora perceptual dos *frames* do sinal de voz.

7.2.1. Limites de frames do PESQ no bloco do codec

Para a avaliação perceptual na implementação do algoritmo do PESQ dentro do *encoder* do codificador do G729a foi testado e validado o padrão mínimo de fluxo de bits que o PESQ é capaz de realizar a pontuação MOS.

Pela recomendação P862, o valor correspondente para análise do sinal de voz no código fonte do algoritmo PESQ é de um intervalo 100 frames de 10ms cada. Cada frame possui 80 amostras de 2 bits, totalizando 16000 bits($100\text{frames} * 2\text{bits} * 80\text{amostras}$). Para isso, a avaliação pretende reproduzir uma análise mais próxima do tempo real, o fluxo de bits mínimo encontrado para análise do PESQ é de 28 frames do PESQ ($28\text{frames} * 2\text{bits} * 80\text{amostras} = 4480\text{bits}$ por intervalo amostrado), sendo este o valor alterado no algoritmo do PESQ.

Com essa alteração, conseguimos uma análise em um período inferior a 1/3 do valor mencionado na recomendação. Constatou-se que, durante o experimento, o intervalo de 28 frames avaliados pelo PESQ foi o resultado máximo em que o algoritmo conseguiu realizar sua função sem que o atraso de conversação e/ou intervalos de expressão ativa(fala) com mais de 25% de silêncio influenciasse no resultado. Assim, a escolha do número de *frames* avaliados pelo PESQ em cada fragmento teve como premissa os períodos de fala ativa superior a 75% do total do *frame*.

7.2.2. Formatação dos arquivos de testes

Para o experimento do algoritmo PESQ alterado e implementado para análise da melhoria perceptual, utilizou-se os arquivos extraídos do banco de dados *Timit Acoustic-Phonetic Contínuos Speech Corpus*, o qual foi projetado para fornecer dados de fala utilizados em reconhecimento de voz. São gravações de banda larga originadas por 630 oradores dos oito principais dialetos do inglês americano. São, ainda, fonéticas escritas e alinhadas ortograficamente em arquivos de 16bits e 16kHz.

Os parâmetros de escolha das sentenças foram as diferenças de oradores masculinos e femininos e suas variações de sons, gerando arquivos de 16bits amostrados à 8kHz. A orientação para formatação dos arquivos de testes foi baseada na recomendação P.830 (ITU P830, 1996) que solicita, no mínimo, dois oradores masculinos e dois oradores femininos para cada teste de condição.

Foram escolhidas 24 frases manipuladas pelo software *Audacity* (Audacity1.3.12,2010), formando quatro arquivos de seis frases e dois arquivos para cada orador (masculino e feminino). O objetivo de estruturar os quatro arquivos foi para avaliar a diferença entre oradores e a variação da frequência de *pitch*. Essa análise pode ser observada nas figuras abaixo, nas ilustrações 7.1, 7.3, 7.5 e 7.7 que mostram os arquivos de teste com oradores distintos. As ilustrações 7.2, 7.4, 7.6 e 7.8 mostram frequência de *pitch*, primeira formante do espectro, conforme abaixo:

