

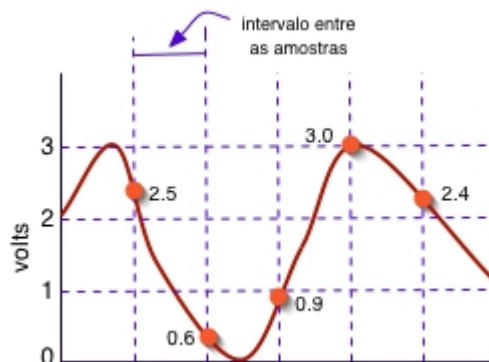
Áudio Digital

O áudio digital é, essencialmente, uma representação numérica do som.

As ondas sonoras se propagam de modo contínuo no tempo e no espaço. Para que sejam representadas no meio digital, seu comportamento analógico (contínuo) tem que ser convertido numa série de valores discretos (descontínuos). Esses valores são números (dígitos) que representam amostras (*samples* em inglês) instantâneas do som. Isso é realizado por meio de um conversor analógico/digital (CAD). Se quisermos ouvir novamente o som, torna-se necessário que os sinais digitais representados por números binários sejam reconvertidos em sinais analógicos por meio de um conversor digital/analógico (CDA).

Amostragem

A conversão do sinal analógico para o digital é realizada por uma sequência de amostras da variação de voltagem do sinal original. Cada amostra é arredondada para o número mais próximo da escala usada e depois convertida em um número digital binário (formado por "uns" e "zeros") para ser armazenado.



valores das amostras				
2.5	0.6	0.9	3.0	2.4
valores quantizados				
2	0	1	3	2
valores convertidos em dígitos binários				
10	00	01	11	10

Taxa de Amostragem

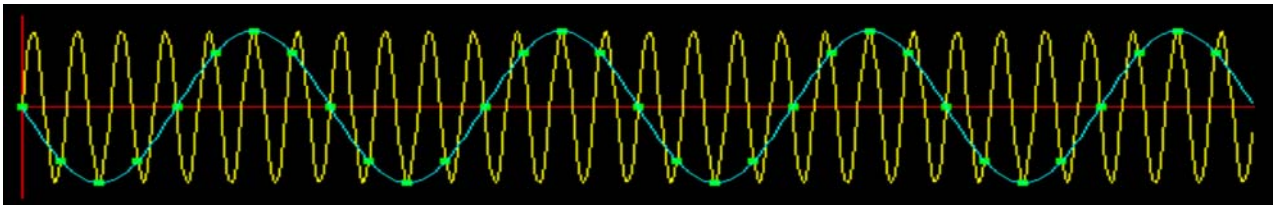
As amostras são medidas em intervalos fixos. O número de vezes em que se realiza a amostragem em uma unidade de tempo é a **taxa de amostragem**, geralmente medida em Hertz. Assim, dizer que a taxa de amostragem de áudio em um CD é de 44.100 Hz, significa que a cada segundo de som são tomadas 44.100 medidas da variação de voltagem do sinal. Dessa maneira, quanto maior for a taxa de amostragem, mais precisa é a representação do sinal, porém é necessário que se realize mais medições e que se utilize mais espaço para armazenar esses valores.

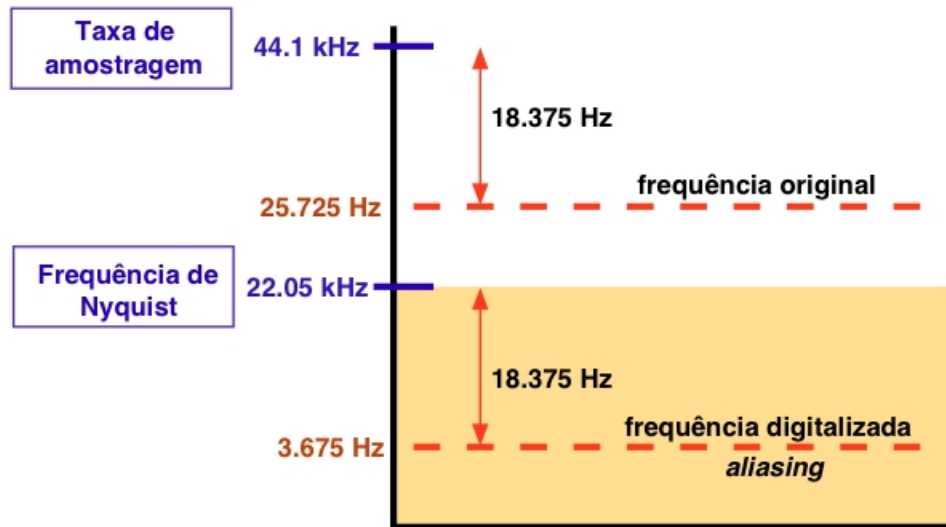
Teorema de Nyquist

A taxa de amostragem deve ser pelo menos duas vezes a maior frequência que se deseja registrar. Esse valor é conhecido como **frequência de Nyquist**. Ao se tentar reproduzir uma frequência maior do que a frequência de Nyquist ocorre um fenômeno chamado *aliasing* (ou *foldover*), em que a frequência é "espelhada" ou "rebatida" para uma região mais grave do espectro.

A figura abaixo representa uma onda de 25.725 Hz (em amarelo) digitalizada com uma taxa de amostragem de 44.100 Hz. Cada amostra é representada pelos pontos verdes. A onda em azul é a onda resultante do efeito de *aliasing*.

