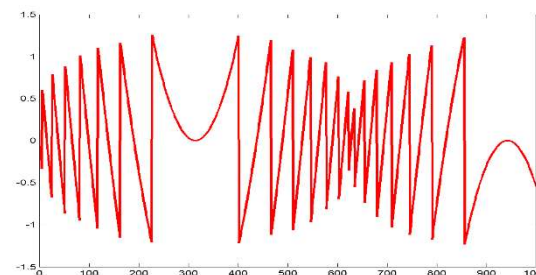


Compressão de Áudio MPEG AAC

PTC3547 – Codificação e
Transmissão Multimídia
Guido Stolfi
EPUSP – 2025

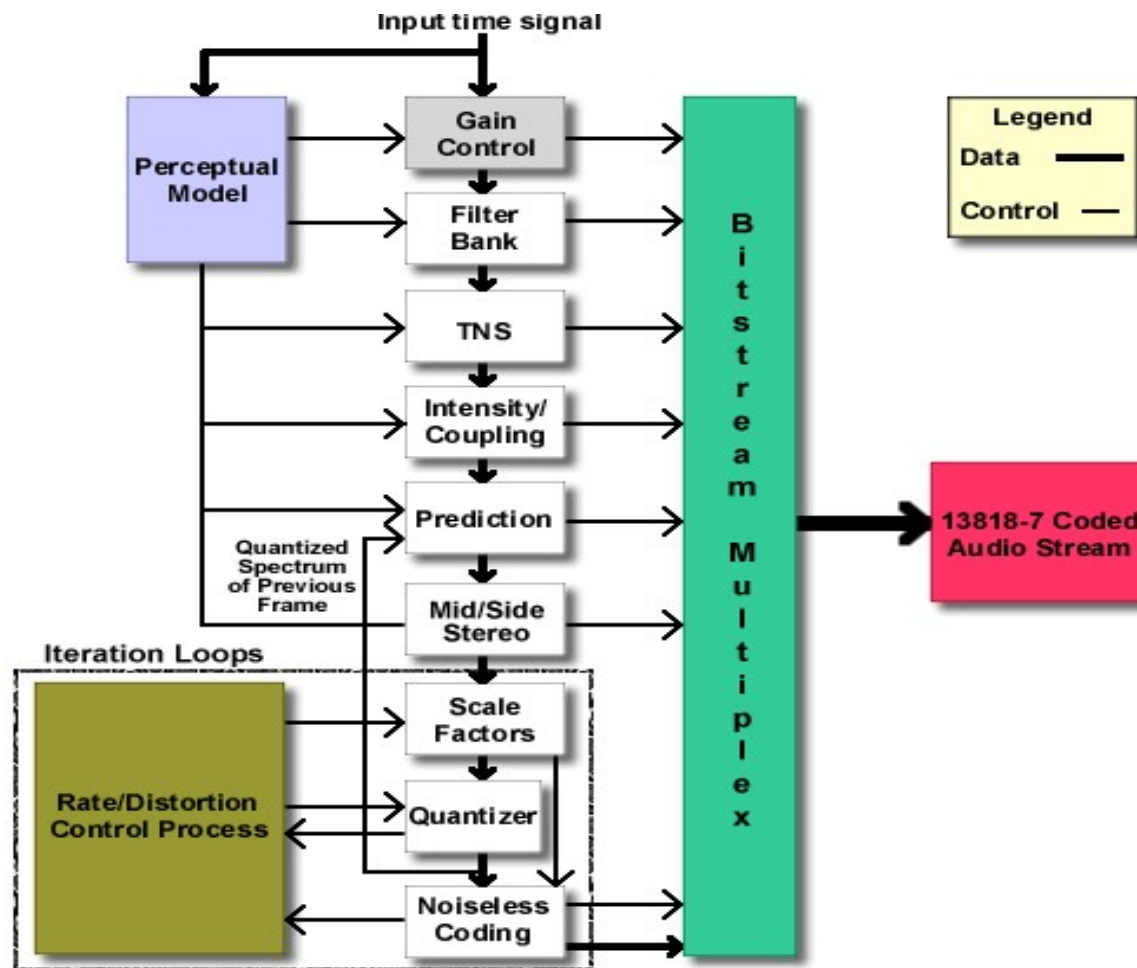


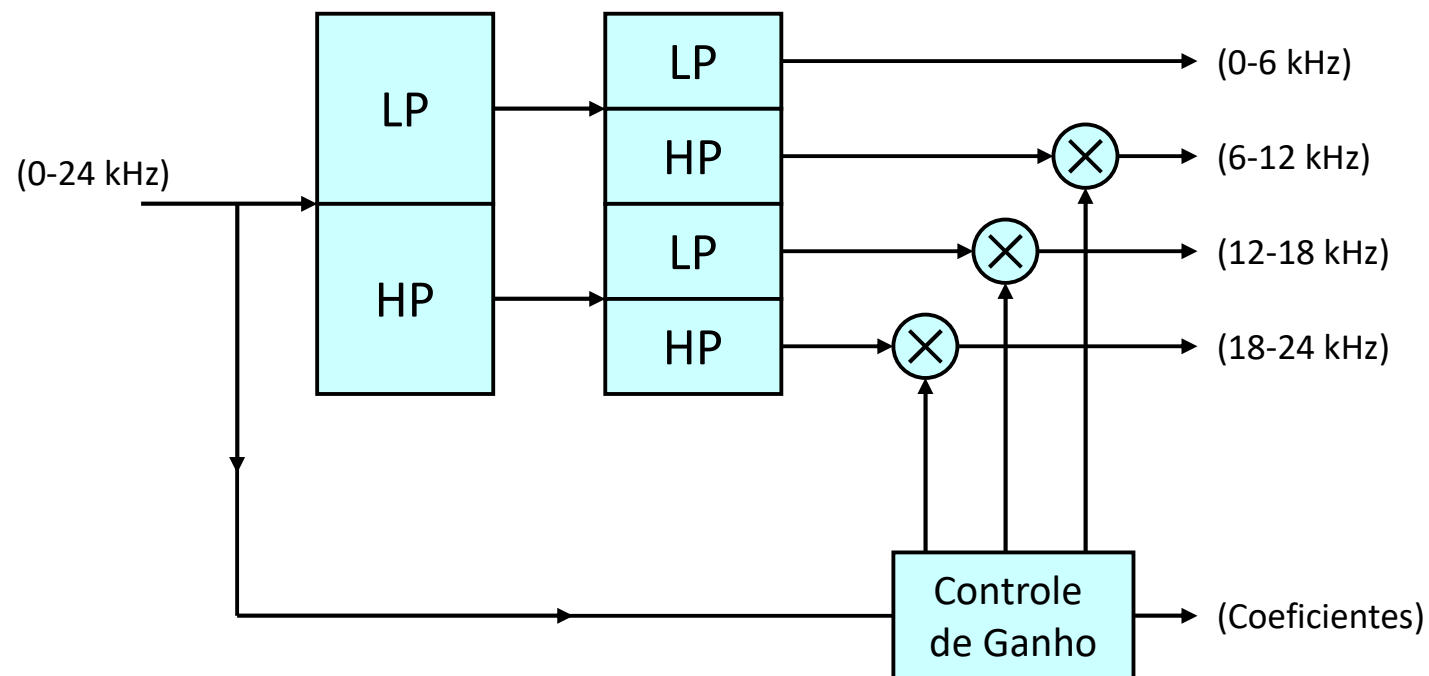
- MPEG-2 AAC – *Advanced Audio Coding*
 - Blocos Funcionais
- MPEG-4 AAC – Ferramentas Adicionais
 - SBR, CELP, HVXC, etc.
 - Conceitos Chave
 - Requisitos Computacionais
- Síntese Musical
- MIDI – *Musical Instruments Digital Interface*

- Derivado do MPEG-2 Nível III
- 1 a 48 canais de áudio
- 8 a 96 kHz de amostragem
- Patentes: AT&T, Sony, Dolby, Fraunhofer
- Usado no sistema ISDB, XM Radio e DRM
 - Low Complexity AAC (AAC-LC)
 - High Efficiency (HE-AAC ou “AACPlus”)
 - Scalable Sample Rate (AAC-SSR)
 - Bit Sliced Arithmetic Coding (BSAC)

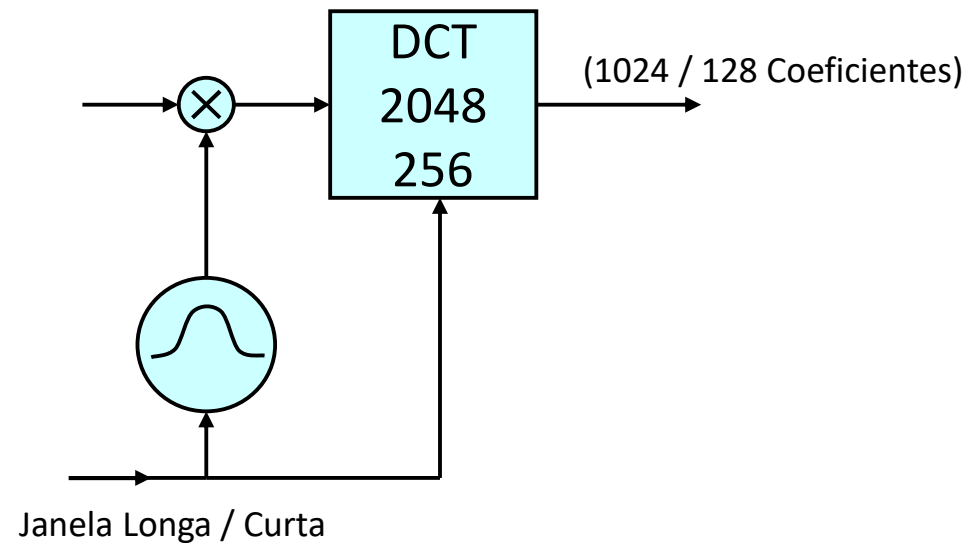
- Banco de filtros implementado por DCT modificada, 1024 pontos (Processamento T / F)
- Não é compatível com MPEG-1
- Utiliza preditor linear para pré-codificação
- Efetua “Noise Shaping” adaptativo
- Pode utilizar replicação espectral (SBR)
- Otimização da codificação estatística dos símbolos transmitidos
 - Resultado: redução de 50% na taxa de bits

