

# Computação Musical - Áudio Digital

*slides do curso MAC0337/5900 – Computação Musical,  
baseados no capítulo 2 do livro de F. Richard Moore  
Elements of Computer Music*

Marcelo Queiroz

Departamento de Ciência da Computação  
Instituto de Matemática e Estatística  
Universidade de São Paulo

Versão de 21/08/2019

# Conteúdo

- Representações sonoras
- Digitalização de som
- Mensuração do espectro
- Filtros digitais

## Objetivos das representações sonoras

- Fidelidade da representação (não perder ou deturpar informação na codificação).
- Fornecer perspectivas diversas sobre a informação (que facilitem por exemplo a reprodução, manipulação e análise).
- Equivalência entre representações (exemplo mapa x instruções).
- Nosso escopo: sinais de áudio analógicos, sinais digitais de áudio e espectros sonoros.

## Transformações entre representações de áudio

- Registro sonoro: amostragem de um campo sonoro 3D em um ponto do captação.
- ADC: conversão analógico -> digital.
- DAC: conversão digital -> analógica.
- FT ou transformada de Fourier: conversão da representação entre os domínios temporal e espectral.
- Existem várias “transformadas de Fourier”: séries de Fourier, CFT, DFT, DTFT, STFT...

# Transdutores

- Propósito: converter energia ou informação de uma forma para outra.
- Ouvido: converte variações da pressão atmosférica em impulsos nervosos.
- Microfone: converte pressão sonora em sinais elétricos contínuos.
- Alto-falante: converte sinais elétricos em energia mecânica (movimento da membrana) e conseqüentemente em variações da pressão do ar.

## Acurácia da transdução

- todo componente eletrônico (e mesmo o ouvido humano) possui respostas diferenciadas nas diversas faixas de frequência (a chamada resposta em frequência)
- No caso do ouvido humano já observamos as distorções de natureza psico-acústica (curvas de phons, sones, mels, etc.).
- Microfones possuem um pequeno diafragma que responde a uma faixa de frequências específica, e mesmo dentro dessa faixa podem introduzir distorções de amplitude e fase.
- Alto-falantes possuem membranas muito maiores do que microfones, e essas membranas não têm flexibilidade para reproduzir grandes faixas de frequência.
- *Alto-falantes* específicos: *subwoofer, woofer, mid-range speaker, tweeter, supertweeter, ...* (todos com distorções em suas respostas em frequência).

## Sinais analógicos

- Seja  $p(t)$  a variação de pressão no ar e considere um transdutor pressão  $\rightarrow$  voltagem. Idealmente:

$$v(t) \propto p(t)$$

Na prática  $v(t) \propto D[p(t)] + r(t)$  (D=distorção, r=ruído).

- Distorção+ruído lineares são fáceis de eliminar. Tipicamente estes efeitos são não-lineares.
- Um dos efeitos do ruído é limitar a faixa dinâmica de operação dos transdutores.

## Formas de onda analógicas

- Funções periódicas:

$$f(t + \tau) = f(t), \quad \forall t.$$

o menor valor de  $\tau$  que verifica esta equação é chamado de *período* da função.

- Não existem sons musicais periódicos (assim como não existe música que dure toda a eternidade).
- Ainda assim, o conceito é muito útil, pois o período de um recorte sonoro (pensando nele repetido ciclicamente) está relacionado à *frequência* e ao conceito de *altura musical*.

## Representação de sinais periódicos

- Sinais senoidais ( $\omega$ =frequência angular,  $f_0$ =frequência em Hz):

$$x(t) = \alpha \cos(\omega t + \varphi_0)$$

$$x(t) = \alpha \cos(2\pi f_0 t + \varphi_0)$$

- Somas de senoides:  $\omega = 2\pi f_0 \Leftrightarrow f_0 = \frac{\omega}{2\pi}$

$$y(t) = \sum \alpha_k \cos(2\pi f_k t + \varphi_k)$$

- Outras formas de onda simples:

quadrada:  $q(t) = \begin{cases} 1 & t \in [k, k+1), k \text{ par} \\ -1 & t \in [k, k+1), k \text{ ímpar} \end{cases}$

dente-de-serra:  $s(2k + t) = t, \forall k \in \mathbb{Z}, \forall t \in [-1, 1]$ .

## Ruído: Definições e dificuldades

- Em teoria da informação (também em DSP) ruído é qualquer sinal indesejado adicionado à informação desejada.
- Em música, ruído é uma categoria sem fronteiras claramente definidas, que abrange uma infinidade de sons.
- Geralmente, ruídos não possuem altura musical definida; porém, alguns sons sem altura musical definida (gongos, sinos, etc.) não são necessariamente denominados ruídos.
- Ruídos musicais costumam não ser periódicos ou quase-periódicos (mas existem exceções: sons de máquinas, veículos, apitos, etc.).