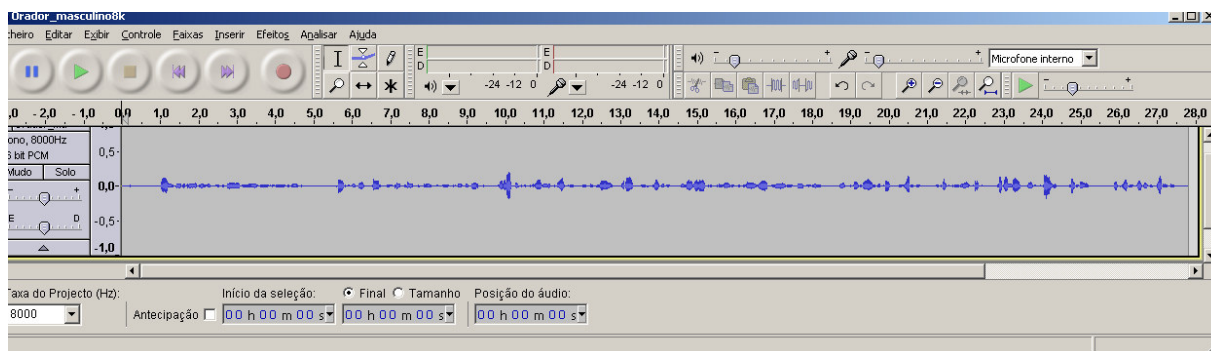


Figura 7-1 Arquivo01 Oradores masculinos – Sinal de voz

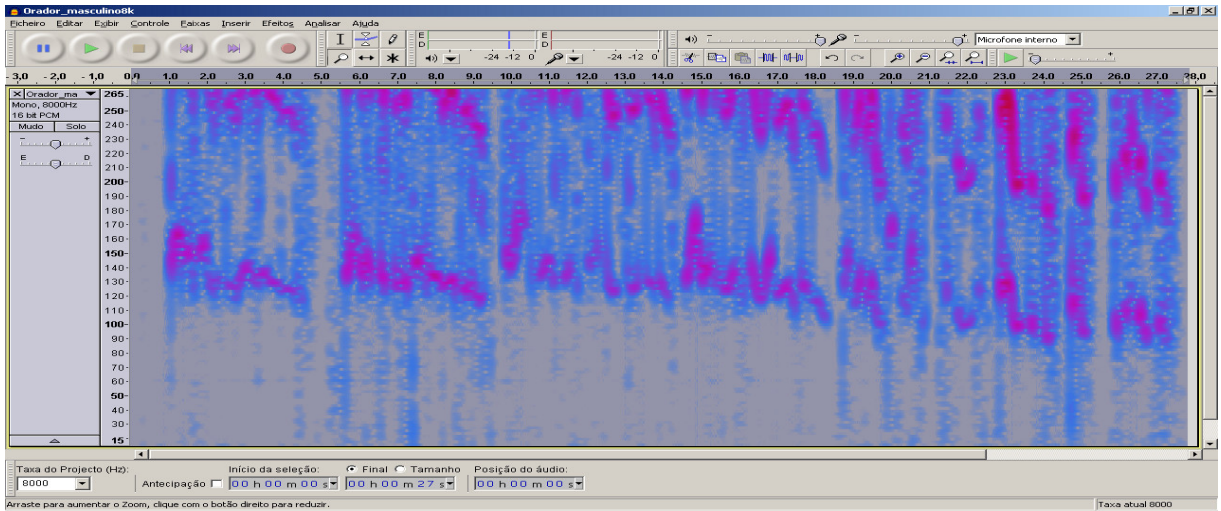


Figura 7-2 Arquivo01 Oradores masculinos – Análise de frequência de pitch

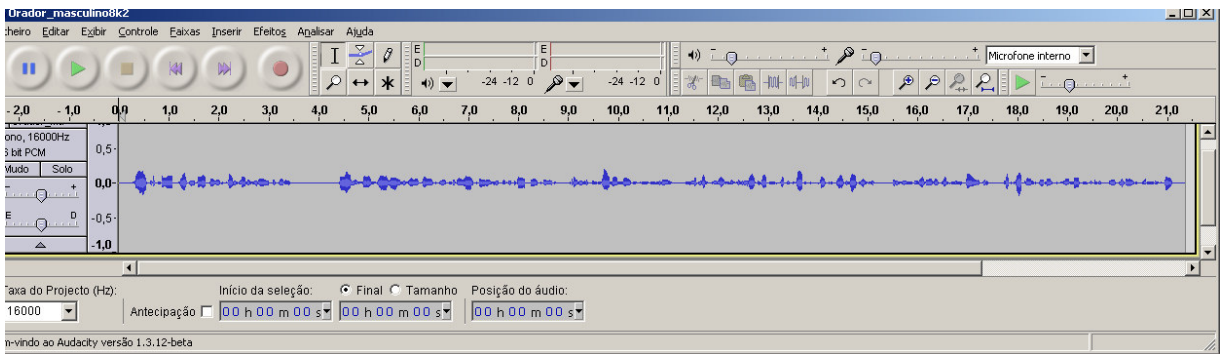


Figura 7-3 Arquivo02 Oradores masculinos – Sinal de voz

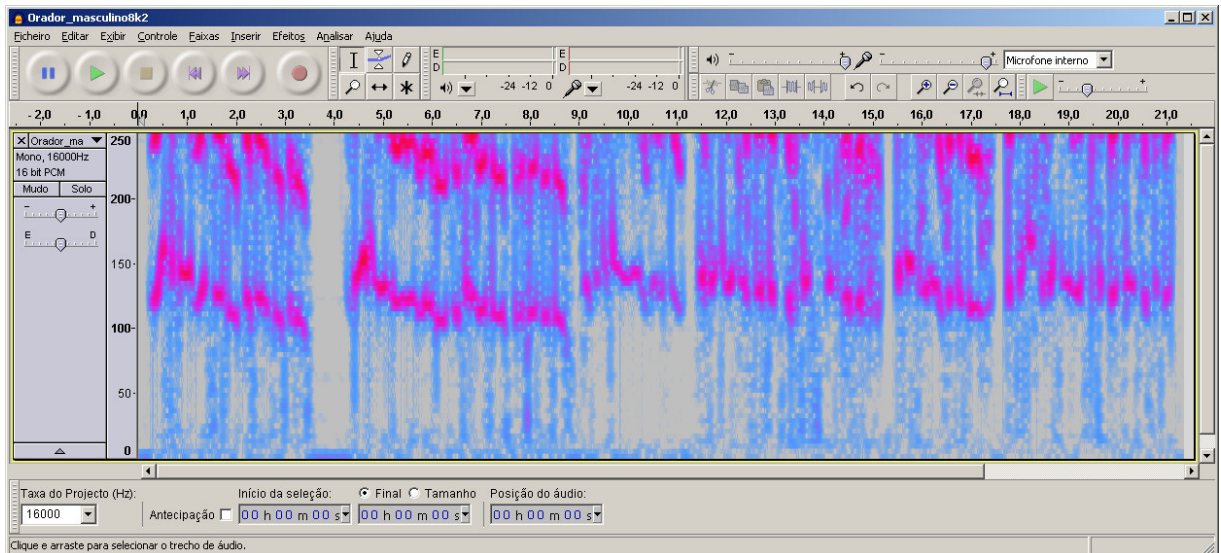


Figura 7-4 Arquivo02 Oradores masculinos – Análise de frequência de pitch

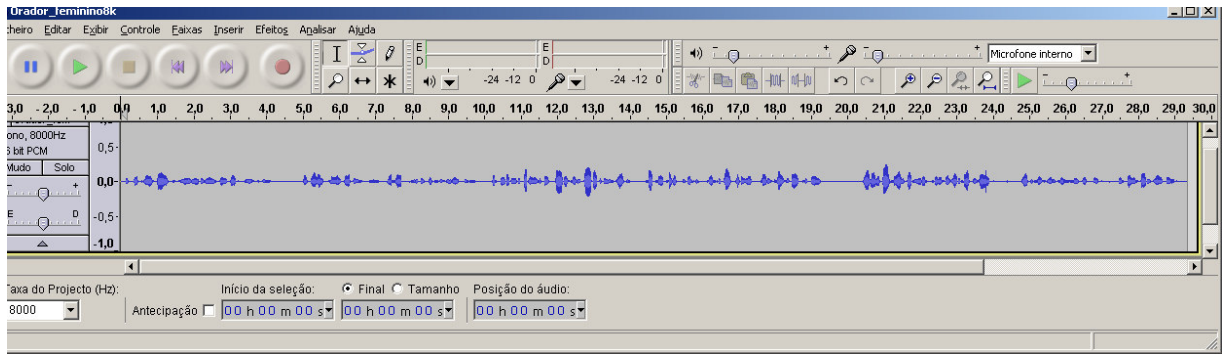


Figura 7-5 Arquivo01 Oradores femininos – Sinal de voz

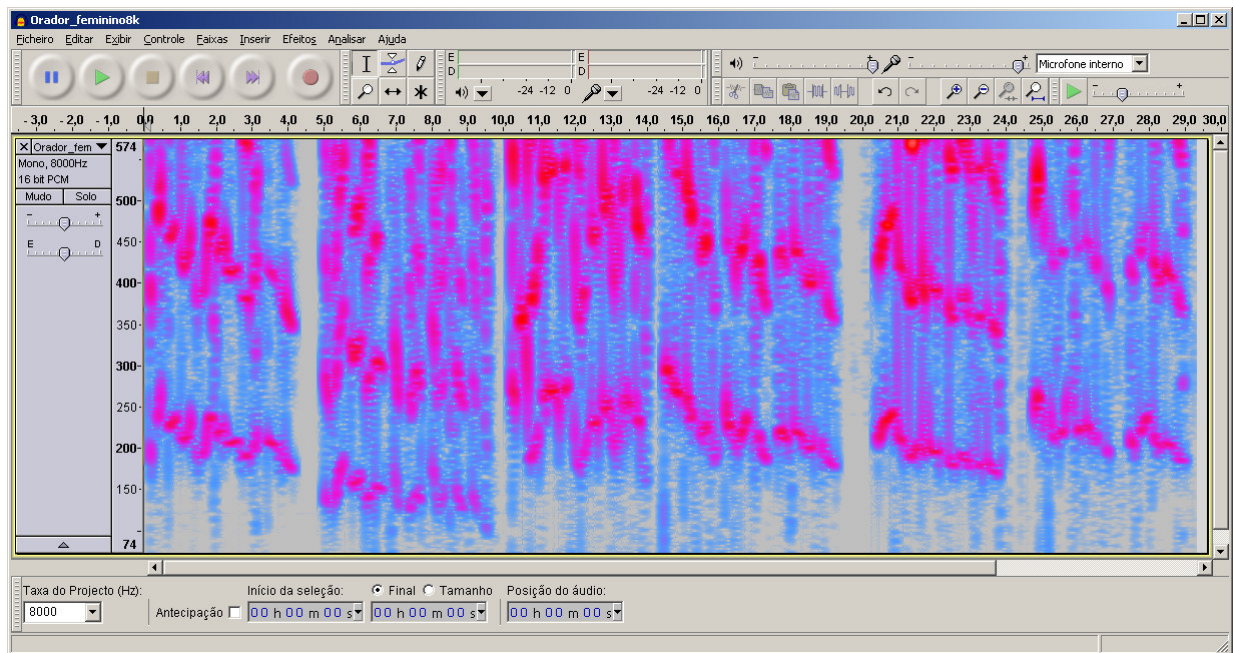


Figura 7-6 Arquivo01 Oradores femininos – Análise de frequência de pitch

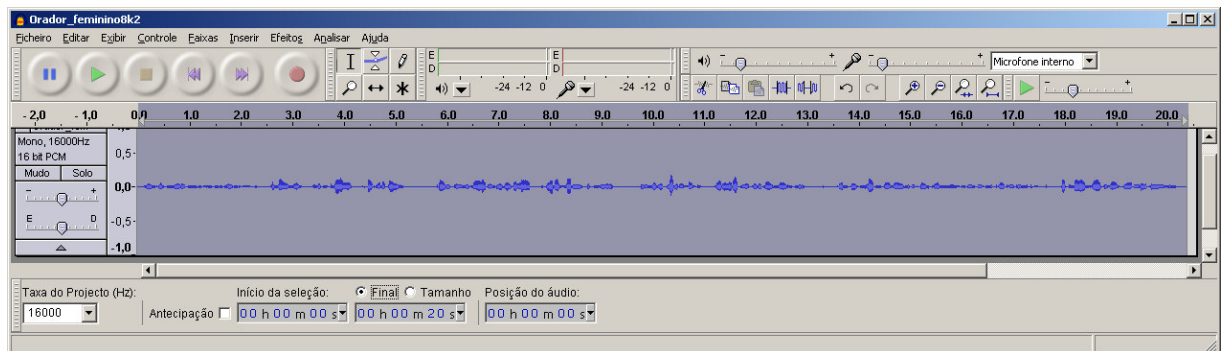


Figura 7-7 Arquivo02 Oradores femininos – Sinal de voz

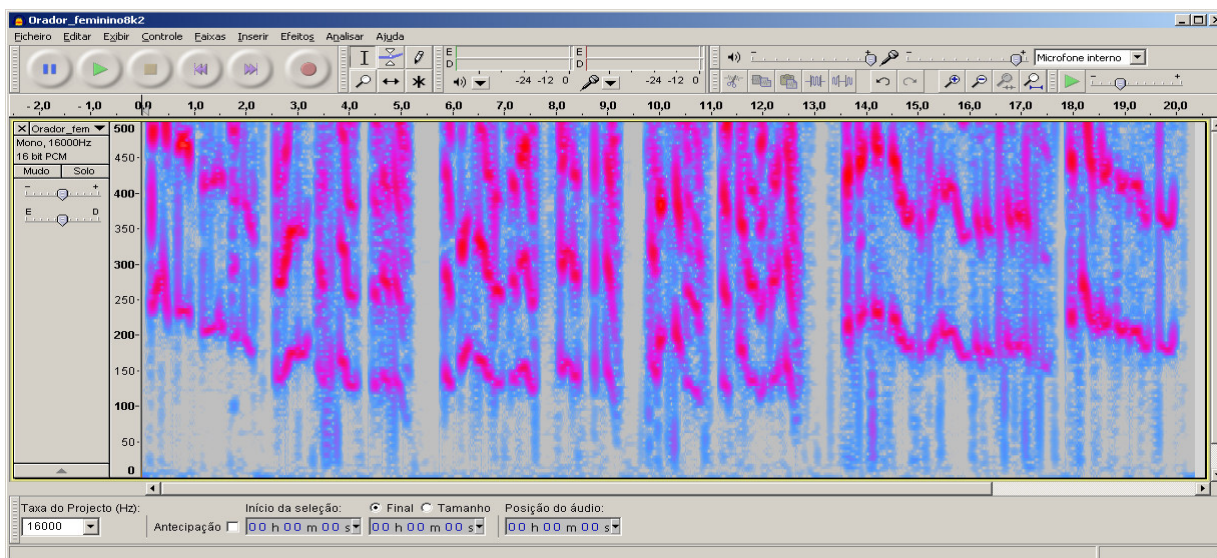


Figura 7-8 Arquivo02 Oradores femininos – Análise de frequência de pitch

7.2.3. Sinais analisados para alteração da codificação

Na análise do codificador, foram avaliados os parâmetros escalares não quantizados, ganho do *pitch* e o ganho do codificação. É possível observar que existem valores para os quais a codificação do sinal possui uma melhora do valor MOS avaliado pelo PESQ, o que consequentemente melhora *frames* avaliados.

Abaixo, a tabela que analisa os valores de alteração do ganho *pitch* e ganho de codificação demonstra um valor fixo, um percentual de adição à cada ganho. Neste caso, foram inseridos no programa fonte do codificador G729a com o algoritmo do PESQ. Como o codificador trabalha com valores estabelecidos pela análise da predição linear e busca dos vetores do *codebook* que geram ganhos, esses parâmetros - os ganhos escalares - foram escolhidos para alteração no processo. Propondo que se houvesse outro vetor muito próximo para a escolha do dos ganhos, o valor do MOS relacionado a cada frame teria uma pequena variação. Os primeiros arquivos de teste foram Orador_masculino8k.wav e Orador_feminino8k.wav. O total refere-se ao somatório das notas do PESQ para cada 28 *frames* do arquivo analisado:

BASE DE ESCOLHA DOS GANHOS MASCULINOS E FEMININOS pitch e code							
Análise do gain code = gain[1]			Análise do gain pitch = gain[0]				
.divisão	% gain adicional	soma PESQ Masc.	soma PESQ. Fem.	.divisão	% gain adicional	soma PESQ Masc.	soma PESQ. Fem.
1	100%	307,175	316,196	1	100%	306,143	313,727
2	50,00%	308,795	315,074	2	50,00%	307,370	312,882
3	33,33%	308,784	316,251	3	33,33%	307,666	314,327
4	25,00%	308,103	317,912	4	25,00%	307,881	316,544
5	20,00%	308,010	316,690	5	20,00%	308,035	314,449
10	10,00%	304,978	314,023	10	10,00%	308,379	317,927
15	6,67%	306,558	315,322	15	6,67%	308,123	314,785
20	5,00%	307,190	316,059	20	5,00%	307,669	315,949
25	4,00%	307,380	315,456	25	4,00%	307,705	317,235
30	3,33%	306,768	315,236	30	3,33%	307,490	317,234
35	2,86%	306,768	314,868	35	2,86%	307,315	317,488
40	2,50%	307,249	315,014	40	2,50%	306,669	317,759
45	2,22%	307,446	316,084	45	2,22%	307,723	316,178
50	2,00%	306,485	316,251	50	2,00%	307,269	314,908
55	1,82%	306,824	316,034	55	1,82%	307,728	314,986
60	1,67%	306,824	316,034	60	1,67%	309,138	315,974
65	1,54%	306,824	316,034	65	1,54%	309,248	314,339
70	1,43%	306,757	316,055	70	1,43%	308,052	314,339
75	1,33%	306,757	316,055	75	1,33%	308,036	314,339
80	1,25%	306,757	316,097	80	1,25%	308,036	315,377
85	1,18%	306,757	316,097	85	1,18%	307,307	315,377
90	1,11%	306,757	316,491	90	1,11%	307,502	314,779
95	1,05%	306,757	316,586	95	1,05%	307,502	314,332
100	1,00%	306,757	316,586	100	1,00%	307,407	315,089
original		307,366	316,328	original		307,366	316,328
pesq médio		3,531	3,515	pesq médio		3,533	3,515
melhor avaliação		308,795	317,912	melhor avaliação		309,248	317,927
pesq médio		3,549	3,532	pesq médio		3,555	3,533
melhora no processo		0,53%	0,50%	melhora no processo		0,61%	0,51%

Tabela 7-1 – Avaliação do PESQ para adição de ganho nos sinais de pitch e codificação

Programação no ganho de *pitch*:

$best_gain[1] = best_gain[1] + ((int)(best_gain[1]/\%));$ Onde %= valor adicional de *pitch*

Programação no ganho de codificação:

$best_gain[0] = best_gain[0] + ((int)(best_gain[0]/\%));$ Onde %=valor adicional de codificação

7.2.4. Ajuste da avaliação Perceptual (alterações significativas)

Constatou-se que, durante a análise dos intervalos de 28 frames avaliados pelo PESQ, existiam intervalos não avaliados, devido ao atraso de conversação e/ou intervalos de expressão ativa(fala) com mais de 25% de silêncio. Portanto, programou a regra de que quando não fosse possível avaliar o valor do PESQ, não seria aplicado o ganho nesse intervalo de 28 frames.

Durante a inicialização do arquivo para codificação, constatou-se que o G729.a inicializa o ganho de *pitch* com o maior valor quantizado 32676. Conclui-se que como não se sabe o instante de entrada do sinal, o primeiro frame recebe o maior valor de ganho de *pitch* possível.

Essa ação estava causando o problema de decisão dos frames subsequentes da análise. Foi implementada a avaliação de não efetuar nenhuma ação nos primeiros três intervalos de *frames* avaliados pelo PESQ.

```
C:\>cd project_test
C:\project_test>g729.exe Orador_masculino8k.wav masculino_original
Inicializa arq1= 1
Inicializa arq2= 1
*****      ITU G.729A 8 Kbits/seg  Speech Coder      *****
-----  Simulacao codec com PESQ  -----
-----  Versao 1.0 (Release 3, TESTE 21_07_2011)  -----

Entrada do arquivo de teste:  Orador_masculino8k.wav
Saida do arquivo bitstream:  masculino_original
vai declarar variaveis novas
Pijz
Orador_masculino8k.wav
Orador_masculino8k.wavdecoded
masculino_originalframe
tamanho do arquivo: 4480 pesq: 2.210000
tamanho do arquivo: 4480 pesq: 3.023000
tamanho do arquivo: 4480 pesq: 3.416000
ganho01_escrital= 55
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.552000
tamanho do arquivo: 4480 pesq: 2.700000
controle= 1
tamanho do arquivo: 4480 pesq: 3.022000
ganho01_escrital= 55
tamanho do arquivo: 4480 pesq: 3.595000
ganho01_escrital= 55
tamanho do arquivo: 4480 pesq: 2.970000
controle= 1
tamanho do arquivo: 4480 pesq: 3.207000
ganho01_escrital= 55
tamanho do arquivo: 4480 pesq: 3.785000
ganho01_escrital= 55
tamanho do arquivo: 4480 pesq: 3.708000
controle= 1
tamanho do arquivo: 4480 pesq: 4.062000
ganho01_escrital= 55
tamanho do arquivo: 4480 pesq: 3.986000
controle= 1
tamanho do arquivo: 4480 pesq: 3.885000
controle= 2
tamanho do arquivo: 4480 pesq: 3.884000
controle= 3
sem ACRESCIMO pitch= 1
tamanho do arquivo: 4480 pesq: 3.827000
controle= 1
```

Figura 7-9 Inicialização dos três primeiros *frames* e ação da avaliação zero

Para cada conjunto de oradores é realizada uma varredura previamente no sinal codificado com intuito de buscar o ganho mais adequado para o segmento de voz dos interlocutores em questão. Esse procedimento foi realizado nos quatro arquivos de testes e em cada um apresentou um ganho diferente, demonstrando a particularidade do sinal variante no tempo amostrado para cada grupo de seis oradores.

Nota-se que a adição de ganho inserido em cada arquivo de teste é aleatória, não depende do gênero do orador e sim do trato vocal do orador conforme tabela 7.2 apresentada a seguir.

ORADORES MASCULINO 1		ORADORES FEMININO 1	
adição pitch	adição codificação	adição pitch	adição codificação
1/65 = 1,54%	1/3 = 33,00%	1/10 = 10%	1/4 = 25%
ORADORES MASCULINO 2		ORADORES FEMININO 2	
adição pitch	adição codificação	adição pitch	adição codificação
1/35 = 2,86%	1/40 = 2,50%	1/55 = 1,82%	1/5 = 20%

Tabela 7-2 – Ganhos avaliados para os arquivos de teste

Os gráficos abaixo mostram como os ganhos foram encontrados. Cabe salientar que a análise de varredura já considerou o critério controle de três *frames*.

As figuras 7.10, 7.11, 7.12 e 7.13 mostram a escolha percentual de adição de ganho de *pitch* e codificação par cada arquivo composto por um conjunto de seis oradores.