Diagrama de Blocos da Codificação AAC

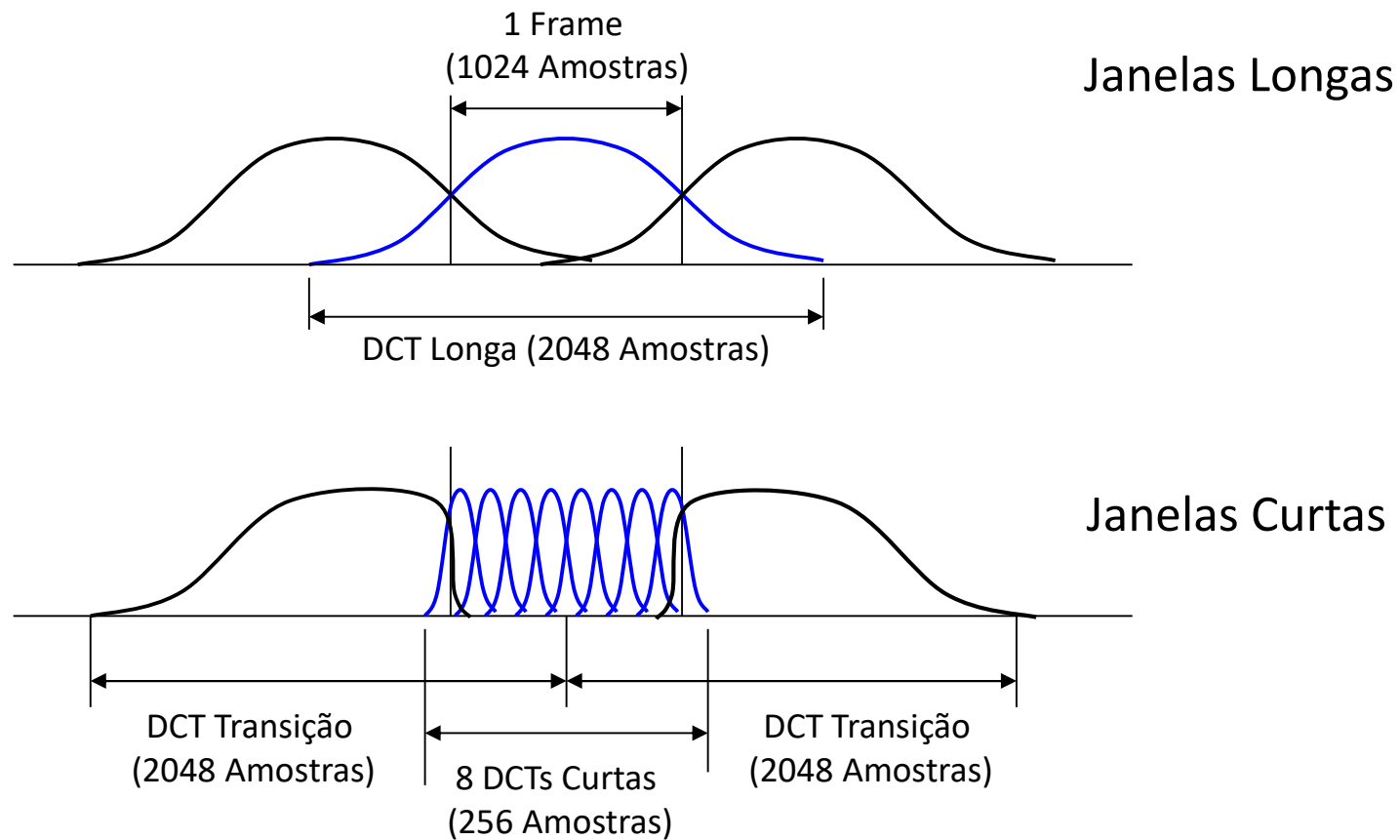


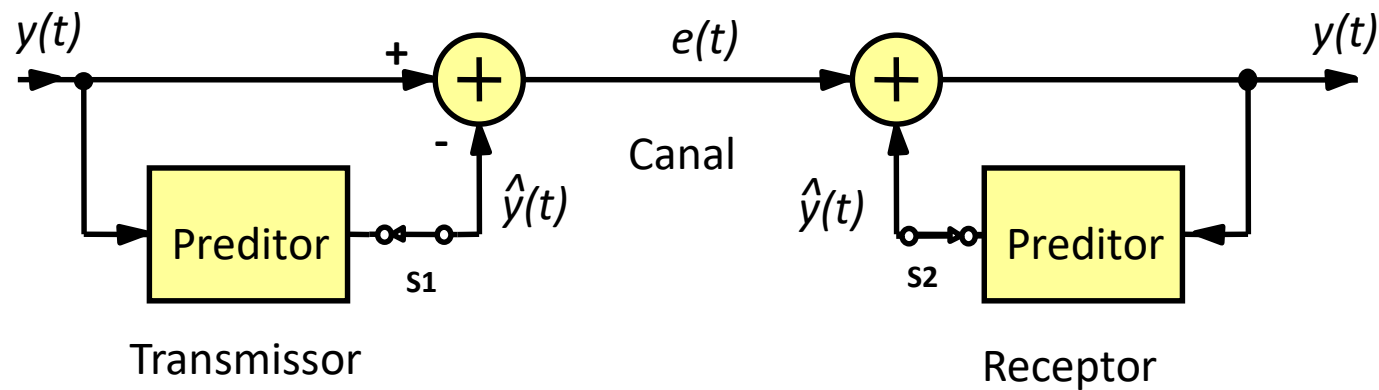


Banco de Filtros: DCT com Janela Temporal



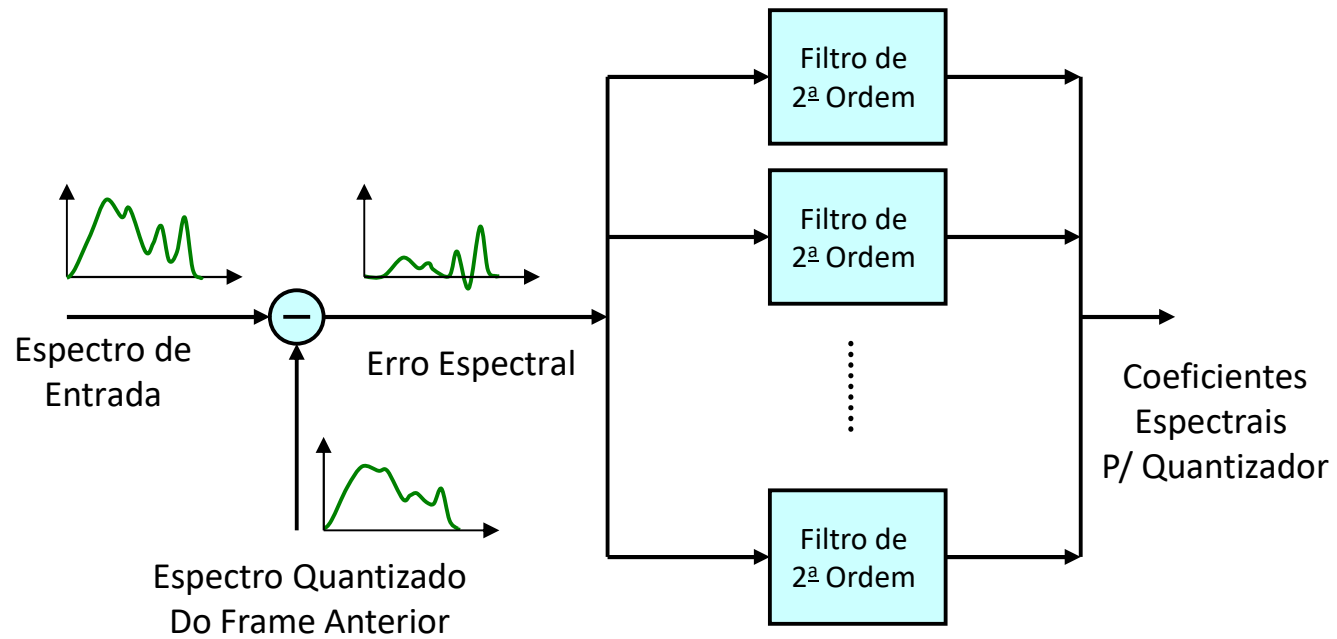
Janelas Temporais Superpostas



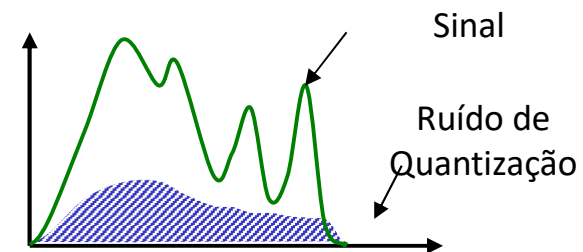
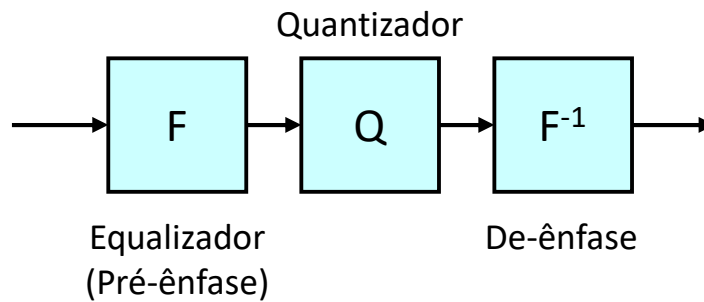
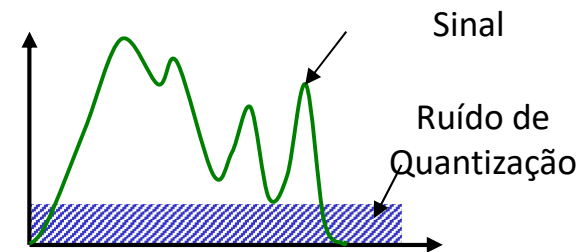
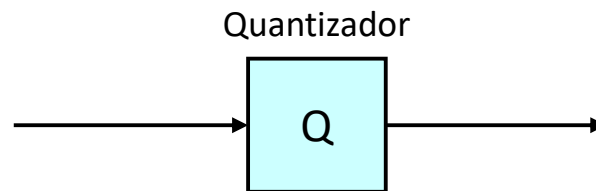