## Caracterização de ruídos

- Caracterizações espectrais: ruídos de banda-larga, banda-estreita, branco, etc..
- Ruído branco: possui energia constante em todas as faixas de frequência.
- Outras cores de ruído:
  - rosa (energia constante por oitava, espectro  $\frac{1}{f}$ ),
  - vermelho (Browniano ou  $\frac{1}{f^2}$ ),
  - cinza (nível perceptual constante de acordo com uma curva como dBA ou ISO 226)

## Relação sinal ruído (SNR)

- A relação sinal-ruído de um sistema é definida como

$$SNR = \frac{\text{máxima amplitude representável}}{\text{amplitude do ruído}}$$

- A mesma medida em dB:

$$SNR = 20 \log \frac{\text{máxima amplitude representável}}{\text{amplitude do ruído}}$$

- Observação: dB é normalmente medido em intensidade/energia; como energia  $\propto$  amplitude<sup>2</sup>,

$$dB_E = 10 \log \frac{E}{E_0} = 10 \log \frac{A^2}{A_0^2} = 10 \log \left( \frac{A}{A_0} \right)^2 = 20 \log \frac{A}{A_0} = dB_A.$$

## Medidas de amplitude

- Amplitude de pico:

$$A_p(x, I) = \max\{|x(t)| \mid t \in I\}$$

- Amplitude média:

$$A_m(x, I) = \frac{1}{|I|} \sum_I |x(t)|$$

- Amplitude RMS (Root-Mean-Square):

$$A_{rms}(x, I) = \sqrt{\frac{1}{|I|} \sum_I x(t)^2}$$

## Medidas de amplitude

Exemplos:

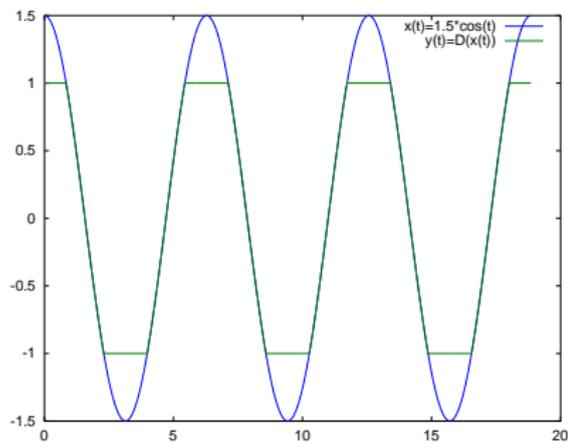
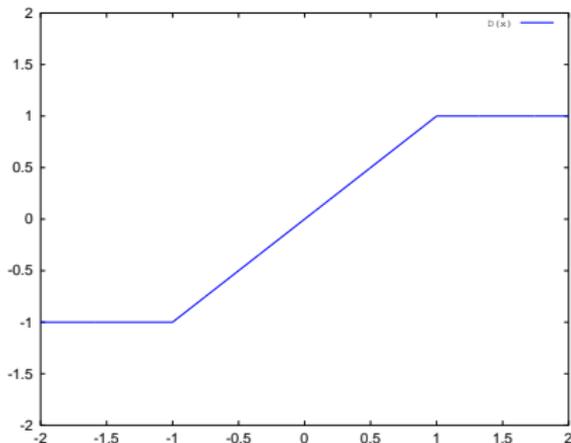
- senoide ( $A_p = 1, A_m = \frac{2}{\pi} \approx 0.637, A_{rms} = \frac{\sqrt{2}}{2} \approx 0.707$ )
- quadrada ( $A_p = A_m = A_{rms} = 1$ )
- dente-de-serra ( $A_p = 1, A_m = \frac{1}{2}, A_{rms} = \frac{1}{\sqrt{3}}$ )
- ruído branco uniforme ( $A_p = 1, A_m = \frac{1}{2}, A_{rms} = \frac{1}{\sqrt{3}}$ )
- 110 V AC possui pico de  $\frac{110}{0.707} \approx 155.56$  V (equivalente DC).

## Tipos de distorção

- Distorção em frequência: dispositivos reais são caracterizados por curvas de resposta que não são perfeitamente planas, e que apresentam quedas abruptas de resposta nas extremidades da faixa de frequência de operação.
- Distorção de amplitude: é caracterizada por não-linearidades na relação entre entrada e saída. Por exemplo, se a entrada varia numa faixa de valores maior do que o dispositivo consegue mapear, o sinal fica “clipado”.