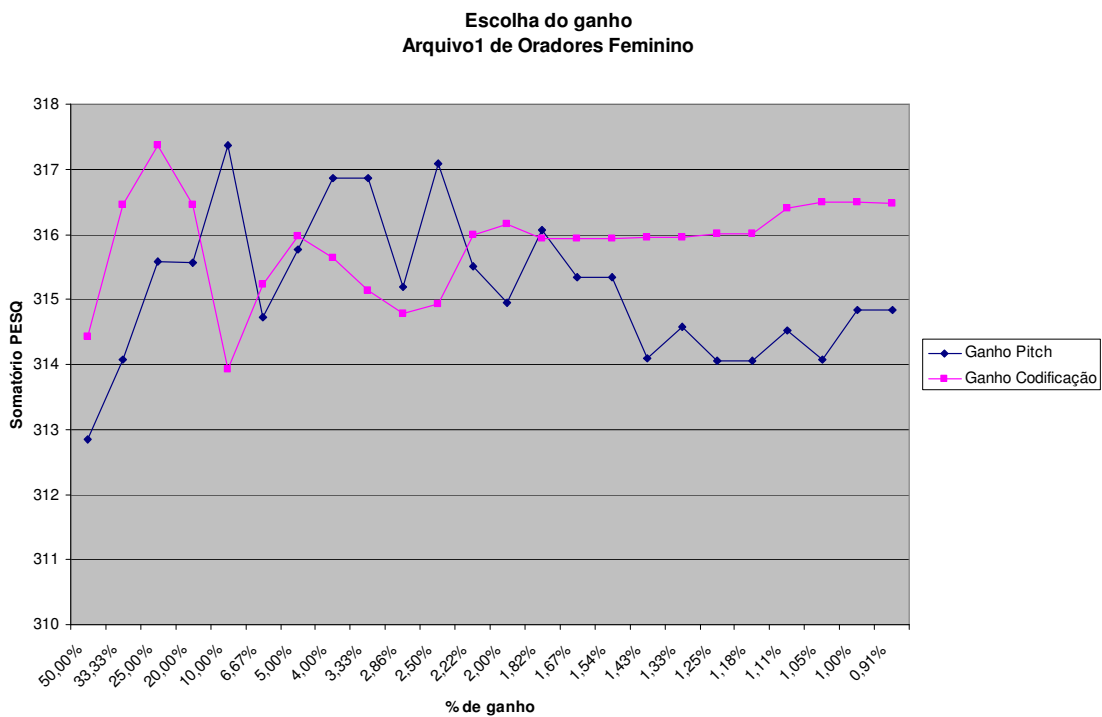


Figura 7-10 Avaliação dos ganhos no arquivo 01 de Oradores Femininos

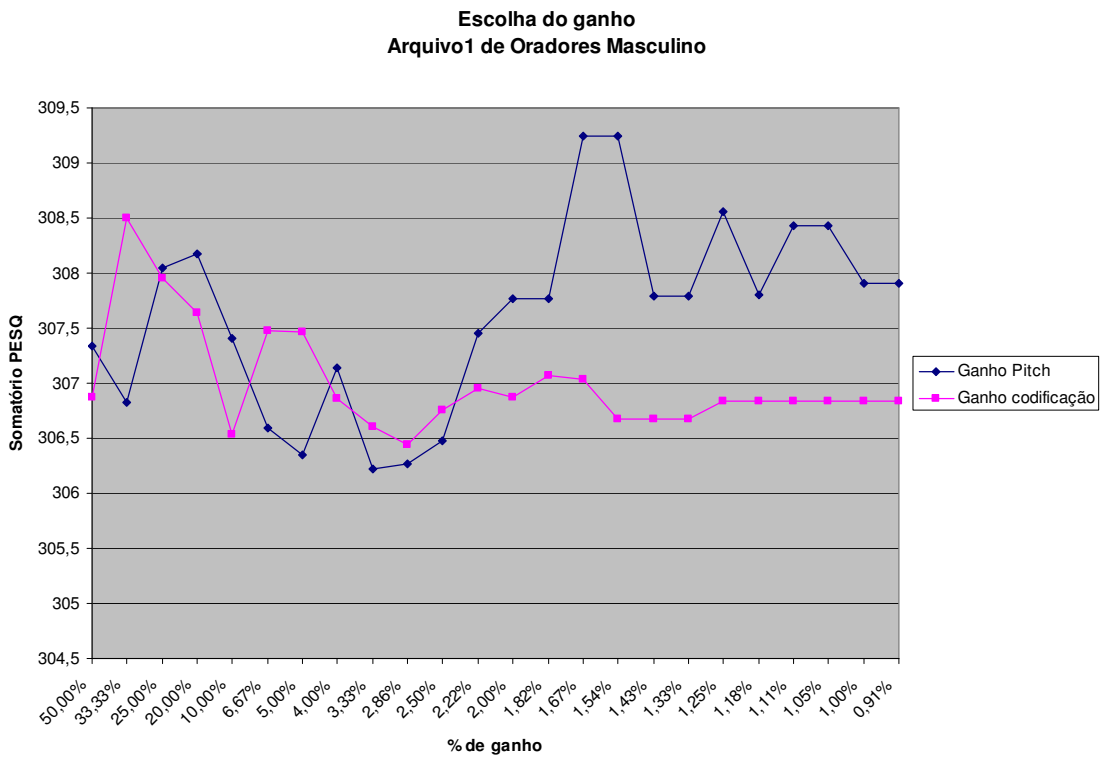


Figura 7-11 Avaliação dos ganhos no arquivo 01 de Oradores Masculinos

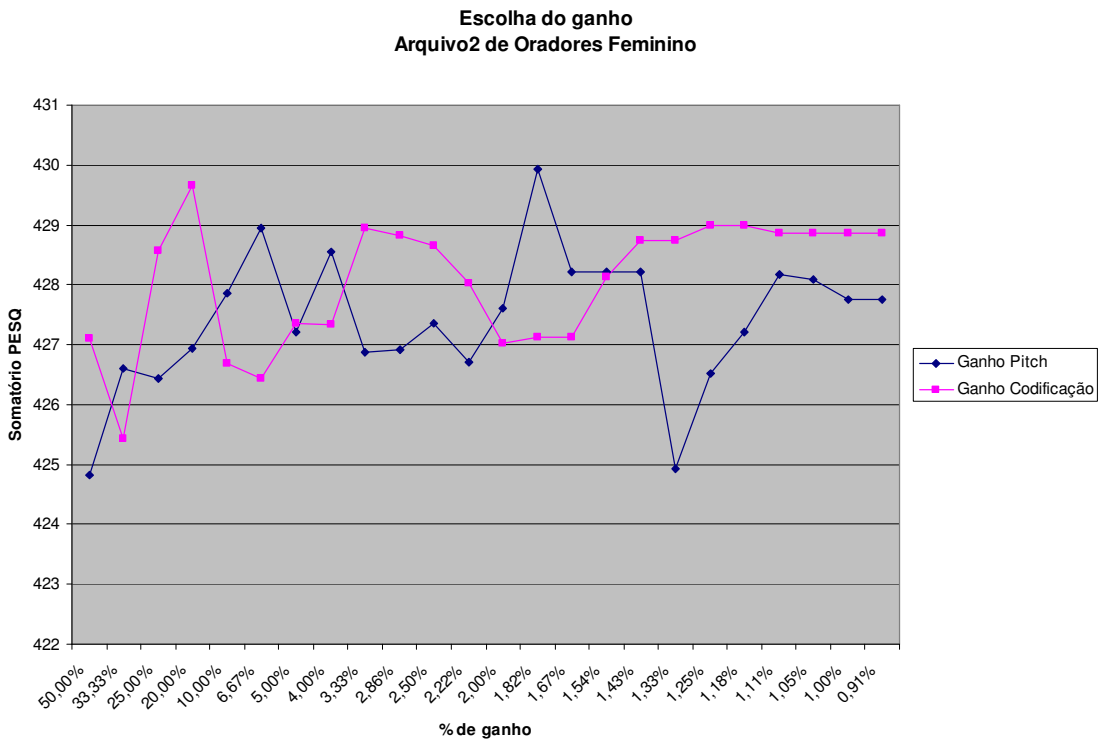


Figura 7-12 Avaliação dos ganhos no arquivo 02 de Oradores Femininos

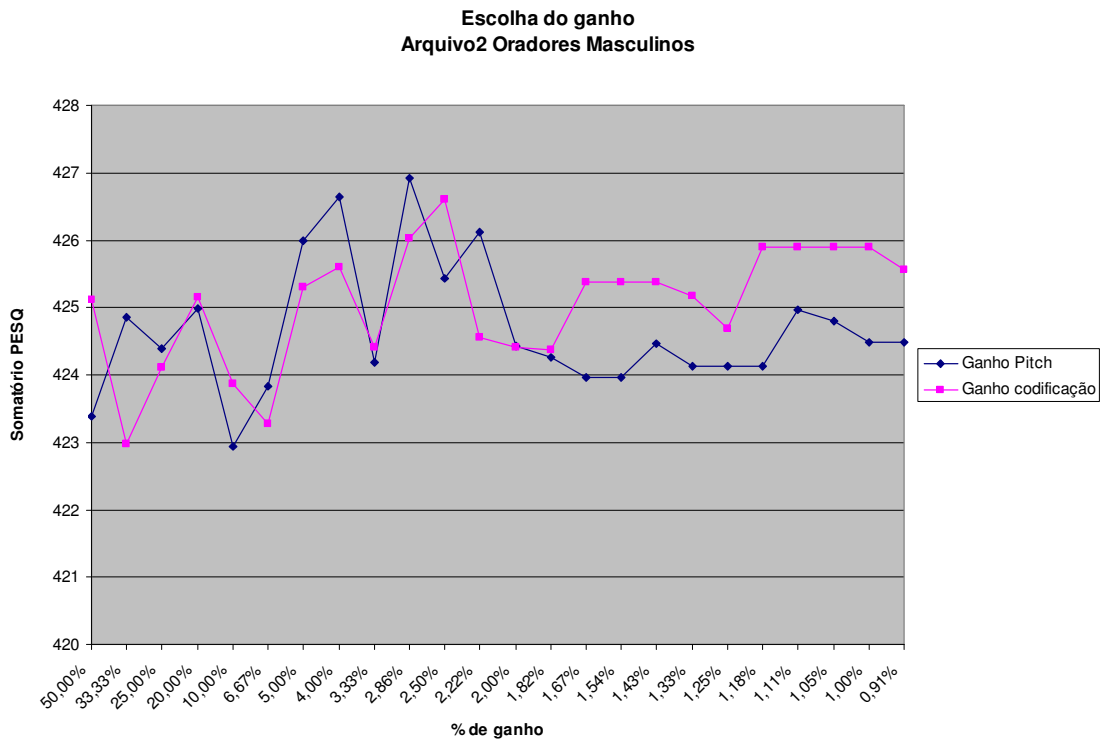


Figura 7-13 Avaliação dos ganhos no arquivo 02 de Oradores Masculinos

No algoritmo do PESQ, a avaliação do MOS ocorre a cada 100 frames, o equivalente à 16000 amostras ($100 \times 2 \times 80$) com tempo de intervalo de avaliação de 2s. Para avaliação de testes, variou o intervalo de *frames* e obteve que o menor valor aceitável é de 28 frames, o equivalente à 4480 amostras ($28 \times 2 \times 80$) com tempo de intervalo de avaliação de 0,56s. Portanto, otimizou-se a comparação do sinal original e do sinal degradado a cada 28 frames.

Nesta análise, como ação de controle, foi testada qual a quantidade de *frames* a ser analisada até que se necessite fazer alguma ação. Devido a formatação original do algoritmo (100 *frames*), optou-se em realizar ação de controle a cada três avaliações de 28 *frames* (84 *frames*) que pelos testes de ajuste do controle obteve a melhor avaliação perceptual.

A tabela 7.3 descrita a seguir apresenta os resultados dos experimentos de ação de controle sobre a tendência de avaliação do PESQ. A coluna controle demonstra quantos *frames* com valores decrescentes consecutivos ocorrem no processo para que aconteça a ação de controle.

arquivos1 de teste				arquivos2 de teste			
ORADORES MASCULINO 1		ORADORES FEMININO 1		ORADORES MASCULINO 2		ORADORES FEMININO 2	
CONTROLE	SOMATÓRIO PESQ	CONTROLE	SOMATÓRIO PESQ	CONTROLE	SOMATÓRIO PESQ	CONTROLE	SOMATÓRIO PESQ
	adição pitch 1/65		adição pitch 1/04		adição pitch 1/35		adição pitch 1/55
5	306,673	5	309,195	5	426,8	5	428,466
4	317,365	4	309,087	4	426,8	4	428,661
3	317,365	3	309,242	3	426,912	3	429,932
2	317,365	2	307,637	2	424,436	2	429,725
1	315,392	1	304,911	1	424,74	1	426,457

Tabela 7-3 – Ajuste do controle de ganho para os arquivos de teste

7.3. Padrão de avaliação perceptual proposto

A elaboração e o desenvolvimento de um padrão de avaliação perceptual aceitável que validasse a proposta basearam-se nos parâmetros avaliados no algoritmo do PESQ e nos parâmetros do codificador G729a. Para isso, a seguinte heurística foi utilizada:

1- A codificação inicializa com os ganhos originais nos três primeiros *frames*, para que o sinal de entrada seja estável, evitando a leitura de erro devido a uma característica interna do codificador.

2- Analisa-se a tendência da pontuação do MOS, como medida numérica de qualidade da voz humana, utilizando-se a leitura e escrita dos ganhos avaliados anteriormente. Esses ganhos são aplicados para cada conjunto de orador, conforme definido no experimento.

3- Como inicialização do modelo de comparação é realizada a tendência do valor da análise do PESQ das amostras sequenciais dos arquivos comparados, verificação do valor de MOS da amostra atual em relação ao valor de MOS da amostra anterior.

4 - Essas amostras são avaliadas em fragmentos sequenciais de 28 *frames* do PESQ, menor quantidade de *frames* avaliados.

5- Não será realizada nenhuma ação de controle quando houver intervalos de *frames* avaliados pelo PESQ sem um valor de MOS.

6- Controle:

i) Se tendência dessas amostras(intervalo de *frames* analisados sequencialmente) apresentarem um valor de MOS do intervalo de *frames* atual maior do que o valor de *frames* anterior, um ganho de *pitch* é adicionado ao processo.

ii) Se houver um valor de MOS regressivo durante três *frames* consecutivos, o codificador deverá alterar seus parâmetros internos para que o ganho retorne ao original, mantendo ao valores de codificação. A escolha por três intervalos de *frames* avaliados consecutivamente pelo PESQ é para manter a ação de controle dentro do tempo de avaliação padrão do algoritmo (100 *frames*).

iii) Quando o ganho estiver recebendo o sinal original, ele retorna para o passo “i” desta heurística e, assim, sucessivamente, até encerrar a avaliação dos *frames* recebidos.

8. Resultados Obtidos

A modificação no codificador CS-ACELP, com a utilização de uma análise perceptual, ponderando a avaliação do algoritmo do PESQ para decisão de atuação sobre a codificação, tem como objetivo aumentar a qualidade para os usuários.

Espera-se que seja possível a melhoria da qualidade de resposta do codec através da alteração do ganho de *pitch*, sem a modificação da essência do codec. Além disso, espera-se que essa implantação seja aceitável para o decodificador do G729a

8.1. Testes realizados

Os testes realizados foram baseados na utilização de arquivos de voz em formato PCM de 16 bits, amostrados a 8KHz, que serviram de entrada para o codificador original CS-ACELP e sua versão modificada.

8.1.1. Análise dos parâmetros escalares

Os ganhos foram avaliados serialmente. Observou-se que um acréscimo no ganho de *pitch* do sinal amostrado foi a forma mais significativa de avaliar o sinal.

Foi verificado que apenas a inserção de ganho no processo não é um parâmetro controlável, visto a variação do sinal, que não está sendo realizado nenhum tratamento.

Na tabela 8.1 a seguir, o primeiro teste que apenas incluiu-se o PESQ e aplicou-se um percentual de ganho de *pitch* foi realizado de forma serial em todos os frames avaliados. Obteve-se um ganho de 10% para o arquivo1 de oradores femininos e ganho de 1,54% para o arquivo1 de oradores masculinos. Isso ocorreu sem a heurística definida para o teste, pois esta foi a primeira avaliação.

ARQUIVOS 1 DE TESTE PARA Oradores masculinos e femininos									
Orador masculino					Orador feminino				
frames amostrados	PESQ Original	best_gain=1	% sucesso	frames melhorados	frames amostrados	PESQ Original	best_gain=1	% sucesso	frames melhorados
1	2.636	2.436	92,41%	0	1	2.208	2.034	92,12%	0
2	3.864	3.873	100,23%	1	2	3.641	3.675	100,93%	1
3	3.453	3.444	99,74%	0	3	3.664	3.687	100,63%	1
4	3.945	3.958	100,33%	1	4	3.798	3.751	98,76%	0
5	3.716	3.847	103,53%	1	5	3.831	3.942	102,90%	1
6	3.872	3.999	103,28%	1	6	3.740	3.564	95,29%	0
7	3.516	3.512	99,89%	0	7	3.794	3.758	99,05%	0
8	3.854	3.899	101,17%	1	8	4.036	4.092	101,39%	1
9	3.960	3.890	98,23%	0	9	3.504	3.572	101,94%	1
10	3.909	4.020	102,84%	1	10	3.688	3.693	100,14%	1
11	3.962	3.916	98,84%	0	11	3.455	3.360	97,25%	0
12	3.963	4.003	101,01%	1	12	3.356	3.185	94,90%	0
13	4.013	4.046	100,82%	1	13	3.800	3.831	100,82%	1
14	3.491	3.537	101,32%	1	14	3.362	3.206	95,36%	0
15	3.609	3.593	99,56%	0	15	3.453	3.624	104,95%	1
16	3.786	3.807	100,55%	1	16	3.833	3.778	98,57%	0
17	3.810	3.880	101,84%	1	17	3.604	3.629	100,69%	1
18	3.279	3.033	92,50%	0	18	3.664	3.692	100,76%	1
19	3.886	3.887	100,03%	1	19	3.644	3.687	101,18%	1
20	4.009	3.965	98,90%	0	20	3.970	4.047	101,94%	1
21	3.353	3.639	108,53%	1	21	3.458	3.449	99,74%	0
22	3.697	3.763	101,79%	1	22	3.983	3.988	100,13%	1
23	4.006	4.025	100,47%	1	23	3.861	3.928	101,74%	1
24	3.917	3.809	97,24%	0	24	4.014	3.905	97,28%	0
25	3.896	3.856	98,97%	0	25	3.139	3.281	104,52%	1
26	3.983	4.011	100,70%	1	26	3.639	3.685	101,26%	1
				0	27	3.754	3.677	97,95%	0
				0	28	3.409	3.326	97,57%	0
				0	29	3.960	4.002	101,06%	1
				61,54%					58,62%
100 frames	97.385				100 frames	105.262			

Tabela 8-1 – Testes somente com o PESQ inserido no codificador

Na análise acima, com a aplicação de um ganho linear em todo o processo de codificação de *pitch*, obteve-se um percentual de 61,54% de melhora de frames avaliados para o orador masculino e um percentual de 58,62% de melhora de frames avaliados para o orador feminino.