- Eficácia do preditor reduz a energia do sinal transmitido $e(t)$
- Chaves $S1$ e $S2$ devem ser abertas em instantes apropriados

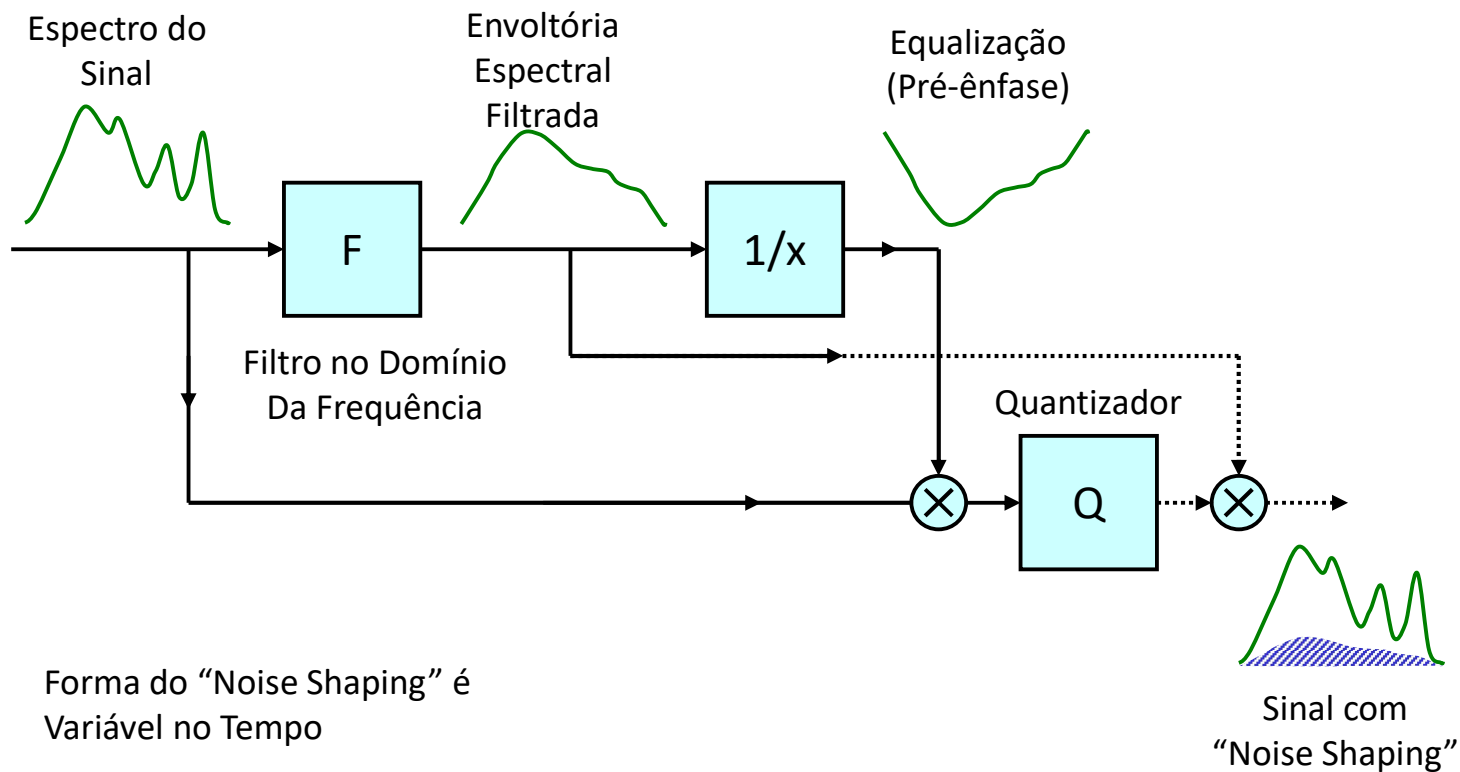
Preditor Espectral



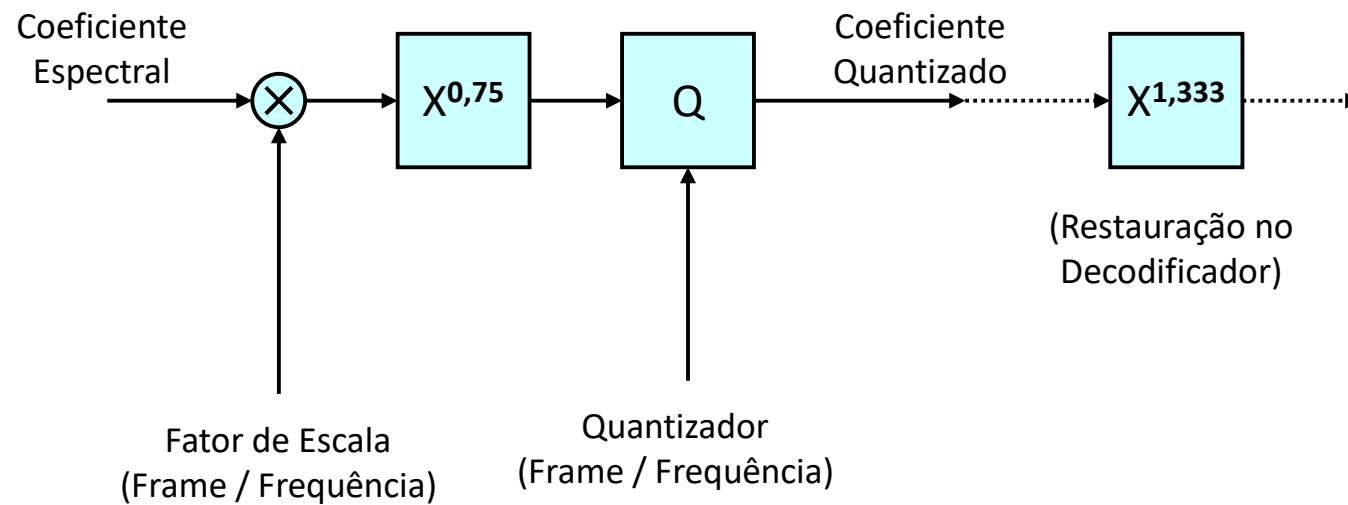
Noise Shaping



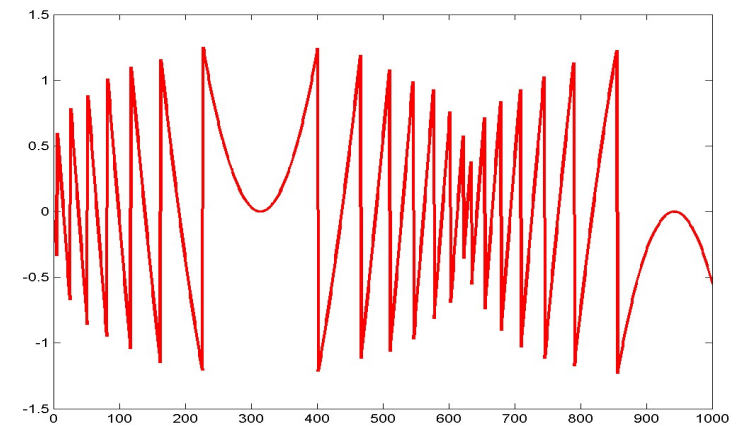
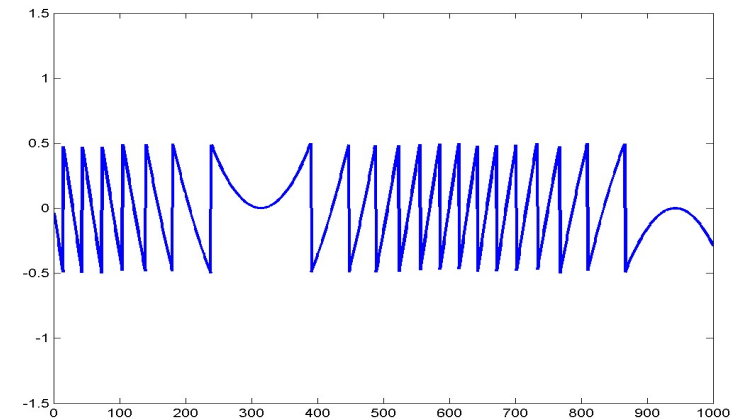
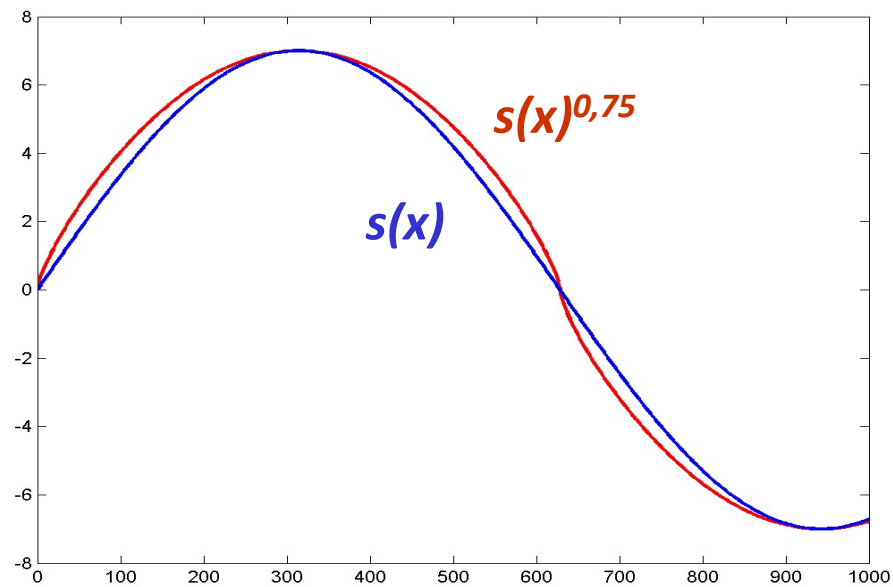
TNS – Temporal Noise Shaping