## Exemplo de distorção de amplitude

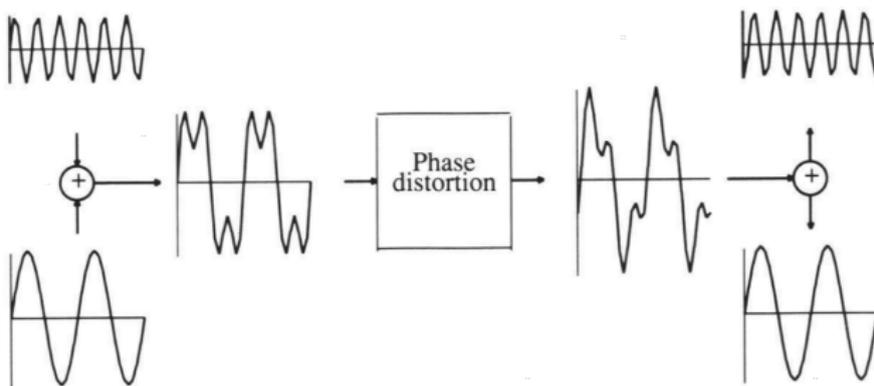
$$\text{Exemplo: } D(x) = \begin{cases} +1 & \text{se } x > 1 \\ x & \text{se } x \in [-1, 1] \\ -1 & \text{se } x < -1 \end{cases} \text{ e } x(t) = 1.5 \cos(t).$$



## Tipos de distorção

- Distorção de fase: é caracterizado por diferenças de atraso entre frequências diferentes.
- Exemplo: num alto-falante com 3 *drivers* (graves, médios e agudos), as pequenas diferenças de posicionamento dos falantes já introduzirá diferenças de atraso na chegada do som ao ouvinte.
- Em 10000 Hz, o comprimento de onda é de 3.44 cm (8.6 mm de diferença dos caminhos já resulta em  $90^\circ$  de desvio de fase)  
Em 1000 Hz, o comprimento de onda é de 34.4 cm (8.6 cm de diferença dos caminhos já resulta em  $90^\circ$  de desvio de fase)

## Exemplo de distorção de fase



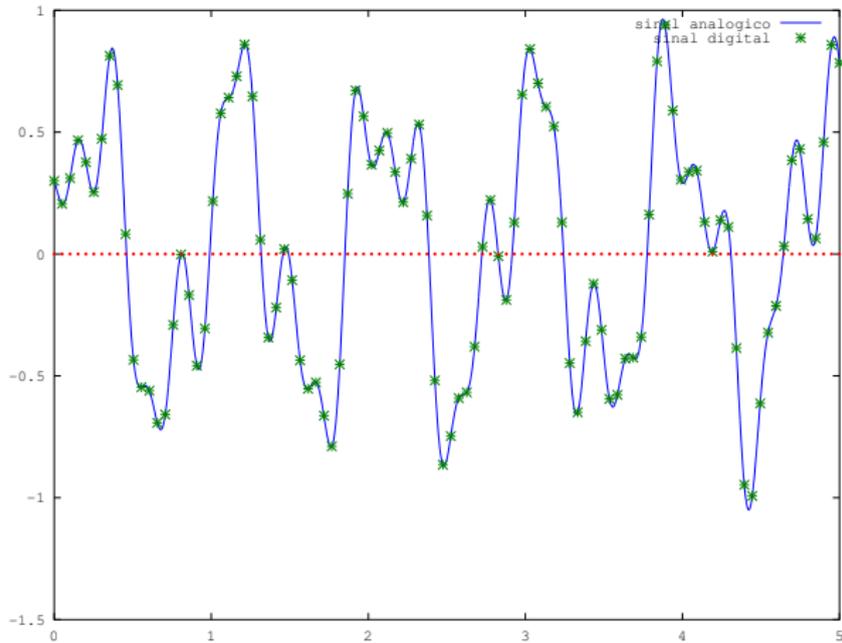
- Exemplo em Pd

## Percepção da distorção

- Em termos perceptuais, considera-se que em tons quasi-periódicos  
distorção de fase  $\ll$  distorção em frequência  $\ll$  distorção de amplitude  
onde  $\ll$  deve ser lido como “é menos pior que”.
- A forma de onda não ajuda muito a entender estas diferenças:  
formas muito diferentes podem soar quase iguais, e formas muito parecidas podem soar bastante diferentes.
- A distorção de fase pode ser muito perceptível em sons transientes (ataques ruidosos, consoantes).
- Ruídos de banda-larga (mesmo com intensidade razoável) podem ser menos perceptíveis que ruídos de banda-estreita (mesmo com intensidade bem mais baixa).

# Digitalização: Sistemas ADC e DAC

- Representação digital: sequências de amostras  $x_1, x_2, \dots, x_n, \dots$  associadas a instantes equi-espaçados.
- Formato mais comum: PCM = Pulse Code Modulation, onde cada  $x_n$  consiste em um vetor de bits representando o valor instantâneo de uma amostra.
- PCM é caracterizado principalmente pela *taxa de amostragem* (número de amostras por segundo, em Hz) e pelo *formato da amostra* (número de bits e tipo de codificação).
- Alternativas: DPCM, PAM, PWM, DM, ADPCM, etc.