Para que o experimento tivesse uma análise criteriosa, foram estudadas e inseridas as premissas avaliadas anteriormente na metodologia conforme ambiente de teste proposto, codificador G729a com avaliação perceptual através do algoritmo PESQ, alterando parâmetros de acordo com a heurística apresentada no capítulo 7.

8.1.2. Avaliação Perceptual – Oradores masculino (Arquivo 01)

Nesta avaliação, foi utilizado o ganho de pitch de 1,54%(1/65), sendo esse o melhor valor avaliado para análise. Abaixo, o resultado da aplicação proposta de avaliação descrita no item 7.3.1.

Resultado do programa executado foi inserido na planilha abaixo. Esses valores foram extraídos da execução da proposta de avaliação com atuação do PESQ.

somatório acumulado de frames do PESQ com ganho do pitch do sistema	307,366
somatório acumulado de frames do PESQ com ganho pitch adicional	308,983
somatório acumulado de frames do PESQ com escolha do melhor valor	311,001
% de frames garantidos com a aplicação do padrão de avaliação perceptual proposto	72,41%
% de frames melhorados com a aplicação do padrão de avaliação perceptual proposto	44,83%

Tabela 8-2 – Resultado do programa executado no arquivo1 de Oradores Masculinos

Ao analisar a implementação proposta, comparando o sinal original com o sinal que teve atuação do controle com o algoritmo PESQ e adição de ganho de pitch, observa-se a melhora do sinal avaliado no experimento.

Quando avaliado perceptualmente o sinal de voz, existem *frames* garantidos (*frames* que mantiveram sua pontuação MOS) e *frames* melhorados (*frames* com pontuação MOS superior ao arquivo original). Abaixo, o sinal avaliado a cada intervalo de 28 *frames* .

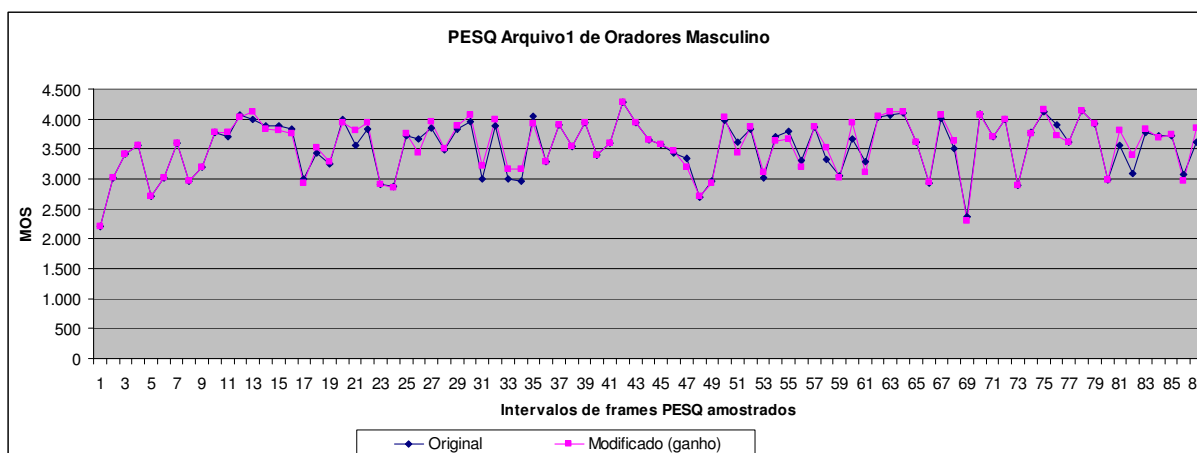


Figura 8-1 Gráfico da aplicação do ganho no arquivo 01 de Oradores Masculinos

Os parâmetros avaliados após a execução no programa com as alterações propostas apresentaram uma melhora no processo garantindo o valor original mantendo o sinal de voz ou melhorando o sinal de voz ao adicionar ganho e utilizar a heurística apresentada.

Arquivo1 de Oradores masculino						
frames PESQ amostrados	gain original	gain pitch=1/65	% de avaliação	Frames Garantidos	Frames Melhorados	Melhor sinal
1	2.210	2.210	100,00%	1	0	2.210
2	3.023	3.023	100,00%	1	0	3.023
3	3.416	3.416	100,00%	1	0	3.416
4	3.552	3.552	100,00%	1	0	3.552
5	2.708	2.708	100,00%	1	0	2.708
6	3.022	3.022	100,00%	1	0	3.022
7	3.595	3.595	100,00%	1	0	3.595
8	2.970	2.970	100,00%	1	0	2.970
9	3.207	3.207	100,00%	1	0	3.207
10	3.785	3.785	100,00%	1	0	3.785
11	3.708	3.782	102,00%	1	1	3.782
12	4.062	4.025	99,09%	0	0	4.062
13	3.986	4.123	103,44%	1	1	4.123
14	3.885	3.834	98,69%	0	0	3.885
15	3.884	3.821	98,38%	0	0	3.884
16	3.827	3.767	98,43%	0	0	3.827
17	3.002	2.931	97,63%	0	0	3.002
18	3.438	3.522	102,44%	1	1	3.522
19	3.249	3.283	101,05%	1	1	3.283
20	3.998	3.934	98,40%	0	0	3.998
21	3.556	3.808	107,09%	1	1	3.808
22	3.834	3.943	102,84%	1	1	3.943
23	2.912	2.916	100,14%	1	1	2.916
24	2.866	2.851	99,48%	0	0	2.866
25	3.716	3.761	101,21%	1	1	3.761
26	3.663	3.437	93,83%	0	0	3.663
27	3.852	3.952	102,60%	1	1	3.952
28	3.483	3.499	100,46%	1	1	3.499
29	3.826	3.880	101,41%	1	1	3.880
30	3.952	4.061	102,76%	1	1	4.061
31	2.998	3.221	107,44%	1	1	3.221
32	3.885	3.994	102,81%	1	1	3.994
33	2.997	3.164	105,57%	1	1	3.164
34	2.958	3.171	107,20%	1	1	3.171
35	4.050	3.919	96,77%	0	0	4.050
36	3.283	3.288	100,15%	1	1	3.288
37	3.900	3.901	100,03%	1	1	3.901
38	3.540	3.540	100,00%	1	0	3.540
39	3.932	3.932	100,00%	1	0	3.932
40	3.391	3.391	100,00%	1	0	3.391
41	3.604	3.604	100,00%	1	0	3.604
42	4.275	4.275	100,00%	1	0	4.275
43	3.943	3.943	100,00%	1	0	3.943
44	3.649	3.649	100,00%	1	0	3.649
45	3.574	3.574	100,00%	1	0	3.574
46	3.426	3.475	101,43%	1	1	3.475
47	3.350	3.192	95,28%	0	0	3.350
48	2.698	2.716	100,67%	1	1	2.716
49	2.959	2.934	99,16%	0	0	2.959
50	3.974	4.024	101,26%	1	1	4.024
51	3.615	3.438	95,10%	0	0	3.615
52	3.831	3.861	100,78%	1	1	3.861
53	3.020	3.114	103,11%	1	1	3.114
54	3.701	3.628	98,03%	0	0	3.701
55	3.800	3.662	96,37%	0	0	3.800
56	3.303	3.192	96,64%	0	0	3.303
57	3.865	3.871	100,16%	1	1	3.871
58	3.333	3.520	105,61%	1	1	3.520
59	3.048	3.026	99,28%	0	0	3.048
60	3.672	3.937	107,22%	1	1	3.937
61	3.281	3.115	94,94%	0	0	3.281
62	4.034	4.050	100,40%	1	1	4.050
63	4.075	4.124	101,20%	1	1	4.124
64	4.106	4.119	100,32%	1	1	4.119
65	3.615	3.606	99,75%	0	0	3.615
66	2.929	2.947	100,61%	1	1	2.947
67	4.010	4.067	101,42%	1	1	4.067
68	3.506	3.630	103,54%	1	1	3.630
69	2.366	2.289	96,75%	0	0	2.366
70	4.087	4.068	99,54%	0	0	4.087
71	3.708	3.708	100,00%	1	0	3.708
72	3.990	3.990	100,00%	1	0	3.990
73	2.893	2.893	100,00%	1	0	2.893
74	3.770	3.767	99,92%	0	0	3.770
75	4.120	4.149	100,70%	1	1	4.149
76	3.911	3.728	95,32%	0	0	3.911
77	3.606	3.607	100,03%	1	1	3.607
78	4.145	4.145	100,00%	1	0	4.145
79	3.914	3.914	100,00%	1	0	3.914
80	2.985	2.985	100,00%	1	0	2.985
81	3.558	3.818	107,31%	1	1	3.818
82	3.094	3.400	109,89%	1	1	3.400
83	3.778	3.840	101,64%	1	1	3.840
84	3.720	3.695	99,33%	0	0	3.720
85	3.717	3.749	100,86%	1	1	3.749
86	3.077	2.963	96,30%	0	0	3.077
87	3.610	3.843	106,45%	1	1	3.843
	307.366	308.983		72,41%	44,83%	311.001

Tabela 8-3 – Avaliação dos parâmetros extraídos do Arquivo1 de Oradores Masculinos

Abaixo, os sinais analisados no arquivo1 de oradores masculinos pelo *software Audacity*. A figura 8.2 descrita abaixo mostra a leitura do sinal de voz:

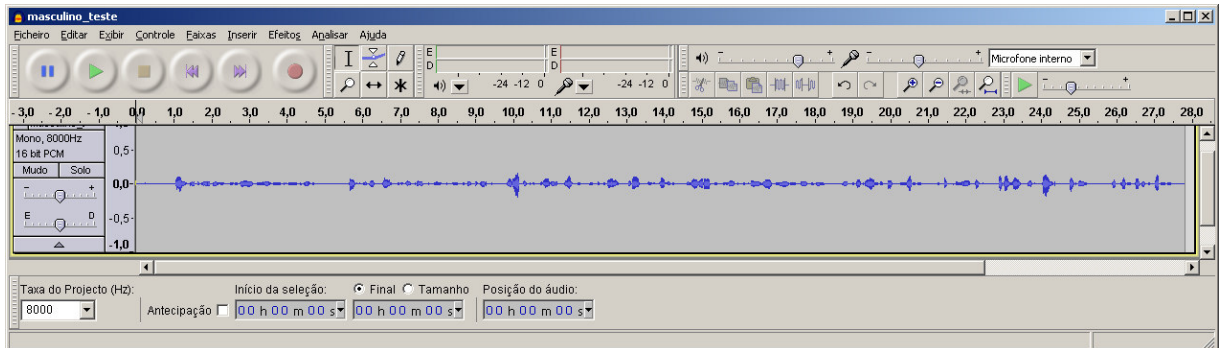


Figura 8-2 Arquivo1 Oradores masculinos, sinal codificado e decodificado com atuação do PESQ

Na análise do espectro abaixo, podemos verificar que a primeira formante, a frequência de *pitch*, varia de acordo com cada orador. Nesta análise, podemos verificar seis frequências distintas.

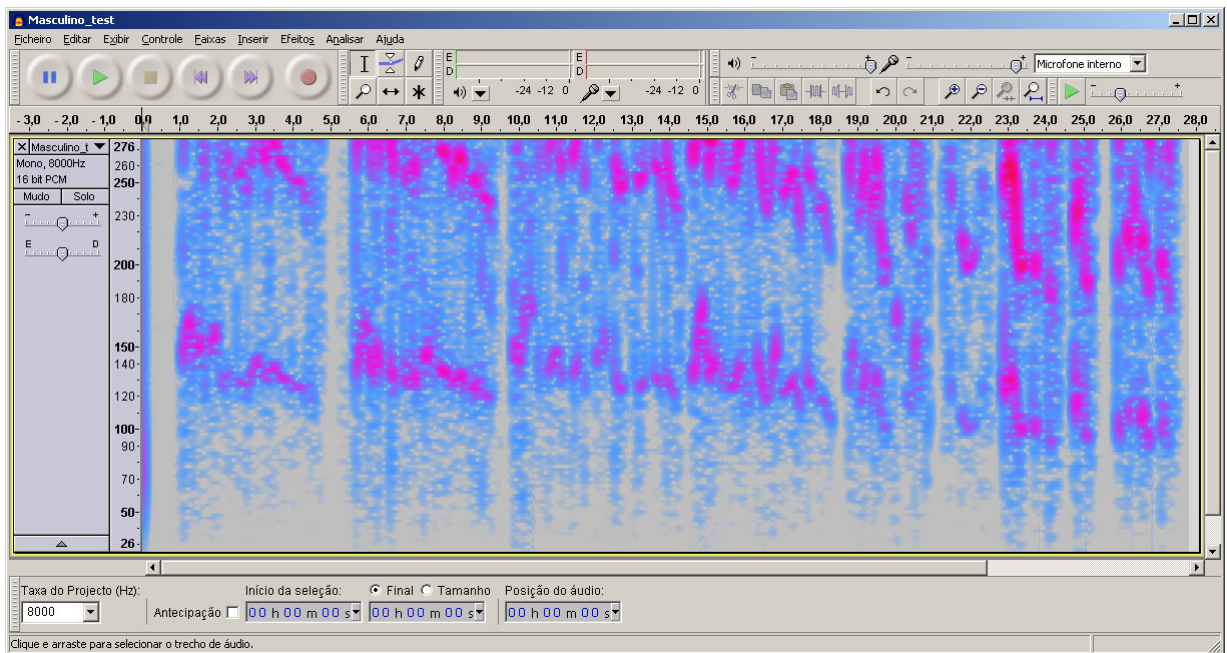


Figura 8-3 Arquivo1 Oradores masculinos, análise da frequência de pitch

8.1.3. Avaliação Perceptual – Oradores feminino (Arquivo 01)

Nesta avaliação, foi utilizado o ganho de *pitch* de 10%(1/10), sendo este o melhor valor avaliado para análise. Abaixo, o resultado da aplicação proposta de avaliação descrita no item 7.3.1.

Resultado do programa executado foi colocado na planilha abaixo. Esses valores foram extraídos da execução da proposta de avaliação com atuação do PESQ.

somatório acumulado de frames do PESQ com ganho do pitch do sistema	316,328
somatório acumulado de frames do PESQ com ganho pitch adicional	317,365
somatório acumulado de frames do PESQ com escolha do melhor valor	329,842
% de frames garantidos com a aplicação do padrão de avaliação perceptual proposto	55,56%
% de frames melhorados com a aplicação do padrão de avaliação perceptual proposto	50,0%

Tabela 8-4 – Resultado do programa executado no arquivo1 de Oradores Femininos

Ao analisar a pontuação do sinal original com o sinal modificado pela atuação do controle com o algoritmo PESQ e adição de ganho de *pitch*, visualiza-se no gráfico abaixo a garantia do sinal quando avaliado perceptualmente. Neste caso, a melhora não foi significativa como ocorreu no arquivo do orador masculino.