Quantizador Não Linear



Quantizador Não Linear



- Fatores de Escala:
 - Codificação Diferencial em relação ao Frame anterior
 - Resolução de 1,5 dB
- Quantizador:
 - Selecionado com resolução de 1,5 dB
- Fator de Escala + Quantizador:
 - Codificado como um símbolo conjunto
 - Vários dicionários (Huffmann) selecionáveis para codificação
- Coeficientes Quantizados:
 - Agrupados (2 ou 4) e codificados com um de 11 dicionários, válidos por faixas de frequência.

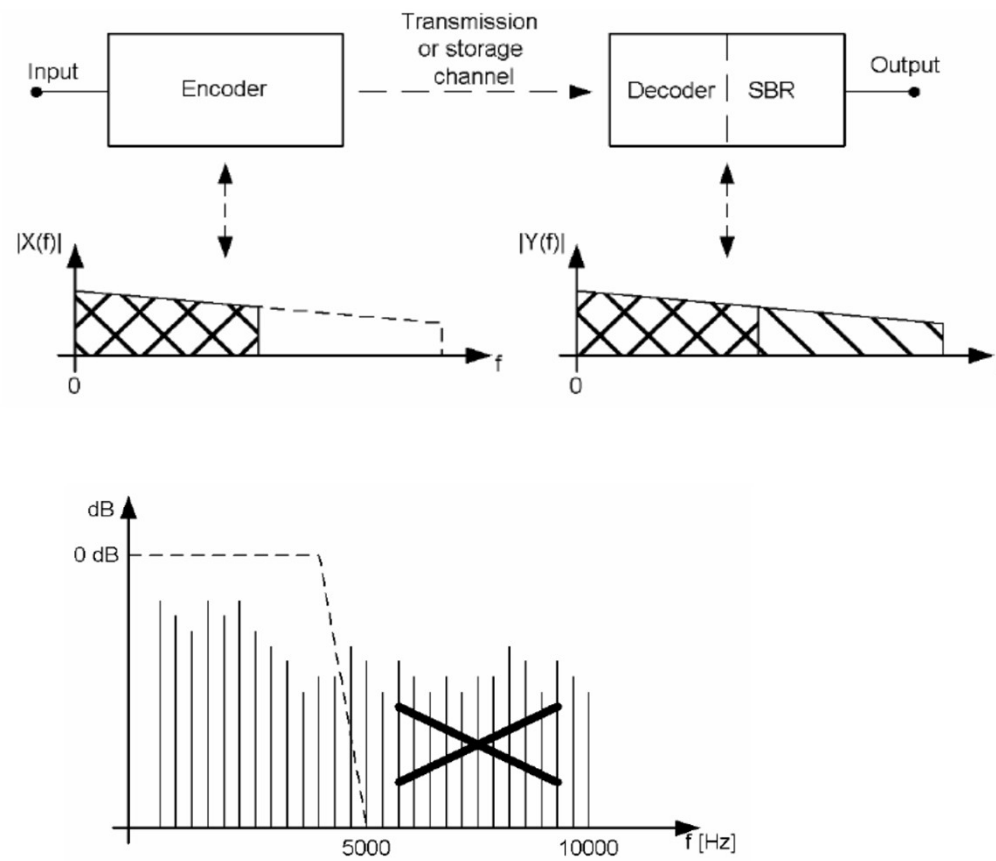
MPEG-4 AAC

Ferramentas Adicionais

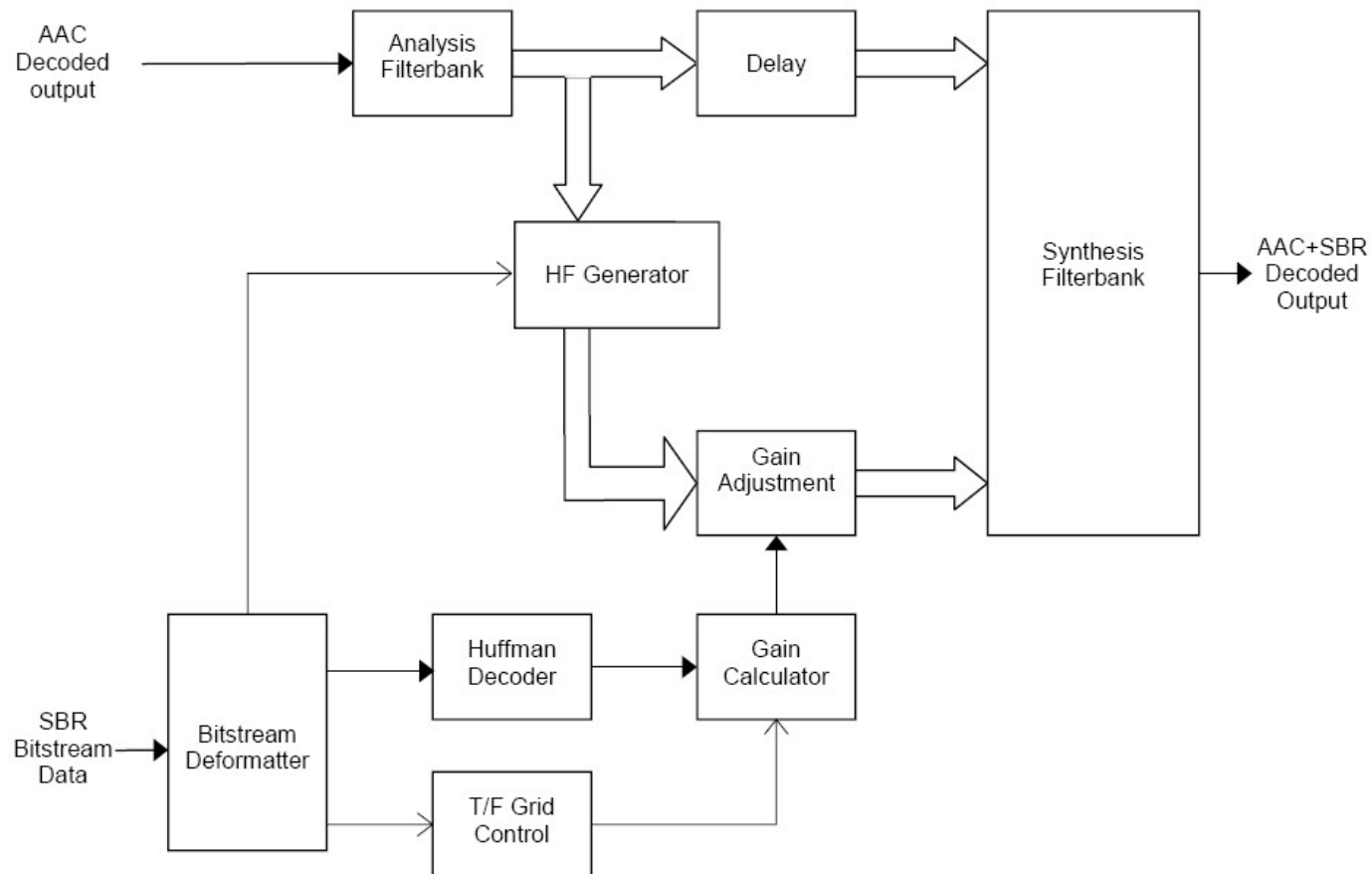
- Codificação de Áudio Natural:
 - MPEG-2 AAC
 - SBR: Spectral Band Replication
 - CELP: Code Excited Linear Prediction
- Codificação paramétrica:
 - HVXC: Harmonic Vector Excitation Coding
 - HILN: Harmonic, Individual Line + Noise
- Áudio Sintético:
 - M-TTS: MPEG Text To Speech
 - SAOL: Structured Audio Orchestra Language
 - SASL: Structured Audio Score Language
 - SASBF: Structured Audio Sample Bank Format
 - MIDI: Musical Instruments Digital Interface

- Norma ABNT NBR 15602-2
- Codificação MPEG-4 AAC (ISO/IEC 14496-3):
 - LC (Perfil básico Low Complexity), níveis L2 e L4;
 - HE (High Efficiency, LC + SBR), níveis L2 e L4;
 - HE + PS (Parametric Stereo), nível L2.