## Limitações da Representação Digital

- **Teorema da amostragem (Shannon-Nyquist):** para representar digitalmente sem perdas um sinal analógico contendo frequências estritamente menores do que  $X$  Hz é suficiente tomar pelo menos  $2X$  amostras por segundo. Mais ainda, para se representar corretamente todas as senóides no intervalo  $[0, X]$  Hz é necessário tomar pelo menos  $2X$  amostras por segundo.
- **Intuição:** uma senoide com frequência  $X$  HZ precisa de pelo menos 2 amostras (+1 e -1) para representar cada ciclo.

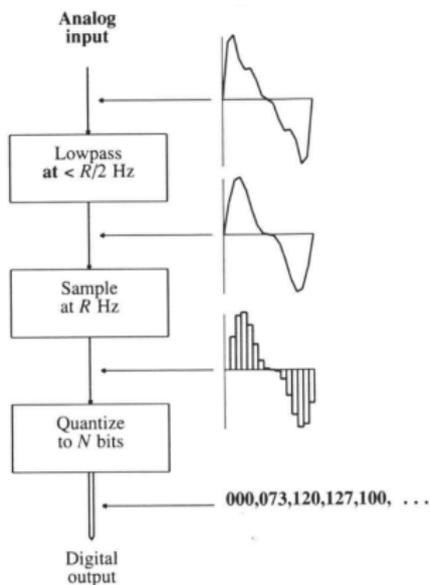
## Limitações da Representação Digital

- Necessidade do teorema: fácil de ver através do fenômeno de rebatimento (*aliasing*, seções 2.2.3 e 2.2.4).
- Suficiência do teorema: não é trivial (inversibilidade da DTFT).
- Se  $R$  Hz é a taxa de amostragem, então a frequência limite de representação  $R/2$  Hz é chamada de *frequência de Nyquist*.

## Conversão Analógico-Digital ( $R$ Hz, $N$ bits)

- Primeiro passo: remover frequências acima de Nyquist ( $R/2$  Hz).
- Segundo passo: colher amostras a cada  $1/R$  segundos.
- Terceiro passo: Quantizar cada amostra usando  $N$  bits (seção 2.2.5).

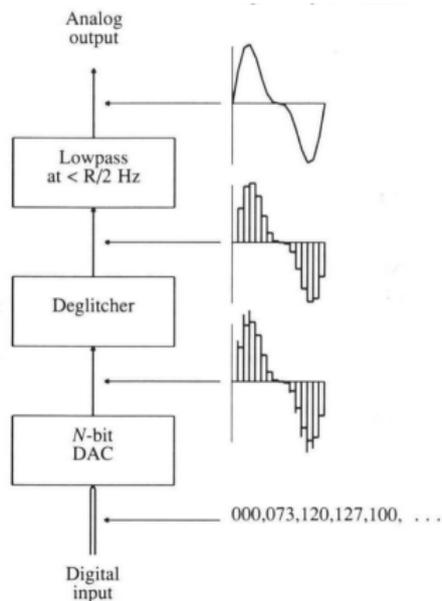
# Conversão Analógico-Digital (R Hz, N bits)



## Conversão Digital-Analógica

- Primeiro passo: construir um sinal analógico “escalonado” (módulo DAC: chave  $k \propto 2^k$ ).
- Segundo passo: eliminar transientes (*deglitcher*: *sample-and-hold*).
- Terceiro passo: remover frequências acima de Nyquist (filtro passa-baixa em  $R/2$  Hz).