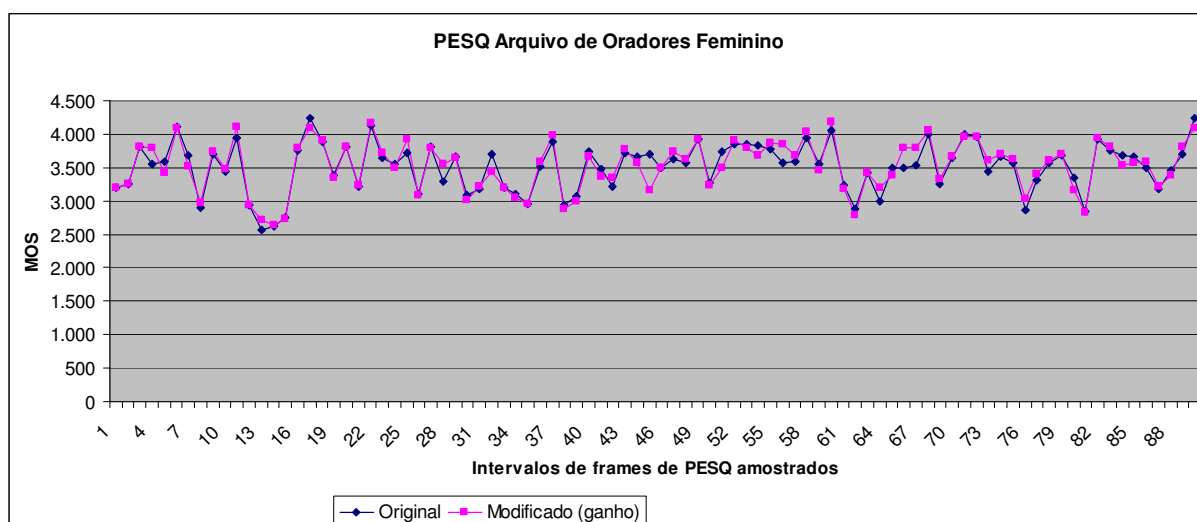


Figura 8-4 Gráfico da aplicação do ganho no arquivo 01 de Oradores Femininos

Os parâmetros avaliados após a execução no programa com as alterações propostas.

Arquivo1 de Oradores feminino						
PESQ amostrados	gain original	gain pitch=1/65	% de avaliação	Frames Garantidos	Frames Melhorados	Melhor sinal
1	3.205	3.205	100,00%	1	0	3.205
2	3.248	3.248	100,00%	1	0	3.248
3	3.819	3.819	100,00%	1	0	3.819
4	3.557	3.795	106,69%	1	1	3.795
5	3.567	3.416	95,23%	0	0	3.567
6	4.104	4.089	99,63%	0	0	4.104
7	3.680	3.519	95,63%	0	0	3.680
8	2.903	2.984	102,79%	1	1	2.984
9	3.693	3.737	101,19%	1	1	3.737
10	3.439	3.471	100,93%	1	1	3.471
11	3.951	4.104	103,87%	1	1	4.104
12	2.940	2.933	99,76%	0	0	2.940
13	2.559	2.723	106,41%	1	1	2.723
14	2.624	2.643	100,72%	1	1	2.643
15	2.743	2.732	99,60%	0	0	2.743
16	3.760	3.785	100,66%	1	1	3.785
17	4.234	4.088	96,55%	0	0	4.234
18	3.889	3.901	100,31%	1	1	3.901
19	3.376	3.355	99,38%	0	0	3.376
20	3.821	3.818	99,92%	0	0	3.821
21	3.218	3.240	100,68%	1	1	3.240
22	4.130	4.171	100,99%	1	1	4.171
23	3.638	3.710	101,98%	1	1	3.710
24	3.555	3.488	98,12%	0	0	3.555
25	3.714	3.919	105,52%	1	1	3.919
26	3.106	3.093	99,58%	0	0	3.106
27	3.809	3.800	99,76%	0	0	3.809
28	3.290	3.547	107,81%	1	1	3.547
29	3.660	3.647	99,64%	0	0	3.660
30	3.091	3.014	97,51%	0	0	3.091
31	3.176	3.211	101,10%	1	1	3.211
32	3.703	3.448	93,11%	0	0	3.703
33	3.200	3.200	100,00%	1	0	3.200
34	3.098	3.052	98,52%	0	0	3.098
35	2.964	2.964	100,00%	1	0	2.964
36	3.520	3.591	102,02%	1	1	3.591
37	3.887	3.975	102,26%	1	1	3.975
38	2.932	2.890	98,57%	0	0	2.932
39	3.068	2.997	97,69%	0	0	3.068
40	3.744	3.671	98,05%	0	0	3.744
41	3.479	3.364	96,69%	0	0	3.479
42	3.212	3.355	104,45%	1	1	3.355
43	3.718	3.784	101,78%	1	1	3.784
44	3.667	3.574	97,46%	0	0	3.667
45	3.698	3.154	85,29%	0	0	3.698
46	3.503	3.488	99,57%	0	0	3.503
47	3.617	3.741	103,43%	1	1	3.741
48	3.573	3.634	101,71%	1	1	3.634
49	3.931	3.921	99,75%	0	0	3.931
50	3.265	3.241	99,26%	0	0	3.265
51	3.736	3.500	93,68%	0	0	3.736
52	3.841	3.903	101,61%	1	1	3.903
53	3.841	3.785	98,54%	0	0	3.841
54	3.837	3.690	96,17%	0	0	3.837
55	3.770	3.868	102,60%	1	1	3.868
56	3.573	3.848	107,70%	1	1	3.848
57	3.593	3.686	102,59%	1	1	3.686
58	3.938	4.034	102,44%	1	1	4.034
59	3.545	3.458	97,55%	0	0	3.545
60	4.062	4.176	102,81%	1	1	4.176
61	3.230	3.183	98,54%	0	0	3.230
62	2.877	2.798	97,25%	0	0	2.877
63	3.423	3.422	99,97%	0	0	3.423
64	2.994	3.193	106,65%	1	1	3.193
65	3.494	3.378	96,68%	0	0	3.494
66	3.488	3.790	108,66%	1	1	3.790
67	3.533	3.802	107,61%	1	1	3.802
68	4.004	4.061	101,42%	1	1	4.061
69	3.249	3.328	102,43%	1	1	3.328
70	3.640	3.658	100,49%	1	1	3.658
71	4.007	3.958	98,78%	0	0	4.007
72	3.963	3.957	99,85%	0	0	3.963
73	3.436	3.605	104,92%	1	1	3.605
74	3.661	3.707	101,26%	1	1	3.707
75	3.567	3.618	101,43%	1	1	3.618
76	2.867	3.023	105,44%	1	1	3.023
77	3.318	3.405	102,62%	1	1	3.405
78	3.567	3.613	101,29%	1	1	3.613
79	3.676	3.698	100,60%	1	1	3.698
80	3.342	3.154	94,37%	0	0	3.342
81	2.847	2.830	99,40%	0	0	2.847
82	3.922	3.949	100,69%	1	1	3.949
83	3.755	3.820	101,73%	1	1	3.820
84	3.691	3.542	95,96%	0	0	3.691
85	3.658	3.566	97,48%	0	0	3.658
86	3.502	3.594	102,63%	1	1	3.594
87	3.189	3.216	100,85%	1	1	3.216
88	3.453	3.393	98,26%	0	0	3.453
89	3.699	3.810	103,00%	1	1	3.810
90	4.242	4.097	96,58%	0	0	4.242
	316.328	317.365		55,56%	50,00%	320.842

Tabela 8-5 – Avaliação dos parâmetros extraídos do Arquivo1 de Oradores Femininos

Abaixo, os sinais analisados no arquivo1 de oradores femininos pelo *software Audacity*:

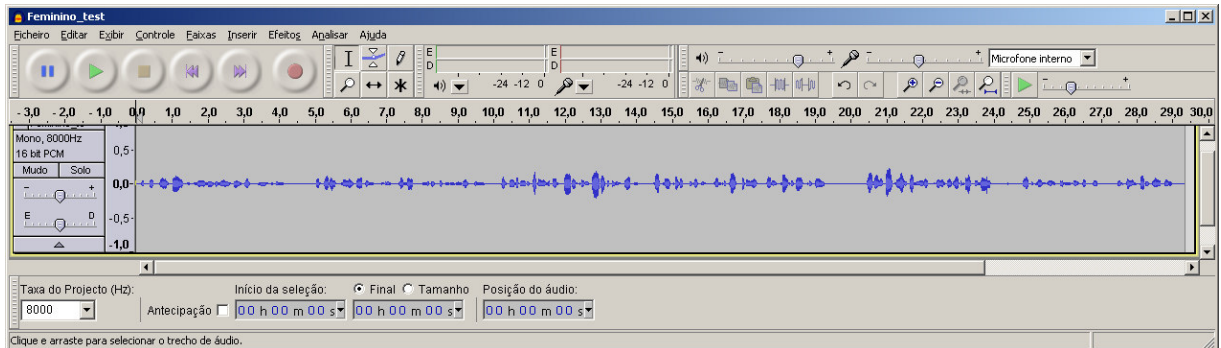


Figura 8-5 Arquivo1 Oradores femininos, sinal codificado e decodificado com atuação do PESQ

Observam-se na figura 8.6 as diferentes formantes para cada um dos seis oradores no arquivo de teste elaborado para o experimento.

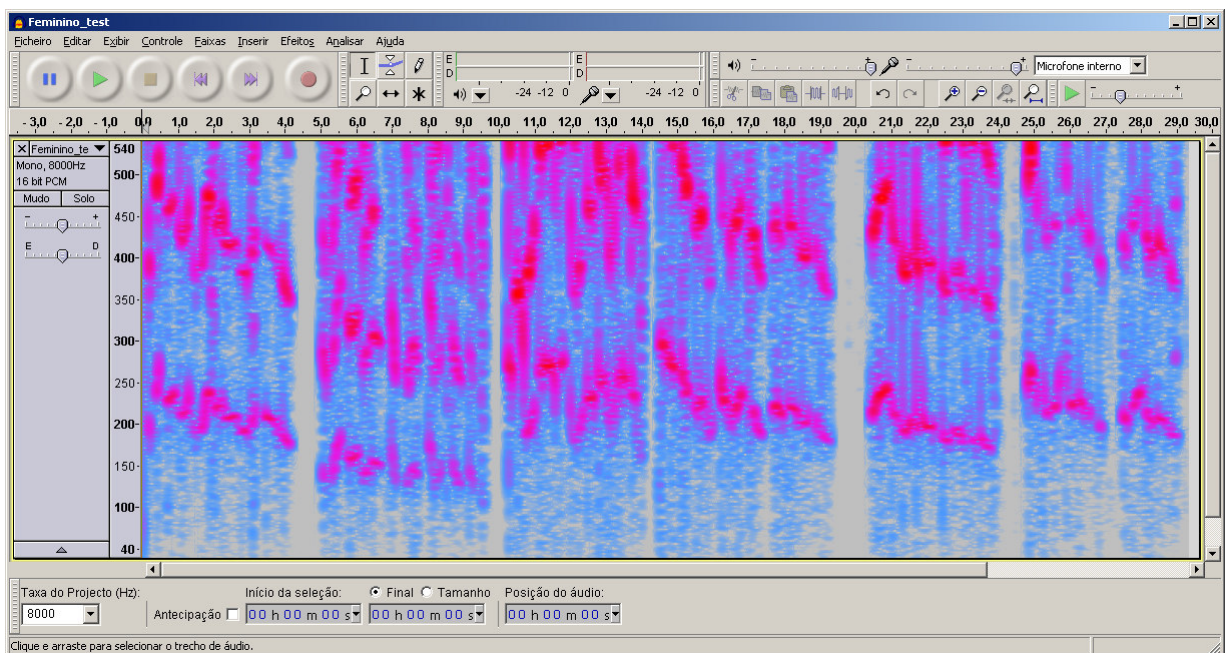


Figura 8-6 Arquivo1 Oradores femininos, análise da frequência de pitch

8.1.4. Avaliação Perceptual – Oradores masculino (Arquivo 02)

Nesta avaliação, foi utilizado o ganho de *pitch* de 2,86% (1/35), sendo este o melhor valor avaliado para análise. Abaixo, o resultado da aplicação proposta de avaliação descrita no item 7.3.1.

Resultado do programa executado foi colocado na planilha abaixo. Esses valores foram extraídos da execução da proposta de avaliação com atuação do PESQ.

somatório acumulado de frames do PESQ com ganho do pitch do sistema	426,227
somatório acumulado de frames do PESQ com ganho pitch adicional	426,799
somatório acumulado de frames do PESQ com escolha do melhor valor	433,666
% de frames garantidos com a aplicação do padrão de avaliação perceptual proposto	52,94%
% de frames melhorados com a aplicação do padrão de avaliação perceptual proposto	46,32%

Tabela 8-6 – Resultado do programa executado no arquivo2 de Oradores Masculinos

Ao analisar a pontuação do sinal original com o sinal que teve atuação do controle com o algoritmo PESQ e adição de ganho de *pitch*, visualizam-se, no gráfico abaixo, a garantia e a melhora do sinal quando avaliado perceptualmente.

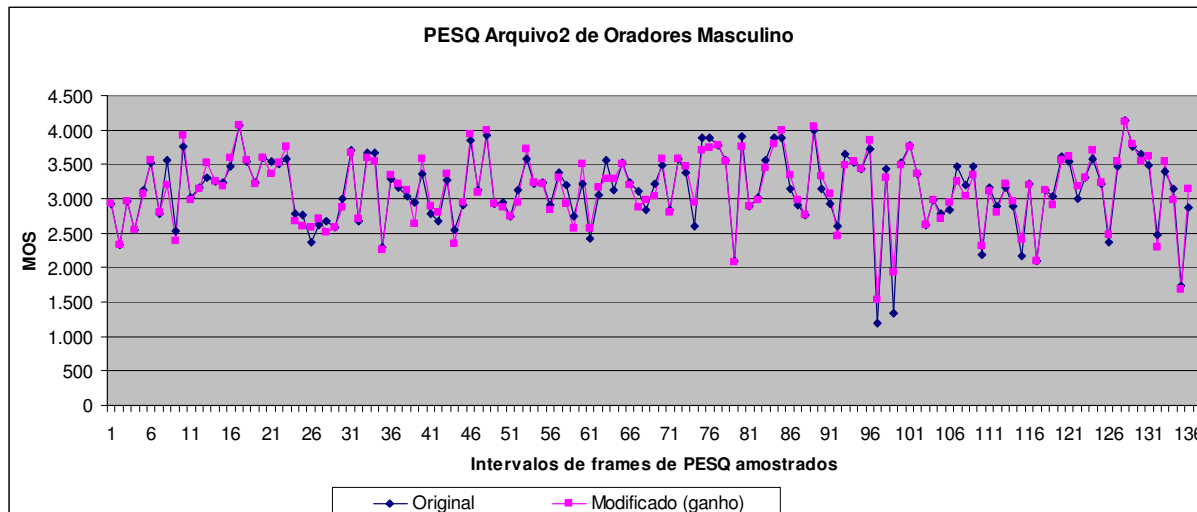


Figura 8-7 Gráfico da aplicação do ganho no arquivo 02 de Oradores Masculinos

Abaixo, os sinais analisados no arquivo2 de oradores masculinos pelo *software Audacity*:

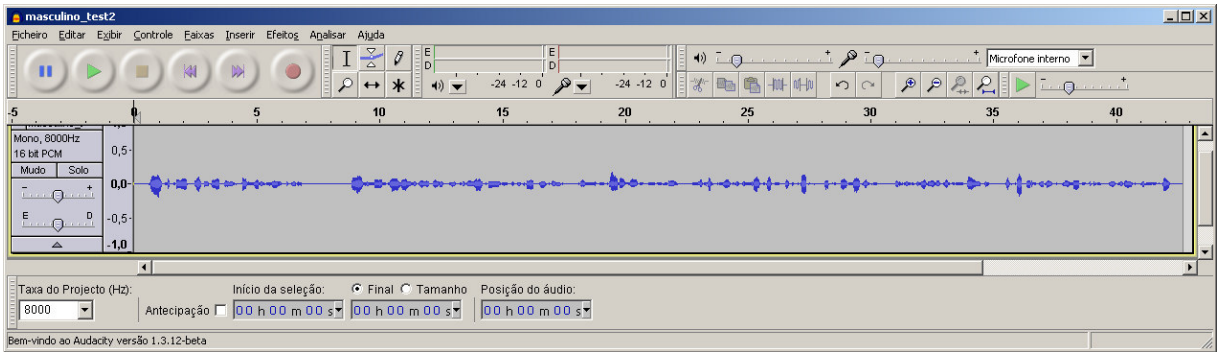


Figura 8-8 Arquivo2 Oradores masculinos, sinal codificado e decodificado com atuação do PESQ

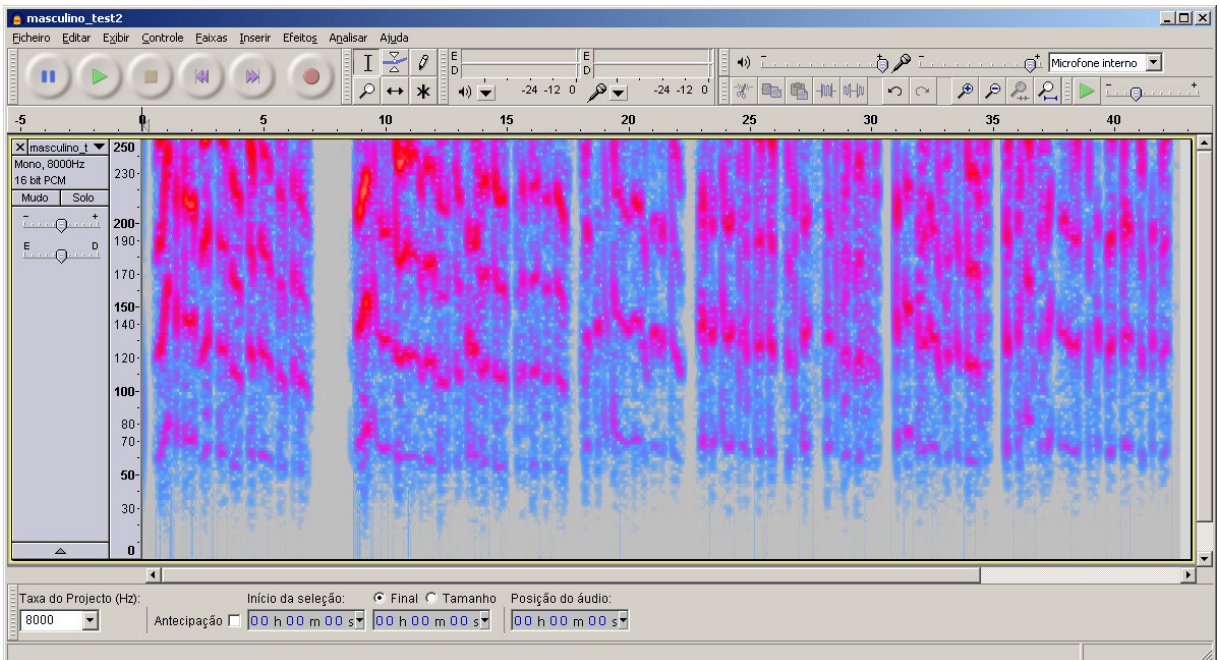


Figura 8-9 Arquivo2 Oradores masculinos, análise da frequência de pitch

8.1.5. Avaliação Perceptual – Oradores feminino (Arquivo 02)

Nesta avaliação foi utilizado o ganho de pitch de 1,82% (1/55), sendo este o melhor valor avaliado para análise. Abaixo, o resultado da aplicação proposta de avaliação descrita no item 7.3.1.

Resultado do programa executado foi colocado na planilha abaixo. Esses valores foram extraídos da execução da proposta de avaliação com atuação do PESQ.

somatório acumulado de frames do PESQ com ganho do pitch do sistema	428,255
somatório acumulado de frames do PESQ com ganho pitch adicional	429,932
somatório acumulado de frames do PESQ com escolha do melhor valor	435,920
% de frames garantidos com a aplicação do padrão de avaliação perceptual proposto	61,65%
% de frames melhorados com a aplicação do padrão de avaliação perceptual proposto	50,38%

Tabela 8-8 – Resultado do programa executado no arquivo2 de Oradores Femininos

Ao analisar a pontuação do sinal original com o sinal modificado pela atuação do controle com o algoritmo PESQ e adição de ganho de *pitch*, visualiza-se, no gráfico abaixo, a garantia do sinal quando avaliado perceptualmente.

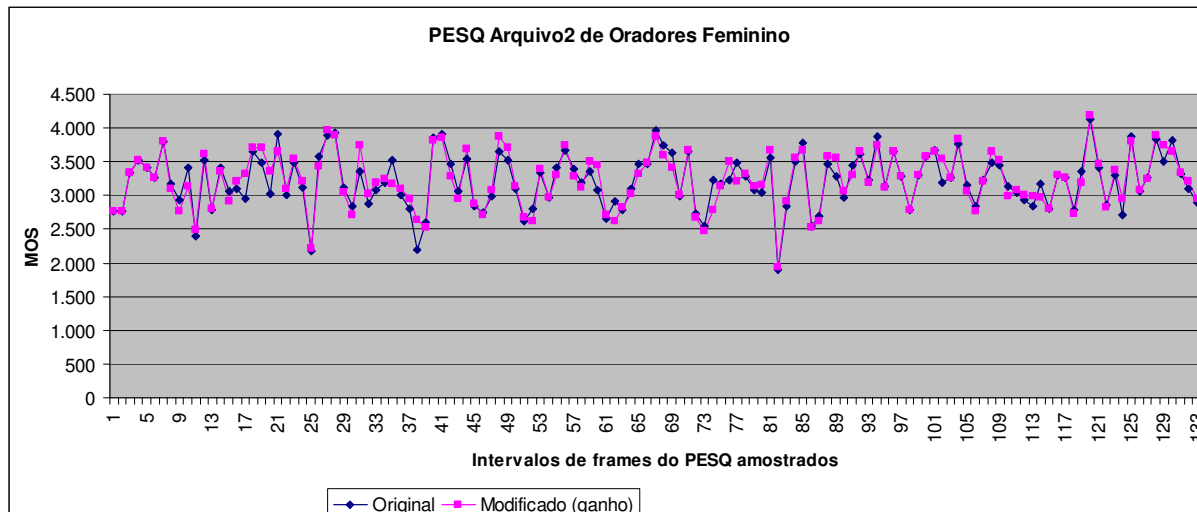


Figura 8-10 Gráfico da aplicação do ganho no arquivo 02 de Oradores Femininos

Abaixo os sinais analisados no arquivo1 de oradores femininos pelo *software Audacity*:

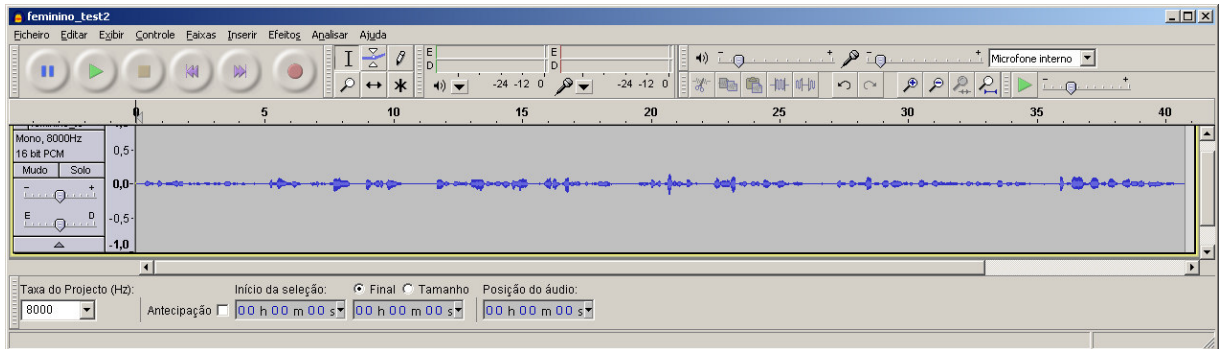


Figura 8-11 Arquivo2 Oradores femininos, sinal codificado e decodificado com atuação do PESQ

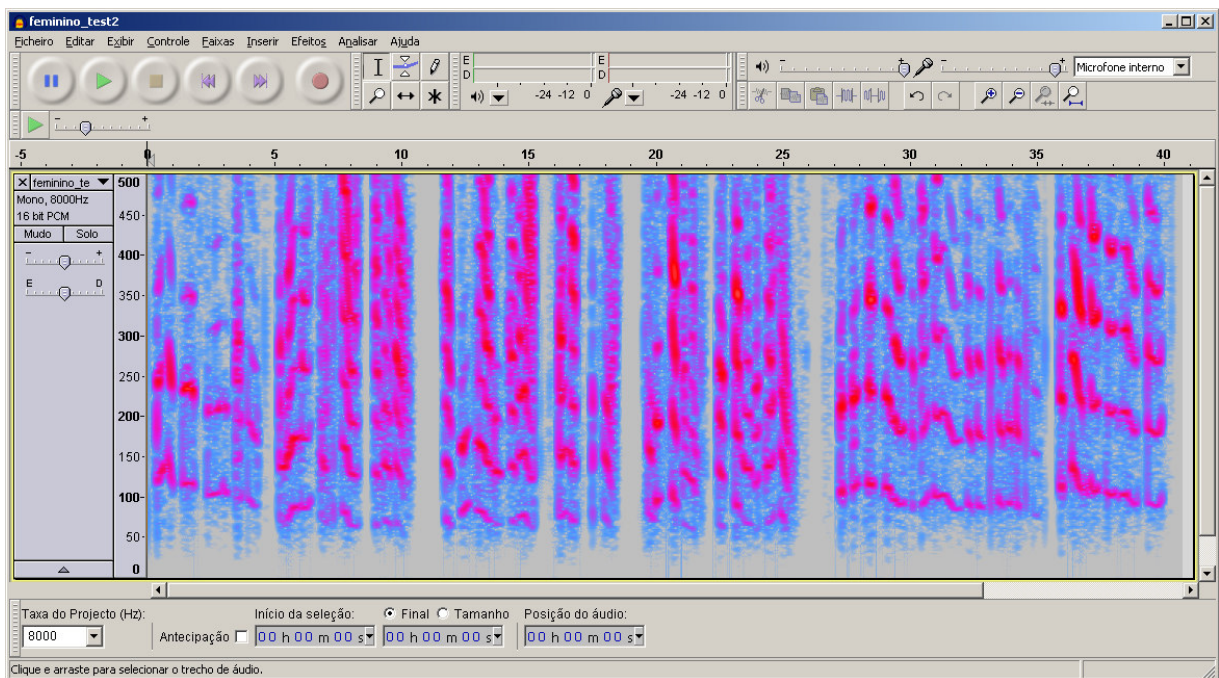


Figura 8-12 Arquivo2 Oradores femininos, análise da frequência de pitch

9. Conclusões

A avaliação foi realizada sobre um modelo de codificação da voz que analisa a qualidade perceptual. Foi inserido dentro do código fonte do codificador G729a, o algoritmo do PESQ. Este critério de decisão para as intervenções no sistema teve importantes considerações a ressaltar.

1-O codificador, ao iniciar o processamento do sinal de voz para codificação, utiliza como referência o valor mais expressivo da escala de quantização do ganho de *pitch*. Esse processo quando constatado foi excluído do critério de avaliação de decisão, retirando a ação de controle dos dois primeiros *frames*, visto que a influência desses *frames* distorcia a análise do sinal, quando executando a heurística proposta.

2- A ausência de sinal também foi um item importante da análise. Segundo (ITU P862, 2001), a avaliação perceptual é obtida pelo algoritmo do PESQ somente quando existe um período de silêncio inferior a 25% do frame de voz analisado. Sendo este o entendimento para os *frames* não avaliados quando executado o algoritmo.

3- O mapeamento dos ganhos de *pitch* e de codificação não quantizados no codificador G729a é importante para avaliar que a pontuação mais significativa. Em todas as análises realizadas, utilizou-se somente o ganho de *pitch*. Foi obtida uma melhor avaliação perceptual quando efetuada adição de ganho. Deste modo, desconsiderou-se a utilização do ganho de codificação como parâmetro para melhoria de qualidade do sinal.

4- Foi realizado teste com os dois ganhos simultâneos, *pitch* e codificação. Entretanto, com o padrão de avaliação adotado pela heurística, não foi obtido sucesso na melhora do processo, pois os valores de ganhos foram verificados separadamente e não estão automatizados.

5- Foram verificados parâmetros específicos do codificador como os valores do “GAMMA”(variável no código fonte do G729a). Esse parâmetro altera as características do filtro de ponderação do codificador. Entretanto, este é um valor fixo no *encoder* e no *decoder*. Assim, tendo como proposta desta dissertação a avaliação do sinal na origem, não foram realizadas alterações em parâmetros tabelados pelo codificador para não alterar o processo de decodificação do sinal.

6- A proposta de modificar *frames a frames* o sinal de voz dinamicamente, ou seja, avaliados a cada intervalo de 28 frames pelo PESQ foi realizada com sucesso. Apesar de um valor baixo para o somatório do PESQ, a melhoria perceptual ocorreu. Constatou-se que o PESQ efetua a média das avaliações do sinal de voz amostrado. Na análise de um arquivo inteiro o valor do PESQ não altera muito a pontuação referida na escala MOS, com variação de 0 à 5 pontos. Neste sentido, cabe ressaltar que o algoritmo do PESQ aplica a pontuação máxima de 4,5 pontos para sinais idênticos.

7- A heurística realizada para obter resultados de melhoria perceptual do sinal de voz foi efetuada com sucesso. Abaixo, um resumo dos parâmetros avaliados para cada conjunto de oradores. O somatório acumulado é a soma do valor da pontuação MOS que cada intervalo de 28 *frames* do PESQ.

	descrição do parâmetros avaliados	masc 01	masc 02	fem 01	fem 02
1	somatório acumulado de frames do PESQ com ganho do pitch do sistema	307,366	426,227	316,328	428,255
2	somatório acumulado de frames do PESQ com ganho pitch adicional	308,983	426,799	317,365	429,932
3	somatório acumulado de frames do PESQ com escolha do melhor valor	311,001	433,666	329,842	435,92
4	% de frames garantidos com a aplicação do padrão de avaliação perceptual proposto	72,41%	52,94%	55,56%	61,65%
5	% de frames melhorados com a aplicação do padrão de avaliação perceptual proposto	44,83%	46,32%	50,00%	50,38%
6	% de ganho de pitch aplicado, valor escolhido pela análise de ganho para os oradores	1,54%	2,86%	10,00%	1,82%

Tabela 9-1 – Análise do ganho de pitch sobre a proposta de avaliação

Na “linha 1” da tabela acima tem o resultado da avaliação do sinal de voz original. Na “linha 2”, o resultado do sinal de voz avaliado com a heurística proposta sempre obteve melhoria no sinal analisado, conforme valores dos parâmetros da “linha 4” e da linha “5”.