SBR (Spectral Band Replication)

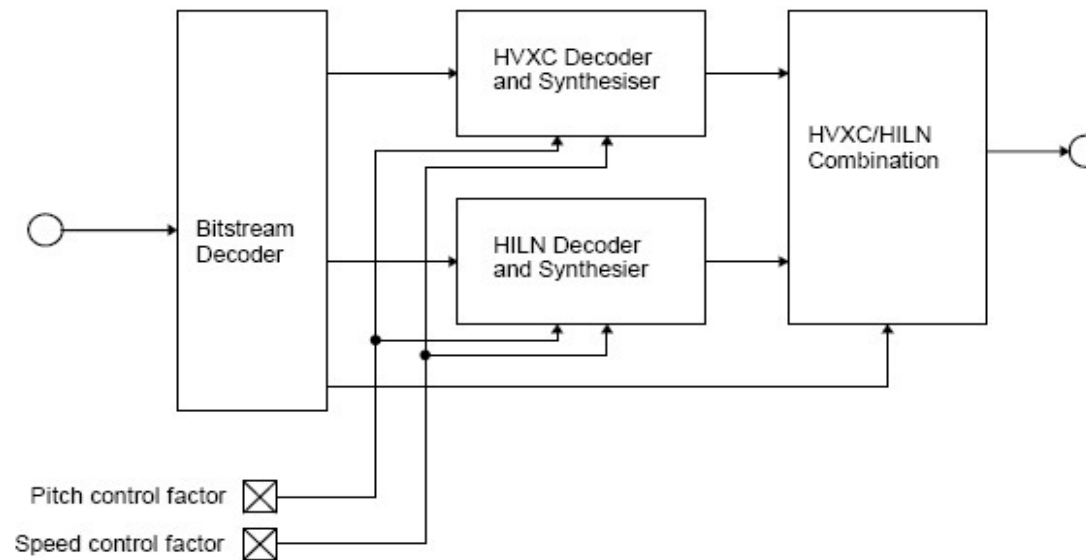


Decodificação SBR (Spectral Band Replication)



- HVXC – *Harmonic Vector Excitation Coding*
 - Codificação de Voz, 1.2 a 4 kb/s
 - Amostragem: 8 kHz
 - Processamento: Frames de 20 ms (160 amostras)
- HILN – *Harmonic and Individual Line + Noise*
 - Voz e música, > 4 kb/s
 - Usa PNS (*Perceptual Noise Substitution*)

MPEG-4 AAC – Decodificador Paramétrico



Modo de Codificação:

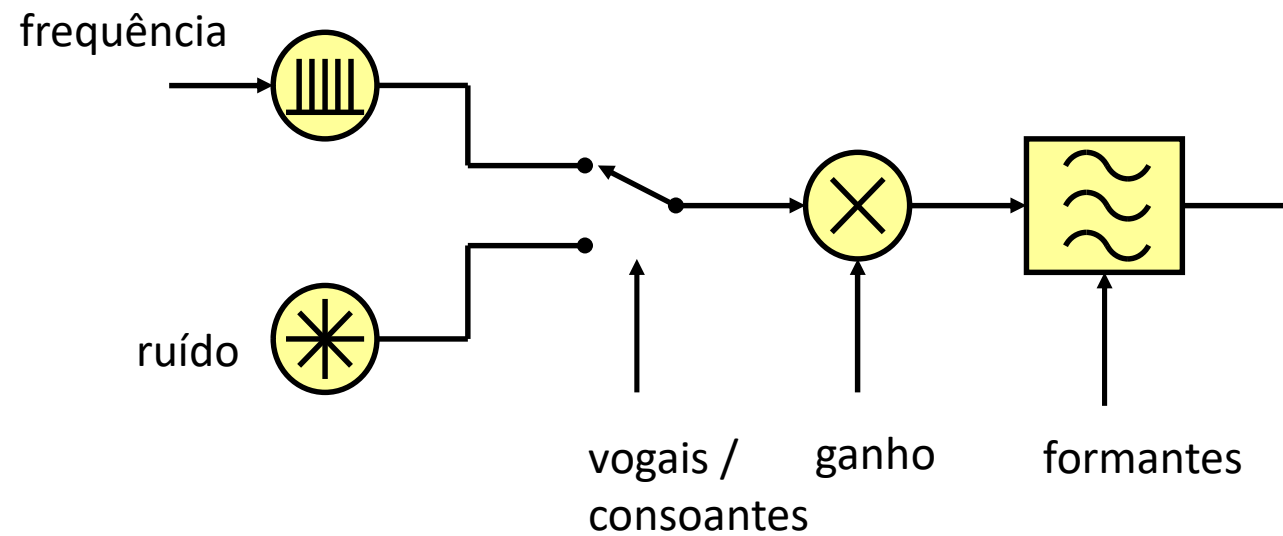
PARAMode = 0	HVXC
1	HILN
2	Comutado HVXC / HILN (fading 20 ms)
3	Combinado HVXC / HILN

- Sinais Vocálicos:
 - Quantização Vetorial da envoltória espectral dos resíduos da codificação LPC (VQ)
- Sinais Não Vocálicos:
 - Codificação por Excitação Vetorial (VXC)

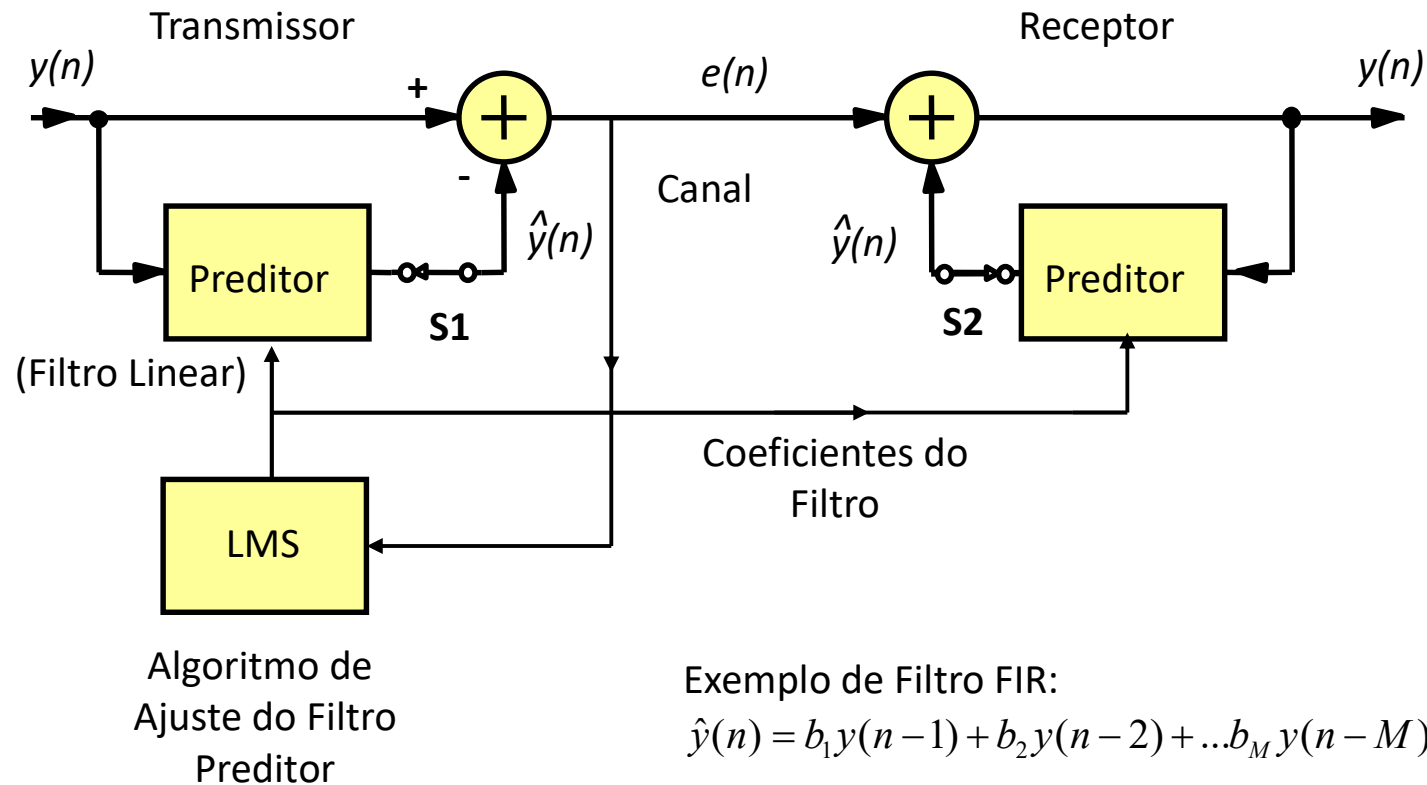
- Vocoder
- LPC (*Linear Predictive Coding*)
- Quantização Vetorial (VQ)
- Decodificação Excitada por Código
- Síntese harmônica
- PNS (*Perceptual Noise Substitution*)

Síntese Básica de Voz (Vocoder)

LCS
EPUSP

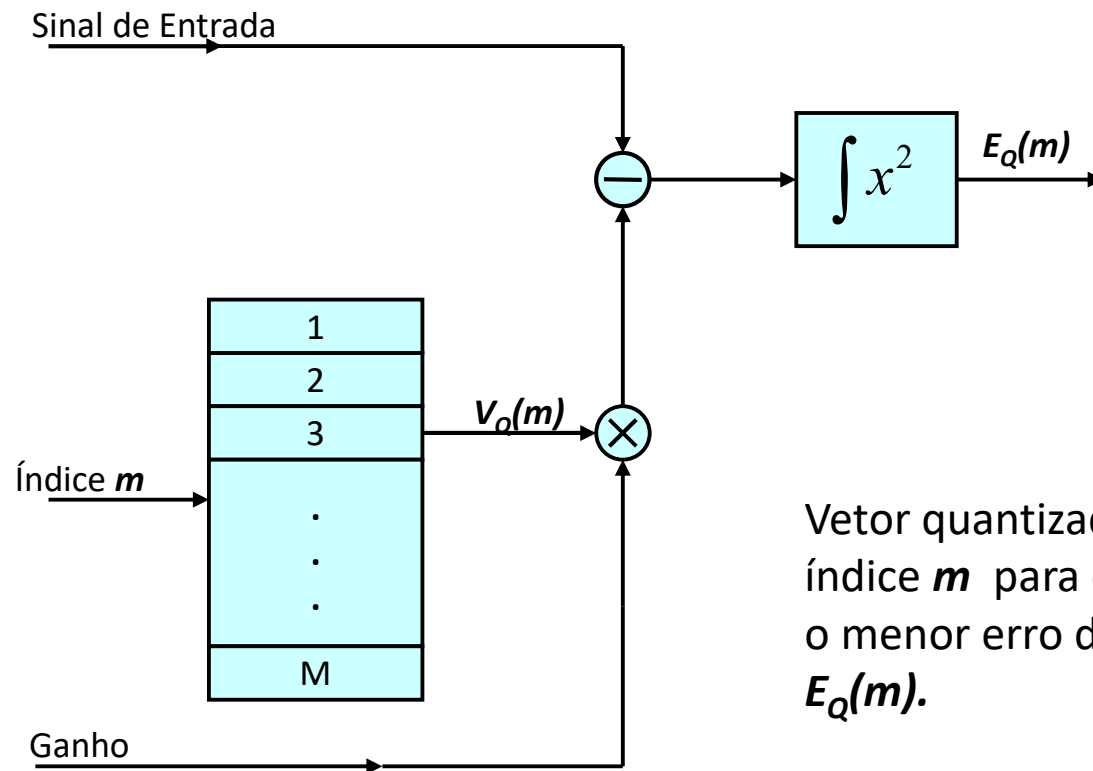


Síntese por LPC (Linear Predictive Coding)