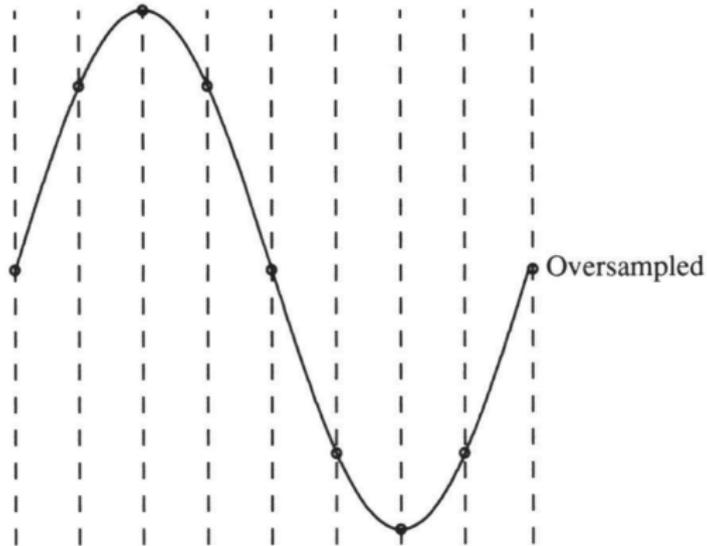
# Conversão Digital-Analógica



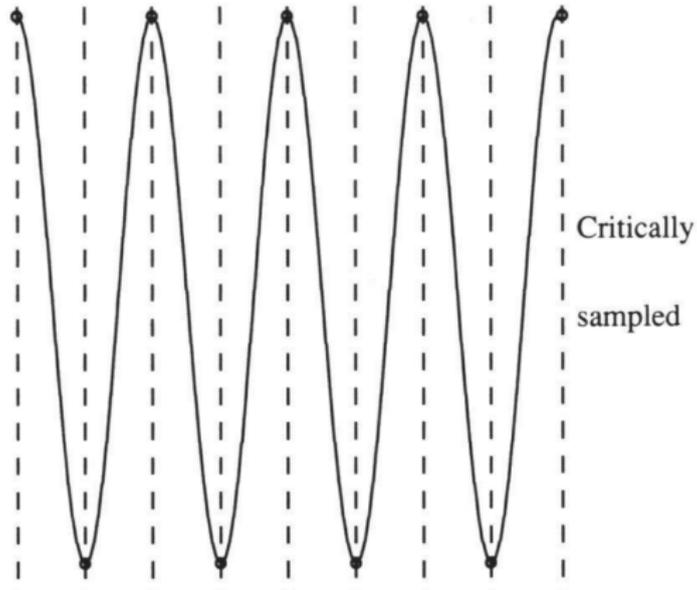
## Aliasing ou Rebatimento

- *Aliasing* é um fenômeno que ocorre quando tentamos representar um sinal que não é representável com a taxa de amostragem usada.
- Uma componente senoidal é considerada *super-amostrada* se é representada por mais de 2 amostras por ciclo, *criticamente amostrada* se é representada por exatamente 2 amostras por ciclo (frequência de Nyquist) e *sub-amostrada* se é representada por menos de 2 amostras por ciclo.
- Senoides sub-amostradas serão erroneamente interpretadas como possuindo frequência mais baixa.

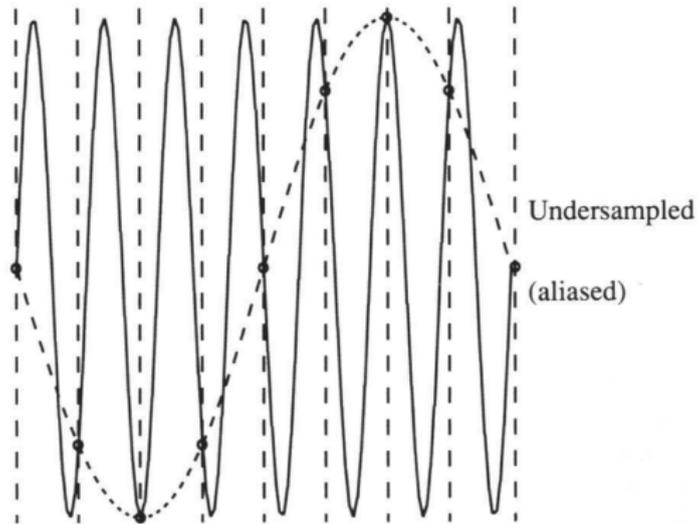
## Senoide super-amostrada



## Senoide criticamente amostrada



## Senoide sub-amostrada (rebatida)



## Rebatimento (ponto de vista algébrico)

- Uma senoide digitalizada é uma função de tempo discreto

$$x[n] = \alpha \cos \left( 2\pi f \frac{n}{R} + \varphi_0 \right).$$

- Se  $f = f_0 + kR$  ( $k \in \mathbb{Z}$ ), então  $\forall n \in \mathbb{Z}$  temos

$$\begin{aligned} x[n] &= \alpha \cos \left( 2\pi f \frac{n}{R} + \varphi_0 \right) \\ &= \alpha \cos \left( 2\pi (f_0 + kR) \frac{n}{R} + \varphi_0 \right) \\ &= \alpha \cos \left( 2\pi f_0 \frac{n}{R} + 2\pi kn + \varphi_0 \right) \\ &= \alpha \cos \left( 2\pi f_0 \frac{n}{R} + \varphi_0 \right) \end{aligned}$$

## Rebatimento (ponto de vista perceptual)

- Todas as frequências  $\{f = f_0 + kR \mid k \in \mathbb{Z}\}$  pertencem a uma mesma *classe de equivalência*, cujo representante natural está no intervalo  $[-\frac{R}{2}, +\frac{R}{2}]$ .
- Senoides com frequência acima de Nyquist simplesmente soam com a frequência rebatida de acordo com a expressão  $f_r = f - \text{round}(\frac{f}{R}) R$ .
- Exemplos de rebatimento em Pd (senoides).
- Algoritmos de síntese e processamento podem produzir componentes acima de Nyquist, com distorções especificamente associadas ao rebatimento.