Não foi implementada a heurística de controle do melhor sinal, devido a necessidade de armazenamento temporário(*buffer*) para escolha do sinal. De acordo com a tabela 9-1, “linha 3”, a melhoria de desempenho seria muito pequena para justificar a carga computacional necessária para aplicação da heurística de controle do melhor sinal.

8- O ganho aplicado ao sinal de voz para avaliação perceptual é verificado para cada orador ou conjunto de oradores, devido a variação do sinal, sendo essa análise, um item particular e importante no processo de melhoria. Conforme descrito na “linha 6”, os ganhos não possuem uma relação direta com o orador e sim com a frequência de *pitch*.

9- A avaliação do sinal de voz, tendo como origem o bloco de codificação do sinal, é uma análise a ser explorada, uma vez que o orador tenha o seu ganho adicional de *pitch* mapeado na fonte de codificação. O sistema tem condições de realizar interações no processo para melhoria do sinal de voz codificado.

Neste sentido, essa avaliação tem uma contribuição para a melhora dos sinais em equipamentos de uso pessoal como *smartphones* e *tablets*. O orador principal (usualmente o proprietário do equipamento codificador) pode ter suas características apreendidas pelo codificador, durante o uso prolongado pela contínua interação com o equipamento.

A realização dessa pesquisa foi motivada pela proposta de análise de melhoria do sinal codificado na fonte do codificador, a avaliação perceptual foi a ferramenta para validação do experimento.

9.1. Sugestões para trabalhos futuros

Como relatado acima, a análise de avaliação perceptual do bloco de origem da codificação do sinal é um tema amplo, pois existem diversos mecanismos para melhorar perceptualmente o sinal de voz codificado.

Abaixo as abordagens sugeridas para propostas de continuidade deste trabalho:

1- Estudo sobre a automatização e a verificação de ganho de *pitch*, efetuando a identificação do gênero do orador, para que o codificador possa apreender o *pitch* e efetuar uma melhoria constante.

2- Estudo de avaliação para guardar momentaneamente (*bufferizar*) o sinal de entrada, em que seja possível mapear a escolha do melhor sinal de voz, possibilitando sempre utilizar o melhor valor. Cabe ressaltar a necessidade de averiguar o impacto de atraso do sinal pelas premissas do PESQ e visto que o objetivo é a melhoria dinâmica do sinal, ou seja, avaliação em tempo real ou mais próximo possível.

3- Estudo da utilização do mapeamento das distâncias dos vetores do *codebook* para avaliação de vetores próximos, como forma de extrair outros valores de ganhos próximos. Isso para que seja possível comparar o resultado desta ação com o mapeamento de adição escalar efetuado neste trabalho, na tentativa de obter uma resposta mais assertiva quando comparada com o a heurística atual implementada.

4- Explorar novos e diferentes algoritmos de pesquisa, bem como novas heurísticas, refinando o processo de controle e tomada de decisão do codificador.

10. Referências Bibliográficas

(HAYKIN e VEEN, 1999) Simon Haykin e Barry Van Veen, **Signal and Systems**, John Wiley & Sons Inc, 1999 (Re-impressão 2007).

(FURUI, 2001) Sadaoki FURUI, **Digital speech processing, synthesis, and recognition**, Marcel Dekker, New York, 2001 (2º edition).

(O'SHAUGHNESSY, 1999) Douglas O'Shaughnessy, **Speech communications, human and machine**, Wiley-IEEE Press, Montreal, 1999 (2 edition).

(RABINER e SCHAFER, 1978) Laurence Rabiner e Ronald W. Schafer, **Digital Processing of Speech Signals**, Prentice Hall, US, 1978 .

(EMBREE e KIMBLE, 1991) Paul M. Embree e Bruce Kimble, **C language Algorithms for Digital Signal Processing**, Prentice Hall, US, 1991.

(PROAKIS e MANOLAKIS, Jhon G. Proakis e Dimitris G. Manolakis, **Digital Signal Processing, Principles, Algorithms. And Applications**, Prentice Hall, 2007 (4º edition).

(BULTHEEL e BAREL,) Adhemar Bultheel and Marc Van Barel, Paper: **Linear prediction: mathematics and engineering**. Department os Computing Science, K.U.Leuven, Belgium. Acessado em junho/2010.

<http://www.emis.de/journals/BBMS/Bulletin/bul941/BULTHEEL.PDF>

(CINNÉIDE, 2008) Alan Ó Cinnéide. Paper: **Linear Prediction The Technique, Its Solution and Application to Speech**, Dublin Institute of Technology. Acessado em junho/2010. <http://eleceng.dit.ie/papers/92.pdf>

(MAKHOUL, 1975) John Makhoul. Publicação: **Linear Prediction: A Tutorial Review**. Proceedings of the IEEE, Volume 63, nro 4, Abril/1975.

(LINDE, BUZO e GRAY, 1985) Yoseph Linde, Andrés Buzo e Robert M. Gray, **An Algorithm for Vector Quantizer Design**, IEEE Transactions on Communications, pp. 702--710, January 1980

(YNOGUTI) Carlos Alberto Ynoguti. Notas de aula: **Psicoacústica e Codificação Perceptual**, Inatel. Disponível em <http://cict.inatel.br/nova2/docentes/ynoguti/e724/Audio.pdf>

(ITU P800, 1996) ITU-T. **Methods for subjective determination of transmission quality**. International Telecommunications Union, 08/1996. (ITU-T Recommendation P.800.0).
<http://www.itu.int/rec/T-REC-P.800-199608-I/en>

(ITU P800.1, 2006) ITU-T. **Mean Opinion Score (MOS) terminology**. International Telecommunications Union, 07/2006. (ITU-T Recommendation P.800.1).
<http://www.itu.int/rec/T-REC-P.800.1/en>

(ITU P830, 1996) ITU-T. **Telephone Transmission Quality, Methods for objective and subjective Assessment of Quality**. International Telecommunications Union, 02/1996. (ITU-T Recommendation P.830). <http://www.itu.int/rec/T-REC-P.830-199602-I>

(ITU P861, 1998) ITU-T. **Objective quality measurement of telephone-band (300-3400Hz) speech codecs**. International Telecommunications Union, 02/1998. (ITU-T Recommendation P.861). <http://www.itu.int/rec/T-REC-P.861/e>

(ITU P862, 2001) ITU-T. **Perceptual Evaluation of Speech Quality**. International Telecommunications Union, 02/2001. (ITU-T Recommendation P.862.0).
<http://www.itu.int/rec/T-REC-P.862/en>

(FERNANDES, 2003) Nelson Luiz Leal. Dissertação, **Relação entre a qualidade das respostas das recomendações G723 e G729, e o comportamento da rede IP de suporte**. COPPE/UFRJ 2003. Acessado em Outubro/2010.
<http://www.ravel.ufrj.br/sites/ravel.ufrj.br/files/publicacoes/voip.pdf>

(MAGRO, 2005) Júlio César Magro. Dissertação, **Estudo da Qualidade de Voz em redes IP**. FEEC/UNICAMP. Acessado em Outubro/2010.

<http://cutter.unicamp.br/document/?code=vtls000365297>

(GOLDBERG, 2000) Randy G. Goldberg. e-book, **A Practical Handbook of Speech Coders**, Ed. Randy Goldberg, Boca Raton: CRC Press LLC, 2000

(VARY e MARTIN, 2006) Peter Vary e Rainer Martin, **Digital Speech Transmission**, Jhon Wiley & Sons Ltda, England, 2006 .

(ITU G711, 1988) ITU-T. **Pulse code modulation (PCM) of voice frequencies**. International Telecommunications Union, 11/1988. (ITU-T Recommendation G.711.0).

<http://www.itu.int/rec/T-REC-G.711-198811-I/en>

(ITU G729, 1996) ITU-T. **Coding of speech at 8 kbit/s using conjugate-structure algebraic-code excited linear-prediction**. International Telecommunications Union, 03/1996. (ITU-T Recommendation G.729. Annex A). <http://www.itu.int/rec/T-REC-G.729-199603-S/en>

(ITU G729a, 1996) ITU-T. **Reduce complexity 8 kbit/s CS-ACELP speech codec**. International Telecommunications Union, 11/1996. (ITU-T Recommendation G.729.0 Annex A). <http://www.itu.int/rec/T-REC-G.729-199611-S!AnnA/en>

(PREGO e NETTO, 2008) Tiago M Prego e Sergio L. Netto. Paper, **Algoritmo de Busca eficiente no Dicionário Adaptativo para Codec ITU-T G.729**. XXVI Simpósio Brasileiro de Telecomunicações – SBrT , 09/2008. Acessado em Setembro/2010.

<http://www02.lps.ufrj.br/~sergioln/papers/BC16.pdf>

(GOMES, 2008) Alexandre Guazzi Gomes. Dissertação, **Codificador de Voz CELPS: Versão estruturada e detecção de silêncio**. LPS- Departamento de Eletrônica e Computação –UFRJ, 12/2008. Acessado em Setembro/2010.

<http://www02.lps.ufrj.br/~sergioln/theses/bsc20alexandreguazzi.pdf>

(CHU, 2003) WAI C. CHU, **Speech coding algorithms. Foundation and evolution of standardized coders**, Mobile Media Laboratory/DoCoMo USA Labs, San Jose, California, 2003.

(KONDOZ, 2003) A.M. KONDOZ, **Coding for low Bit rate Communication Systems**,a University of Surrey, UK, New York, John Wiley & Sons, 2004 .

(SCHROEDER e ATAL, 1985) M. R. Schroeder and B. S. Atal, **Code-excited linear prediction (CELP): high-quality speech at very low bit rates**, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 10, pp. 937-940, 1985.

(Jayant e Noll, 1984) N.S. Jayant e P.Noll, **Digital Coding of Waveforms, Principles and Applications to Speech and Video**, Prentice Hall, Englewood Cliffs, NJ, 1984.

(Woodard, J. P., 1995) Woodard, J. P., **Commonly speech codecs.**, Acessado em Setembro/2011 http://wwwmobile.ecs.soton.ac.uk/speech_codecs/common_classes.html

(Gomes, A. G., 2008) Alexandre Guazzi Gomes, Dissertação **Codificador de Voz CELPS: Versão estruturada e detecção de silêncio**, DEL UFRJ, acessado em Dezembro 2010.

(Audacity1.3.12,2010) Audacity Software de Avaliação de Sinais de Audio. **Release of Audacity is 1.3.12 (Beta)**. Acessado em Junho/2010.
<http://audacity.sourceforge.net/download/>

11. Anexo

11.1. Arquivos de testes do Banco de dados Timit

Arquivo1 de teste dos oradores masculino: Orador_masculino8k.wav

- Arquivo 01: sx327.wav
Texto: Al received a joint appointment in the biology and the engineering departments.
- Arquivo 02: si824.wav
Texto: Positive results start when it goes towards the hand you use to make your mark.
- Arquivo 03: si045.wav
Texto: It offered to surrender its right to exclusive trade, but asked an indemnity
- Arquivo 04: si918.wav
Texto: We know that actors can learn to portray a wide variety of character roles
- Arquivo 05: si675.wav
Texto: Both eventualities are possible logically, but practically they are impossible
- Arquivo 06: si631.wav
Texto: Program note reads as follows: take hands; this urgent visage beckons us

Arquivo1 de teste dos oradores feminino: Orador_feminino8k.wav

- Arquivo 01: sx451.wav
Texto: The thick elm forest was nearly overwhelmed by Dutch Elm Disease
- Arquivo 02: sx333.wav
Texto: An adult male baboon's teeth are not suitable for eating shellfish
- Arquivo 03: si434.wav
Texto: Conceivably the submarine defense problem can be solved by sufficient forces
- Arquivo 04: si418.wav
Texto: Personal predispositions tend to blunt the ear and, in turn, the voice as well
- Arquivo 05: si836.wav
Texto: No manufacturer has taken the initiative in pointing out the cost involved
- Arquivo 06: si788.wav
Texto: We have become amateur insurance experts and fine-feathered yard birds

Arquivo2 de teste dos oradores masculino: Orador_masculino8k2.wav

Arquivo 01: sx449.wav
Texto: Ralph prepared red snapper with fresh lemon sauce for dinner

Arquivo 02: si1079.wav
Texto: Boys and men go along the riverbank or to the alcoves in the top arcade

Arquivo 03: si2305.wav
Texto: As you can count on me to do the same

Arquivo 04: si2004.wav
Texto: She had jumped away from his shy touch like a cat confronted by a sidewinder

Arquivo 05: si1374.wav
Texto: This leaves the ordering of entries variable

Arquivo 06: si774.wav
Texto: Get copper or earthenware mugs that keep beer chilled or soup hot.

Arquivo2 de teste dos oradores feminino: Orador_feminino8k2.wav

Arquivo 01: sx397.wav
Texto: Tim takes Sheila to see movies twice a week.

Arquivo 02: sx117.wav
Texto: The mango and the papaya are in a bowl

Arquivo 03: si1377.wav
Texto: As these maladies overlap, so must the cure.

Arquivo 04: si2007.wav
Texto: To his puzzlement, there suddenly was no haze

Arquivo 05: si614.wav
Texto: This big, flexible voice with uncommon range has been superbly disciplined.