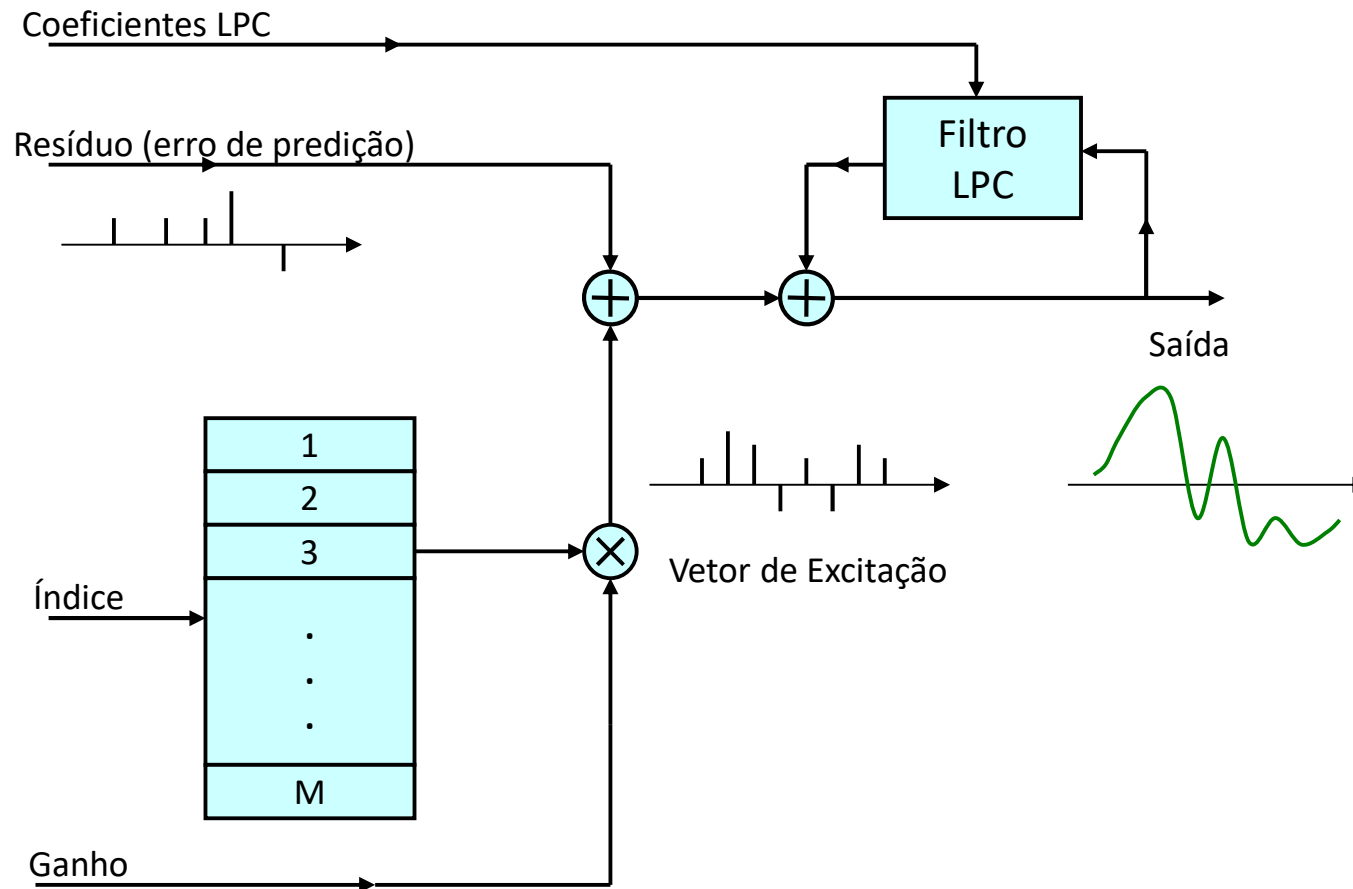
- Conjunto N-dimensional de valores de entrada é quantizado como um vetor (de N dimensões), escolhido entre um conjunto de M vetores possíveis, de acordo com um critério de “mínima distância”;
- A lista dos M vetores possíveis é o “dicionário de vetores” (*codebook*);
- O conjunto de valores de entrada é convertido em um número ***m*** (índice) que identifica o vetor quantizado dentro da lista.