## Exemplo de rebatimento: onda quadrada

- Podemos sintetizar uma onda quadrada perfeita pela expressão

$$q(t) = \begin{cases} 1 & t \in [k, k+1), k \text{ par} \\ -1 & t \in [k, k+1), k \text{ ímpar} \end{cases}$$

- Porém esta expressão viola a condição de limitação de banda:

$$q(t) = \left. \begin{aligned} &\text{sen}(2\pi \frac{1}{2} t) + \frac{1}{3} \text{sen}(2\pi \frac{3}{2} t) + \frac{1}{5} \text{sen}(2\pi \frac{5}{2} t) \\ &+ \frac{1}{7} \text{sen}(2\pi \frac{7}{2} t) + \frac{1}{9} \text{sen}(2\pi \frac{9}{2} t) + \frac{1}{11} \text{sen}(2\pi \frac{11}{2} t) \\ &+ \frac{1}{13} \text{sen}(2\pi \frac{13}{2} t) + \frac{1}{15} \text{sen}(2\pi \frac{15}{2} t) + \dots \end{aligned} \right\} \begin{array}{l} \text{Decomposição} \\ \text{por} \\ \text{Fourier} \end{array}$$

- Exemplos de rebatimento em Pd (ondas quadradas).

## Exemplo de rebatimento: onda quadrada

- Por causa do rebatimento, a onda gerada é bem diferente!
- A solução é usar uma aproximação da onda quadrada com limitação na faixa de frequências:

$$\tilde{q}(t) = \sum_{k=1,3,5,\dots,K} \frac{1}{k} \text{sen}(2\pi k f_0 t)$$

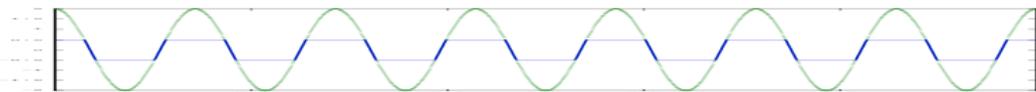
de tal modo que  $K f_0 < \frac{R}{2} \implies K < \frac{R}{2f_0}$ .

## Exemplo de rebatimento: *overdrive*

- Sinais com infinitos harmônicos podem ser produzidos, inadvertidamente, tanto na síntese quanto no processamento sonoro.
- Um exemplo muito próximo do anterior é o da distorção do tipo *overdrive*, onde a amplitude é distorcida por uma função do tipo

$$D(x) = \text{clip}(x, \alpha) = \min\{\alpha, \max\{-\alpha, x\}\}$$

- Quanto maior a diferença entre a amplitude de uma senoide e o valor de *clipping*  $\alpha$ , mais parecida será a saída desse processo com uma onda quadrada (infinitos harmônicos  $\implies$  *aliasing*):



## Quantização Linear (LQ)

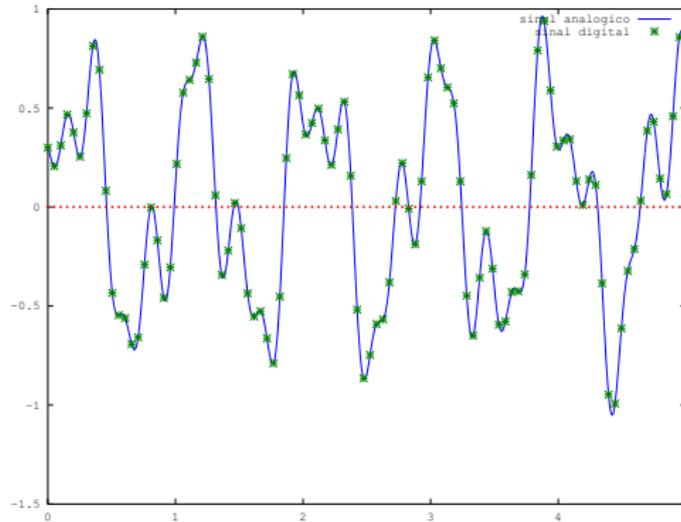
- Objetivo: Mapear uma faixa contínua de valores de entrada  $[-E, +E]$  a um conjunto finito de  $2^N$  valores de amplitude discreta.
- Divisão linear: o intervalo  $[-E, +E]$  é dividido em pedaços de tamanho  $\frac{2E}{2^N} = \frac{E}{2^{N-1}}$  com fronteiras

$$\left[ k \frac{E}{2^{N-1}}, (k+1) \frac{E}{2^{N-1}} \right], \quad k = -2^{N-1}, \dots, -1, 0, 1, \dots, 2^{N-1}-1.$$

- Cada intervalo destes será mapeado em um único código de  $N$  bits que normalmente representa o valor médio do intervalo.
- Assimetria na atribuição: pontos na fronteira de dois intervalos devem ser “arredondados” para cima ou para baixo.