Arquivo 06: si2048.wav
Texto: The kid has no manners, boys

12. Apêndices

12.1. Apêndice A - Resultado do Arquivo1 de Oradores masculinos

```
C:\>cd project_test

C:\project_test>G729.exe Orador_masculino8k.wav masculino_teste1
Inicializa arq1= 1
Inicializa arq2= 1

***** ITU G.729A 8 Kbits/seg Speech Coder *****
----- Simulacao codec com PESQ -----
----- Versao 1.0 (Release 3, TESTE 21_07_2011) -----
Entrada do arquivo de teste: Orador_masculino8k.wav
Saida do arquivo bitstream: masculino_teste1
vai declarar variaveis novas
Pÿ%
Orador_masculino8k.wav
Orador_masculino8k.wavdecoded
masculino_teste1frame
tamanho do arquivo: 4480 pesq: 2.210000
tamanho do arquivo: 4480 pesq: 3.023000
tamanho do arquivo: 4480 pesq: 3.416000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.552000
tamanho do arquivo: 4480 pesq: 2.708000
controle= 1
tamanho do arquivo: 4480 pesq: 3.022000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.595000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 2.970000
controle= 1
tamanho do arquivo: 4480 pesq: 3.207000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.785000
ganho01_escrita1= 65
```

tamanho do arquivo: 4480 pesq: 3.782000
controle= 1
tamanho do arquivo: 4480 pesq: 4.025000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 4.123000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.834000
controle= 1
tamanho do arquivo: 4480 pesq: 3.821000
controle= 2
tamanho do arquivo: 4480 pesq: 3.767000
controle= 3
sem ACRESCIMO pitch= 1
tamanho do arquivo: 4480 pesq: 2.931000
controle= 1
tamanho do arquivo: 4480 pesq: 3.522000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.283000
controle= 1
tamanho do arquivo: 4480 pesq: 3.934000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.808000
controle= 1
tamanho do arquivo: 4480 pesq: 3.943000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 2.916000
tamanho do arquivo: 4480 pesq: 2.851000
controle= 1
tamanho do arquivo: 4480 pesq: 3.761000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.437000
controle= 1
tamanho do arquivo: 4480 pesq: 3.952000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.499000
controle= 1
tamanho do arquivo: 4480 pesq: 3.880000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 4.061000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.221000
controle= 1
tamanho do arquivo: 4480 pesq: 3.994000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.164000
controle= 1
tamanho do arquivo: 4480 pesq: 3.171000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.919000
ganho01_escrita1= 65

tamanho do arquivo: 4480 pesq: 3.288000
controle= 1
tamanho do arquivo: 4480 pesq: 3.901000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.540000
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.932000
tamanho do arquivo: 4480 pesq: 3.391000
controle= 2
tamanho do arquivo: 4480 pesq: 3.604000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 4.275000
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.943000
tamanho do arquivo: 4480 pesq: 3.649000
controle= 1
tamanho do arquivo: 4480 pesq: 3.574000
controle= 2
tamanho do arquivo: 4480 pesq: 3.475000
controle= 3
sem ACRESCIMO pitch= 1
tamanho do arquivo: 4480 pesq: 3.192000
controle= 1
tamanho do arquivo: 4480 pesq: 2.716000
controle= 2
tamanho do arquivo: 4480 pesq: 2.934000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 4.024000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.438000
controle= 1
tamanho do arquivo: 4480 pesq: 3.861000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.114000
controle= 1
tamanho do arquivo: 4480 pesq: 3.628000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.662000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.192000
controle= 1
tamanho do arquivo: 4480 pesq: 3.871000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.520000
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.026000
tamanho do arquivo: 4480 pesq: 3.937000
ganho01_escrita1= 65

tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.115000
tamanho do arquivo: 4480 pesq: 4.050000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 4.124000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 4.119000
controle= 1
tamanho do arquivo: 4480 pesq: 3.606000
controle= 2
tamanho do arquivo: 4480 pesq: 2.947000
controle= 3
sem ACRESCIMO pitch= 1
tamanho do arquivo: 4480 pesq: 4.067000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.630000
controle= 1
tamanho do arquivo: 4480 pesq: 2.289000
controle= 2
tamanho do arquivo: 4480 pesq: 4.068000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.708000
controle= 1
tamanho do arquivo: 4480 pesq: 3.990000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 2.893000
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.767000
tamanho do arquivo: 4480 pesq: 4.149000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.728000
controle= 1
tamanho do arquivo: 4480 pesq: 3.607000
controle= 2
tamanho do arquivo: 4480 pesq: 4.145000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.914000
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 2.985000
tamanho do arquivo: 4480 pesq: 3.818000
ganho01_escrita1= 65
tamanho do arquivo: 4480 pesq: 3.400000
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.840000
tamanho do arquivo: 4480 pesq: 3.695000
controle= 2
tamanho do arquivo: 4480 pesq: 3.749000
ganho01_escrita1= 65

tamanho do arquivo: 4480 pesq: 0.000000

tamanho do arquivo: 4480 pesq: 2.963000

tamanho do arquivo: 4480 pesq: 3.843000

ganho01_escrita1= 65

total frames:2783

Somatorio das notas dos frames PESQ : 308.983000

C:\project_test>

12.2. Apêndice B - Resultado do Arquivo1 de Oradores femininos

```
C:\project_test>G729.exe Orador_feminino8k.wav feminino_teste1
```

```
Inicializa arq1= 1
```

```
Inicializa arq2= 1
```

```
***** ITU G.729A 8 Kbits/seg Speech Coder *****
```

```
----- Simulacao codec com PESQ -----
```

```
----- Versao 1.0 (Release 3, TESTE 21_07_2011) -----
```

```
Entrada do arquivo de teste: Orador_feminino8k.wav
```

```
Saida do arquivo bitstream: feminino_teste1
```

```
vai declarar variaveis novas
```

```
ÿ%
```

```
Orador_feminino8k.wav
```

```
Orador_feminino8k.wavdecoded
```

```
feminino_teste1frame
```

```
tamanho do arquivo: 4480 pesq: 0.000000
```

```
tamanho do arquivo: 4480 pesq: 0.000000
```

```
tamanho do arquivo: 4480 pesq: 3.205000
```

```
tamanho do arquivo: 4480 pesq: 3.248000
```

```
tamanho do arquivo: 4480 pesq: 3.819000
```

```
ganho01_escrita1= 10
```

```
tamanho do arquivo: 4480 pesq: 3.795000
```

```
ganho01_escrita1= 10
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 3.416000
```

```
ganho01_escrita1= 10
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 4.089000
```

```
ganho01_escrita1= 10
```

```
tamanho do arquivo: 4480 pesq: 3.519000
```

```
ganho01_escrita1= 10
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 2.984000
```

```
ganho01_escrita1= 10
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 0.000000
```

```
ganho01_escrita1= 10
```

```
tamanho do arquivo: 4480 pesq: 3.737000
```

```
tamanho do arquivo: 4480 pesq: 0.000000
```

```
ganho01_escrita1= 10
```

tamanho do arquivo: 4480 pesq: 3.471000
tamanho do arquivo: 4480 pesq: 4.104000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 2.933000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 2.723000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 2.643000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 2.732000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.785000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 4.088000
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.901000
tamanho do arquivo: 4480 pesq: 3.355000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.818000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.240000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 4.171000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.710000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.488000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.919000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.093000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.800000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.547000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.647000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.014000

ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.211000
tamanho do arquivo: 4480 pesq: 3.448000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.200000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.052000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 2.964000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.591000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.975000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 2.890000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 2.997000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.671000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.364000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.355000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.784000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.574000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.154000
tamanho do arquivo: 4480 pesq: 3.488000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.741000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.634000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.921000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.241000

ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.500000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.903000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.785000
tamanho do arquivo: 4480 pesq: 3.690000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.868000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.848000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.686000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 4.034000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.458000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 4.176000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.183000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 2.798000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.422000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.193000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.378000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.790000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.802000
tamanho do arquivo: 4480 pesq: 4.061000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.328000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.658000

ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.958000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.957000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.605000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.707000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.618000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.023000
tamanho do arquivo: 4480 pesq: 3.405000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.613000
tamanho do arquivo: 4480 pesq: 3.698000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.154000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 2.830000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.949000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.820000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.542000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 3.566000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.594000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.216000
ganho01_escrita1= 10
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 10
tamanho do arquivo: 4480 pesq: 3.393000
tamanho do arquivo: 4480 pesq: 3.810000
ganho01_escrita1= 10

tamanho do arquivo: 4480 pesq: 4.097000
ganho01_escrita1= 10
total frames:2937

Somatorio das notas dos frames PESQ : 317.365000

C:\project_test>

12.3. Apêndice C - Resultado do Arquivo2 de Oradores masculinos

```
C:\project_test1>G729.exe orador_masculino8k2.wav masculino_teste2
```

```
Inicializa arq1= 1
```

```
Inicializa arq2= 1
```

```
***** ITU G.729A 8 Kbits/seg Speech Coder *****
```

```
----- Simulacao codec com PESQ -----
```

```
----- Versao 1.0 (Release 3, TESTE 21_07_2011) -----
```

```
Entrada do arquivo de teste: orador_masculino8k2.wav
```

```
Saida do arquivo bitstream: masculino_teste2
```

```
vai declarar variaveis novas
```

```
Pÿ%
```

```
orador_masculino8k2.wav
```

```
orador_masculino8k2.wavdecoded
```

```
masculino_teste2frame
```

```
tamanho do arquivo: 4480 pesq: 2.933000
```

```
tamanho do arquivo: 4480 pesq: 2.324000
```

```
tamanho do arquivo: 4480 pesq: 2.960000
```

```
ganho01_escrita1= 35
```

```
tamanho do arquivo: 4480 pesq: 2.555000
```

```
ganho01_escrita1= 35
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 3.074000
```

```
ganho01_escrita1= 35
```

```
tamanho do arquivo: 4480 pesq: 3.559000
```

```
ganho01_escrita1= 35
```

```
tamanho do arquivo: 4480 pesq: 2.797000
```

```
ganho01_escrita1= 35
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 3.202000
```

```
ganho01_escrita1= 35
```

```
tamanho do arquivo: 4480 pesq: 2.389000
```

```
ganho01_escrita1= 35
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 3.918000
```

```
ganho01_escrita1= 35
```

```
tamanho do arquivo: 4480 pesq: 2.974000
```

```
ganho01_escrita1= 35
```

```
controle= 1
```

tamanho do arquivo: 4480 pesq: 3.144000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.532000
tamanho do arquivo: 4480 pesq: 3.254000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.182000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.603000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 4.066000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.553000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.216000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.592000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.368000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.523000
tamanho do arquivo: 4480 pesq: 3.753000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.683000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.604000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.590000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.711000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.517000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.578000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.879000
tamanho do arquivo: 4480 pesq: 3.668000

ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.714000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.603000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.540000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.261000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.348000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.214000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.131000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.645000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.583000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.889000
tamanho do arquivo: 4480 pesq: 2.807000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.362000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.354000
tamanho do arquivo: 4480 pesq: 2.942000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.933000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.091000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.993000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.919000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.882000

tamanho do arquivo: 4480 pesq: 2.749000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.947000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.714000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.233000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.222000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.829000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.304000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.934000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.562000
tamanho do arquivo: 4480 pesq: 3.507000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.567000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.162000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.287000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.298000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.508000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.196000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.873000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.990000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.044000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.579000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.806000
ganho01_escrita1= 35

controle= 1
tamanho do arquivo: 4480 pesq: 3.582000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.462000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.944000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.703000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.745000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.773000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.537000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.083000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.760000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.888000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.978000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.455000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.796000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.985000
tamanho do arquivo: 4480 pesq: 3.345000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.987000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.769000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 4.045000
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.617000
tamanho do arquivo: 4480 pesq: 3.285000

ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.599000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.355000
tamanho do arquivo: 4480 pesq: 3.634000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.498000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.813000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 1.469000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.043000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 1.606000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.349000
tamanho do arquivo: 4480 pesq: 3.781000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.355000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.616000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.979000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.714000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.937000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.261000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.033000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.346000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.312000
ganho01_escrita1= 35
controle= 1

tamanho do arquivo: 4480 pesq: 3.100000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.796000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.209000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.969000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.399000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.200000
tamanho do arquivo: 4480 pesq: 2.089000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.120000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.910000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.563000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.610000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.172000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.316000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.704000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.226000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 2.471000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.544000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 4.126000
tamanho do arquivo: 4480 pesq: 3.804000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.544000

ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.608000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.291000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 3.549000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 2.985000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 1.686000
ganho01_escrita1= 35
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 35
tamanho do arquivo: 4480 pesq: 3.146000
total frames:4275

Somatorio das notas dos frames PESQ : 426.799000

C:\project_test1>

12.4. Apêndice D - Resultado do Arquivo2 de Oradores femininos

```
C:\project_test1>G729.exe Orador_feminino8k2.wav feminino2_test
```

```
Inicializa arq1= 1
```

```
Inicializa arq2= 1
```

```
***** ITU G.729A 8 Kbits/seg Speech Coder *****
```

```
----- Simulacao codec com PESQ -----
```

```
----- Versao 1.0 (Release 3, TESTE 21_07_2011) -----
```

```
Entrada do arquivo de teste: Orador_feminino8k2.wav
```

```
Saida do arquivo bitstream: feminino2_test
```

```
vai declarar variaveis novas
```

```
ÿ%
```

```
Orador_feminino8k2.wav
```

```
Orador_feminino8k2.wavdecoded
```

```
feminino2_testframe
```

```
tamanho do arquivo: 4480 pesq: 2.762000
```

```
tamanho do arquivo: 4480 pesq: 2.764000
```

```
tamanho do arquivo: 4480 pesq: 3.335000
```

```
ganho01_escrita1= 55
```

```
tamanho do arquivo: 4480 pesq: 3.521000
```

```
ganho01_escrita1= 55
```

```
tamanho do arquivo: 4480 pesq: 3.403000
```

```
ganho01_escrita1= 55
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 3.256000
```

```
ganho01_escrita1= 55
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 3.794000
```

```
ganho01_escrita1= 55
```

```
tamanho do arquivo: 4480 pesq: 3.102000
```

```
ganho01_escrita1= 55
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 2.758000
```

```
ganho01_escrita1= 55
```

```
controle= 1
```

```
tamanho do arquivo: 4480 pesq: 3.127000
```

```
ganho01_escrita1= 55
```

```
tamanho do arquivo: 4480 pesq: 2.496000
```

```
ganho01_escrita1= 55
```

controle= 1
tamanho do arquivo: 4480 pesq: 3.609000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.796000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.356000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.923000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.217000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.319000
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.705000
tamanho do arquivo: 4480 pesq: 3.699000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.357000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.653000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.097000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.544000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.213000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.215000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.431000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.956000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.892000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.034000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.703000
ganho01_escrita1= 55
controle= 1

tamanho do arquivo: 4480 pesq: 3.737000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.028000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.191000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.253000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.163000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.106000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.947000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.644000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.531000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.819000
tamanho do arquivo: 4480 pesq: 3.850000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.289000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.953000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.681000
tamanho do arquivo: 4480 pesq: 2.872000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.719000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.029000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.635000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.687000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.261000

ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.672000
tamanho do arquivo: 4480 pesq: 2.623000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.394000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.970000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.304000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.750000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 0.000000
tamanho do arquivo: 4480 pesq: 3.284000
tamanho do arquivo: 4480 pesq: 3.118000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.501000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.445000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.718000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.623000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.825000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.029000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.326000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.493000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.881000
tamanho do arquivo: 4480 pesq: 3.605000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.412000
ganho01_escrita1= 55

controle= 1
tamanho do arquivo: 4480 pesq: 3.014000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.671000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.681000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.474000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.785000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.137000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.511000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.202000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.319000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.142000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.154000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.661000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 1.941000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.907000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.554000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.674000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.527000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.619000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.571000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.554000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.061000
ganho01_escrita1= 55

controle= 1
tamanho do arquivo: 4480 pesq: 3.301000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.645000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.198000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.741000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.123000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.644000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.281000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.787000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.298000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.578000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.653000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.535000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.267000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.844000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.066000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.768000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.215000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.655000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.525000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.987000
ganho01_escrita1= 55
controle= 1

tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.072000
tamanho do arquivo: 4480 pesq: 3.007000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.991000
tamanho do arquivo: 4480 pesq: 2.863000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.733000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.136000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.408000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.631000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.335000
tamanho do arquivo: 4480 pesq: 4.020000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.480000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.805000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.234000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.719000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.818000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 2.982000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 3.229000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.852000
ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.605000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 0.000000

ganho01_escrita1= 55
tamanho do arquivo: 4480 pesq: 3.646000
tamanho do arquivo: 4480 pesq: 3.436000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.996000
ganho01_escrita1= 55
controle= 1
tamanho do arquivo: 4480 pesq: 2.883000
ganho01_escrita1= 55
controle= 1
total frames:4077

Somatorio das notas dos frames PESQ : 428.556000

C:\project_test1>