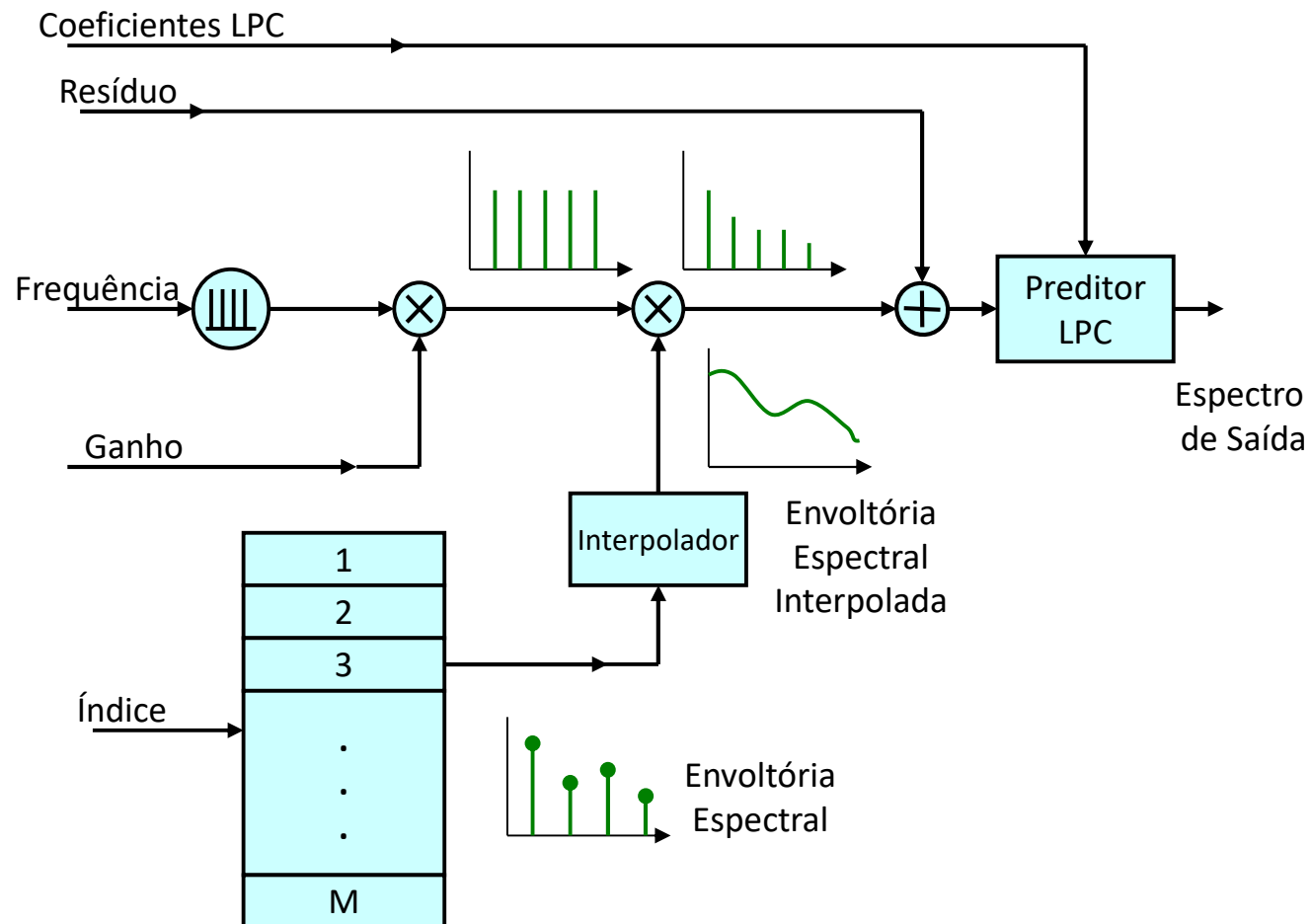
Quantização Vetorial



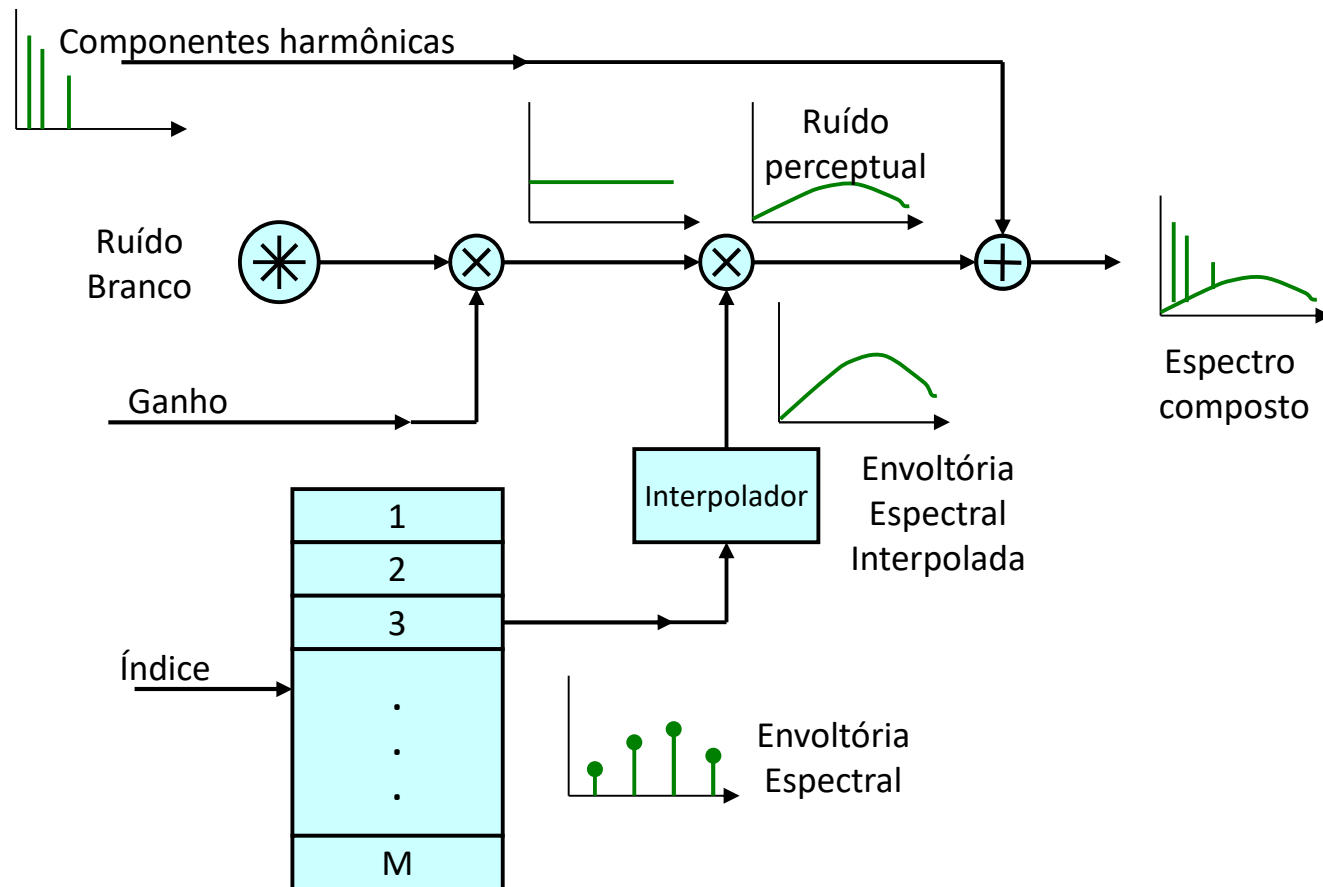
Vetor quantizado é dado pelo índice m para o qual foi obtido o menor erro de quantização $E_Q(m)$.

Decodificação Excitada por Código (CELP)



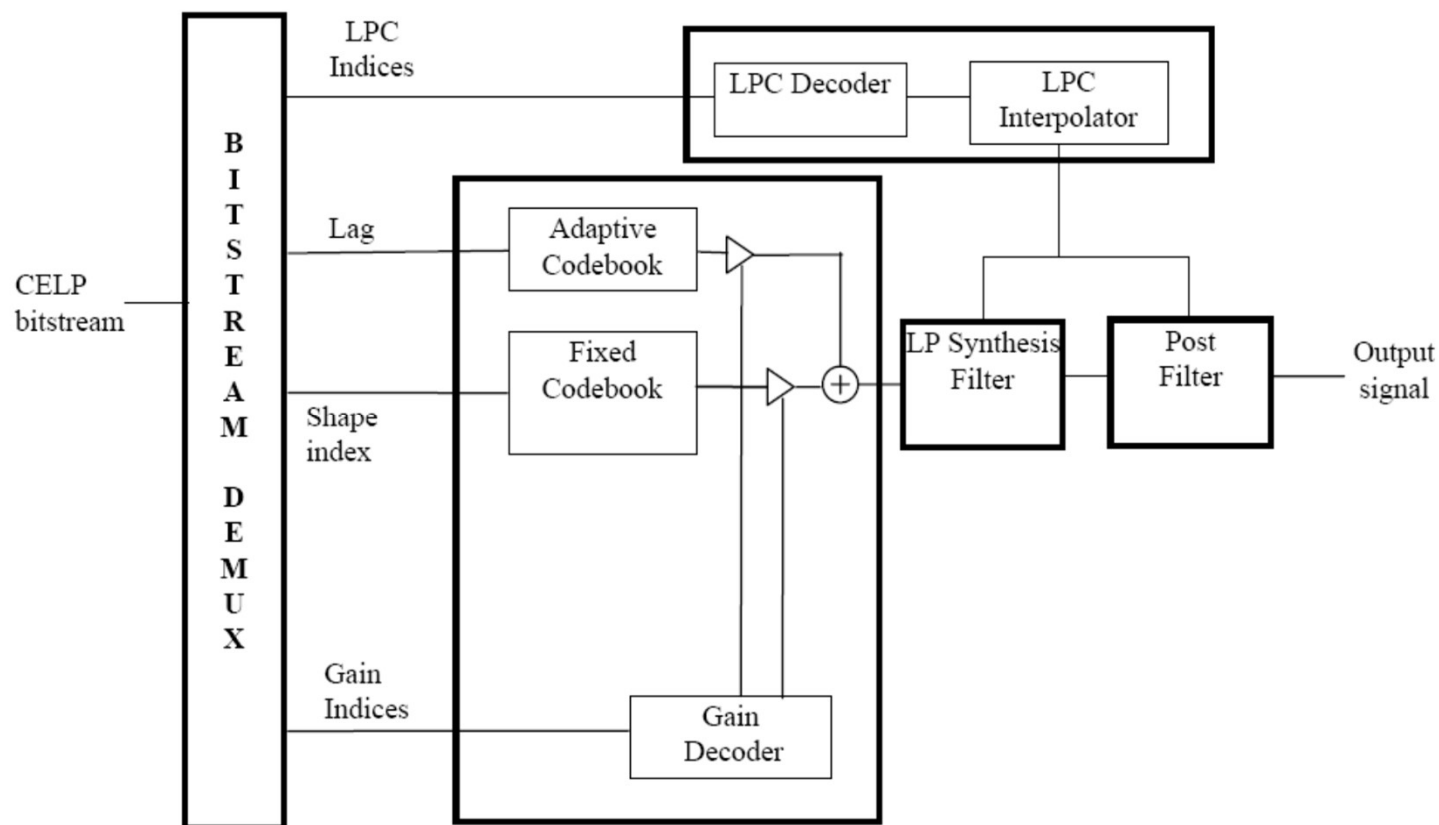


PNS (Perceptual Noise Substitution)



- Taxa de Amostragem de 8 kHz:
 - Quantizador Vetorial
 - Taxas de Bits de 3850 a 12200 bps
- Taxa de Amostragem de 16 kHz:
 - Quantizador Escalar
 - Taxas de Bits continuamente variáveis de 13667 a 20133 bps e de 21800 a 24000 bps

Ferramentas Específicas do Decodificador CELP MPEG-4

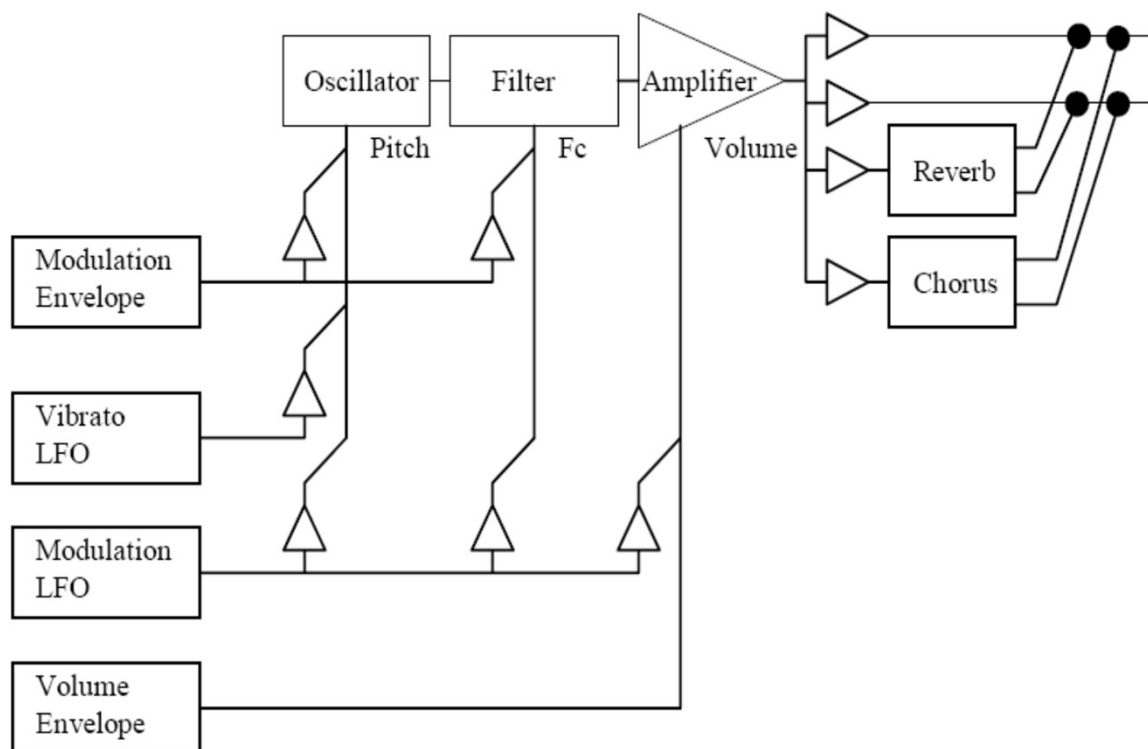


Requisitos Computacionais Típicos (MPEG-4)