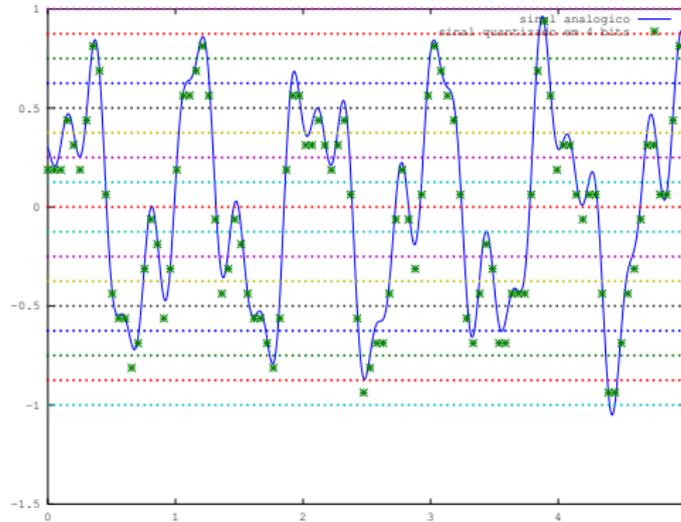
## Exemplo: Quantização Linear

Sinal amostrado sem quantização:



## Exemplo: Quantização Linear

Sinal amostrado e quantizado:



## Erro de quantização

- Seja  $q(x)$  o mapa (não-linear) que transforma uma amplitude  $x \in [-E, +E]$  em um dos códigos  $q \in \{0, 1, \dots, 2^N - 1\}$  e  $d(q)$  o mapa do sentido contrário, que atribui a cada código  $q \in \{0, 1, \dots, 2^N - 1\}$  um valor  $x \in [-E, +E]$ .
- A composição quantização+dequantização  $d(q(x))$  representa um processo com perda de informação, já que  $q(x)$  não é inversível. Em particular,  $d(q(x)) \neq x$  para quase todo  $x \in [-E, +E]$ .
- O sinal  $x[n] - d(q(x[n]))$  é chamado de ruído de quantização, sendo sua amplitude máxima  $\frac{E}{2^N}$  no caso da LQ.

## Relação Sinal/Ruído da LQ

- Podemos medir a relação sinal/ruído no caso específico do ruído de LQ pela expressão

$$SQNR = \frac{\text{ampl. máx. representável}}{\text{ampl. máx. ruído quant.}} = \frac{E}{E/(2^N)} = 2^N,$$

ou em decibéis:

$$SQNR(dB) = 20 \log_{10} 2^N = 20N \log_{10} 2 \approx 6N.$$

(cada bit a mais acrescenta 6 dB na relação sinal/ruído).

## Relação Sinal/Ruído da LQ

- Exemplos: CD usa 16 bits, com  $SQNR \approx 96 \text{ dB}$ ; áudio com 24 bits possui  $SQNR \approx 145.5 \text{ dB}$ , faixa dinâmica maior do que a dos limites da percepção humana ( $\approx 120 \text{ dB}$ ).
- A medida de SQNR linear se refere à amplitude máxima representável: um sinal pianíssimo em um CD poderia facilmente estar  $60 \text{ dB}$  abaixo do nível máximo, e nesse caso particular a relação sinal/ruído seria de  $36 \text{ dB}$  (perceptível).

## Características do ruído de quantização

- O ruído de quantização **NÃO É** independente da entrada, e por isso pode ter características espectrais indesejáveis, que o tornem mais ou menos perceptíveis de acordo com contexto: como exemplo, um sinal periódico (com período inteiro em amostras) terá ruído de quantização periódico. (uma solução para isso é a técnica de *dithering*)
- Em um exemplo limite de distorção, um sinal senoidal com amplitude  $\frac{E}{2^N}$  seria mapeado em uma onda quadrada.
- Uma solução para melhorar a relação sinal/ruído para sons de menor amplitude é usar técnicas de quantização *não-linear*.

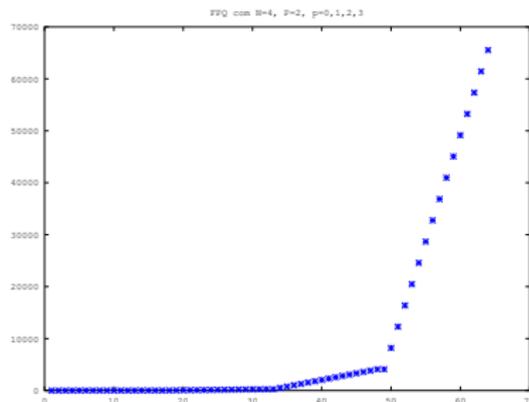
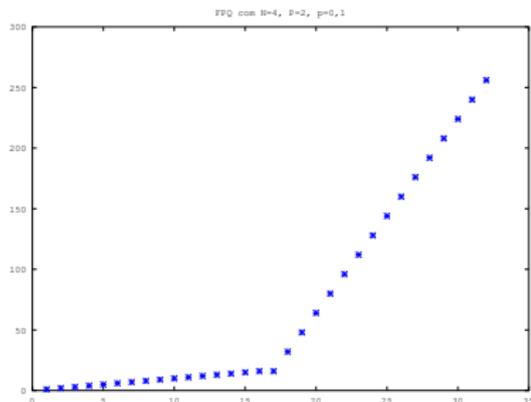
## Quantização Não-Linear