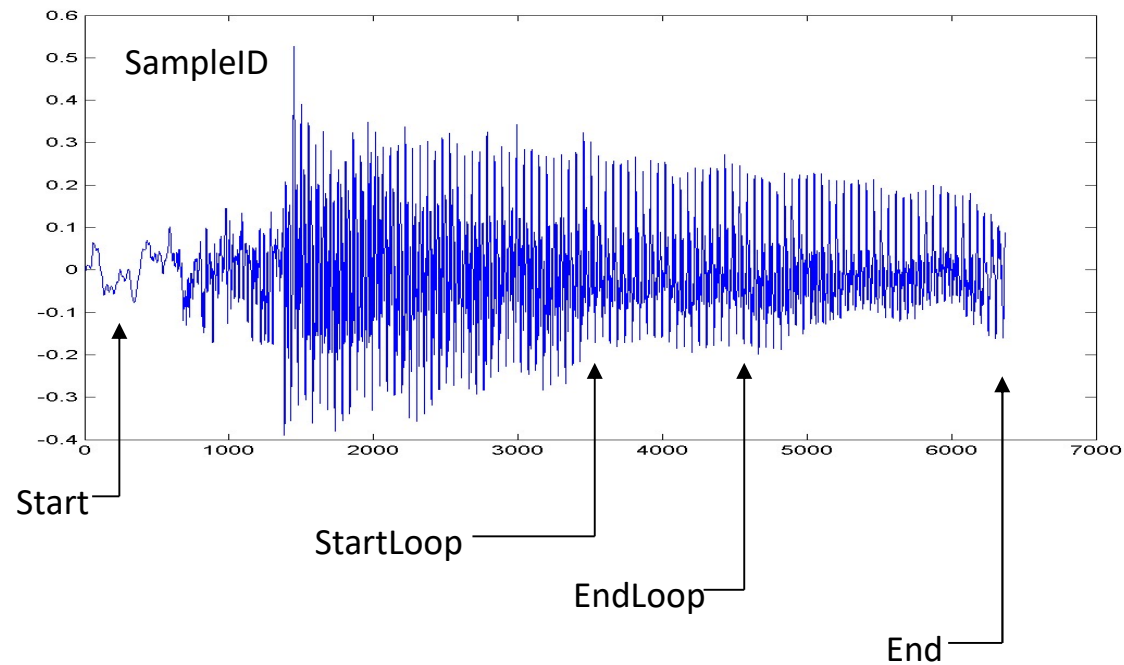
	f_A (kHz)	MOPS	kWords
AAC Principal	48	5	5
AAC-LC	48	3	3
AAC-SSR	48	4	3
LTP	48	4	4
AAC Escalável	48	5	4
VQ Duplo	24	2	3
CELP	8	1	1
CELP	16	2	1
CELP	8/16	3	1
HVXC	8	2	1

- Síntese Musical
 - SAOL: Structured Audio Orchestra Language
 - SASL: Structured Audio Score Language
 - SASBF: Structured Audio Sample Bank Format
 - MIDI: Musical Instruments Digital Interface
- Síntese de Fala
 - TTS (*Text To Speech*)

Modelo de Síntese Musical



- Geradores de:
 - **Índice**: apontam para outra estrutura de dados
 - **Range**: indicam uma faixa de valores onde um parâmetro (nota, intensidade) é definido
 - **Substituição**: atribuem um valor numérico a um evento
 - **Amostra**: afetam as propriedades de um som amostrado
 - **Valor**: afetam diretamente parâmetros do som sintetizado.

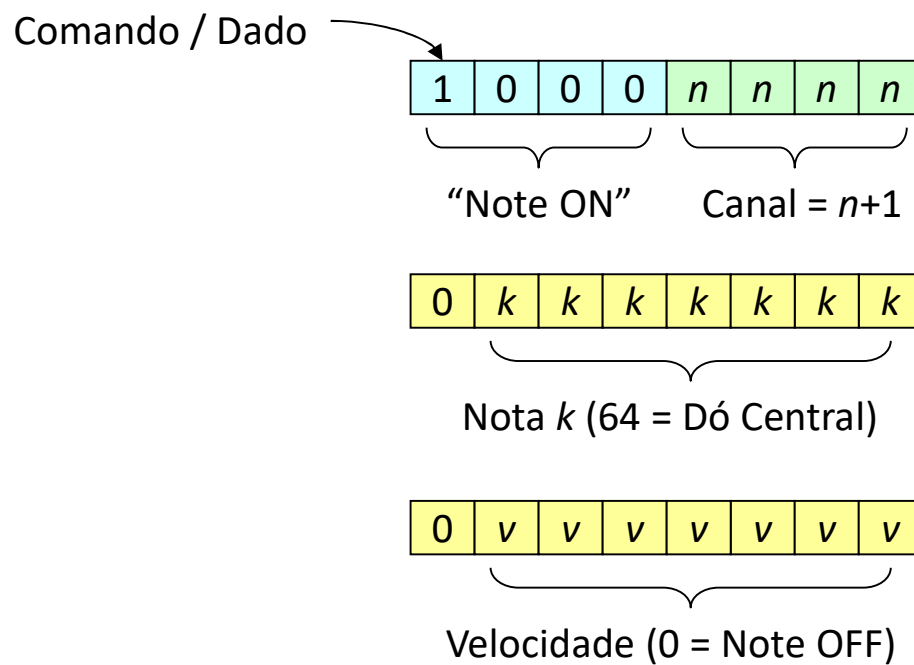


PITCH: Afeta frequência de amostragem,
após modificado por Vibrato, Modulação etc.

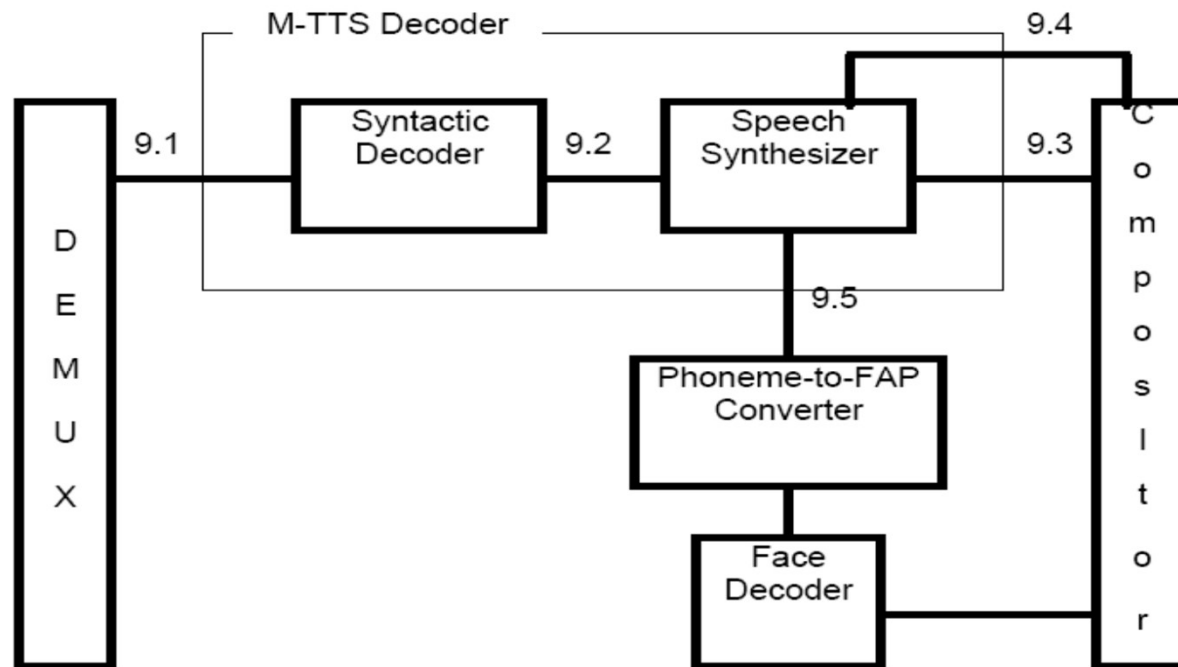
- **Canal MIDI:** 1 a 16 – identifica módulo que deve receber o comando.
- **Note ON:** identifica o Canal MIDI, a nota musical (0 a 127) e a velocidade de acionamento (1 a 127, associada geralmente à intensidade).
- **Note OFF:** idem, com velocidade de “desacionamento”.
- **Controlador:** CC 0 a 127: atribui valores contínuos (14 bits) a parâmetros pré-determinados (volume, *pitch bend*, filtro, etc.)
- **Program Change:** seleciona configuração pré-programada para conjunto de parâmetros de síntese (ex. *General Midi* estabelece lista de instrumentos padronizados).

- **Aftertouch:** parâmetro que afeta uma nota de um dado canal, após o evento **Note ON** (p. ex, alterando a modulação ou intensidade).
- **System Exclusive:** Informação específica para um determinado equipamento (p. ex., programação de parâmetros).
- **Standard Midi File (SMF):** sequência de comandos MIDI com informações de temporização relativa entre os comandos

Mensagem MIDI



Codificador MPEG-4 TTS (Text To Speech)



Aplicação do Codificador TTS