- Ideia: usar regiões mais finas para amplitudes mais baixas, em detrimento de regiões mais grosseiras nas amplitudes maiores
- Justificativa: um erro de quantização com amplitude maior tem mais chance de passar despercebido (mascaramento) se o sinal é mais forte.
- Resultado: melhor relação sinal/ruído para vários níveis de amplitude simultaneamente.
- Abordagens: quantização em ponto flutuante e quantização logarítmica.

## Quantização em Ponto Flutuante (FPQ)

- Considere que usaremos  $N$  bits para representar a *mantissa* (dígitos significativos, incluindo o sinal) e  $P$  bits para representar o *expoente*.
- O maior valor representável pela mantissa será então  $2^{N-1}$ . Consideraremos que essa faixa está associada a  $P = 0$ .
- Para  $P = 1$  os valores de mantissa serão multiplicados por  $2^N$ ; para  $P = 2$ , os mesmos valores serão multiplicados por  $2^N 2^N = 2^{2N}$ .
- Em geral, se  $m$  é a mantissa e  $p$  é o expoente, então o inteiro representado será  $m \cdot 2^{pN}$ .

## Exemplo de FPQ $c/N=4$ , $P=2$



## Relação Sinal/Ruído para FPQ

- O maior sinal representável em FPQ tem amplitude  $2^{N-1}2^{(2^P-1)N}$ . Mas o ruído de quantização nesta faixa também será grande:  $\frac{1}{2}2^{(2^P-1)N} = 2^{-1}2^{(2^P-1)N}$ .
- A relação sinal/ruído será então:

$$SQNR_{dB} = 20 \log_{10} \frac{2^{N-1}2^{(2^P-1)N}}{2^{-1}2^{(2^P-1)N}} = 20 \log_{10} 2^N \approx 6N.$$

## Relação Sinal/Ruído para FPQ

- Na faixa mais fina de representação ( $P = 0$ ), a amplitude máxima representável é  $2^{N-1}$  e o ruído de quantização tem amplitude máxima  $\frac{1}{2}$ , com uma relação sinal/ruído de:

$$SQNR_{dB} = 20 \log_{10} \frac{2^{N-1}}{1/2} = 20 \log_{10} 2^N \approx 6N.$$

- Em todas as faixas intermediárias (expoentes  $2^{pN}$ ) teremos sempre

$$SQNR_{dB} = 20 \log_{10} \frac{2^{pN+N-1}}{2^{pN-1}} = 20 \log_{10} 2^N \approx 6N.$$

## Vantagens em relação à LQ

- Manter uma mesma relação sinal-ruído em todos os níveis de amplitude (correspondentes a valores diferentes de  $p$ ) já é por si só uma vantagem em relação à LQ.
- Além desta vantagem, a faixa dinâmica de representação, definida como a relação entre a máxima e a mínima amplitudes representáveis, será neste caso:

$$\begin{aligned} \text{Faixa Dinâmica}(dB) &= 20 \log_{10} \frac{2^{N-1}2^{(2^P-1)N}}{1/2} \\ &= 20 \log_{10} 2^{N+2^P N-N} \approx 6N \cdot 2^P. \end{aligned}$$

- Exemplo: tomando  $N = 14$  e  $P = 2$  ( $N + P = 16$ ) teríamos  $SQNR = 84 \text{ dB}$  em todas os níveis de amplitude, e uma faixa dinâmica de  $336 \text{ dB}$ .

## Quadro comparativo entre LQ e FPQ

- Podemos comparar FPQ e LQ com o mesmo número de bits se pensarmos em  $N$  bits particionados em  $P$  bits de expoente e  $N - P$  bits de matissa:

	Linear	Ponto Flutuante
$SQNR_{dB}$	$\approx 6N$	$\approx 6(N - P)$
Faixa Dinâmica( $dB$ )	$\approx 6N$	$\approx 6(N - P) \cdot 2^P$

## Codificação

- Após a quantização, cada amostra digital precisa ser colocada em um formato específico de representação:
  - inteiros sem sinal,
  - inteiros com sinal,
  - inteiros com sinal em complemento de 2,
  - bytes ordenados na sequência natural (*little-endian*) ou inversa (*big-endian*),
  - ponto flutuante
- Além disso, os códigos podem receber acréscimos para detectar ou corrigir erros no armazenamento ou transmissão (*error-correction codes*).