A figura abaixo apresenta o efeito de *aliasing* (ou *foldover*) descrito acima:



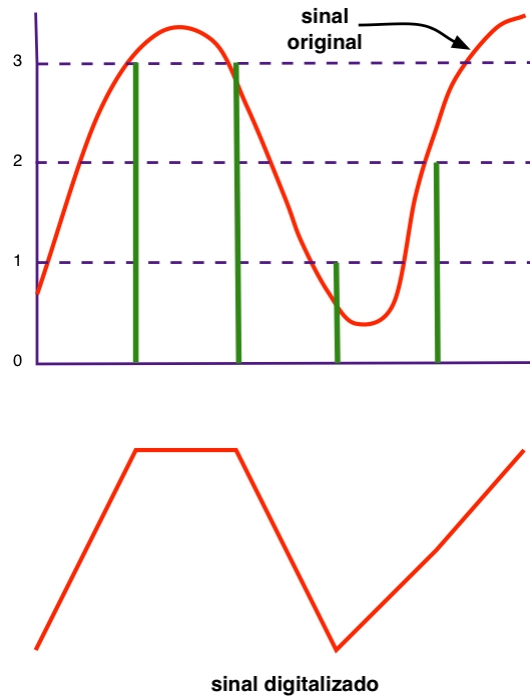


Assim, como ouvimos numa faixa que vai aproximadamente de 20 a 20kHz, uma taxa de amostragem deveria ser de pelo menos 40kHz para que todas as frequências audíveis pudessem ser registradas.

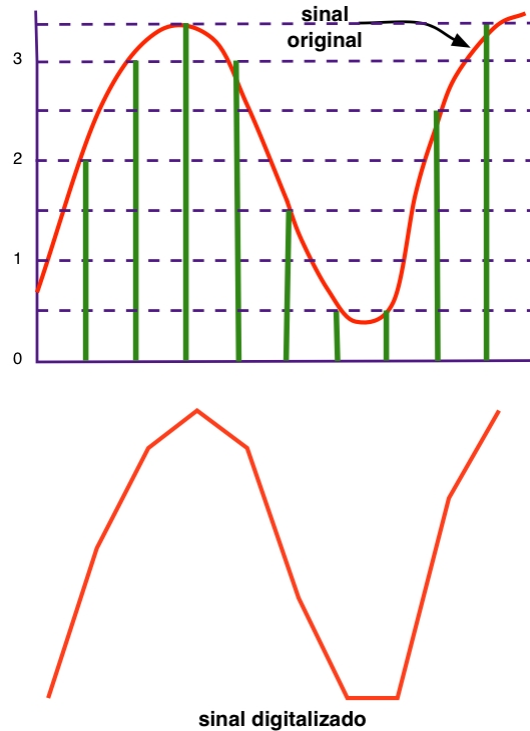
Taxas maiores permitem o uso de filtros com decaimentos mais suaves que causam menos distorções de fase, especialmente nas frequências mais agudas.

Resolução

Refere-se ao número de bits usados para representar cada amostra. Uma amostra representada por apenas um bit poderia receber apenas dois valores: "0" ou "1". Já uma representação com 3 bits poderia receber 8 valores diferentes ($2^3 = 8$): 000, 001, 010, 100, 110, 101, 011, 111. Um CD tem uma resolução de 16 bits o que permite uma resolução binária com 65.534 (2^{16}) valores.



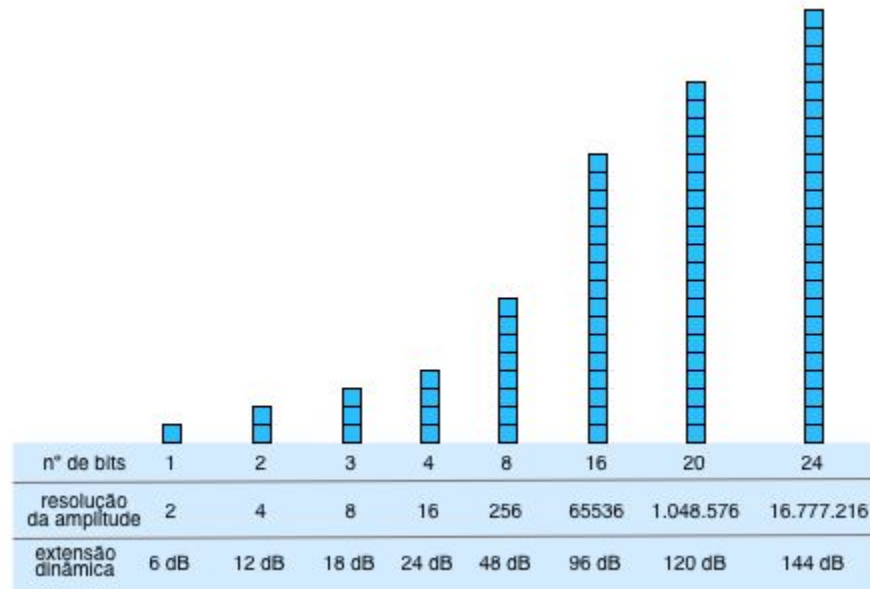
No gráfico acima, a digitalização com uma taxa de amostragem e resolução muito baixas gera uma representação muito distorcida do sinal original.



Com o aumento da taxa de amostragem e da resolução, a onda representada se aproxima cada vez mais da forma de onda do sinal original.

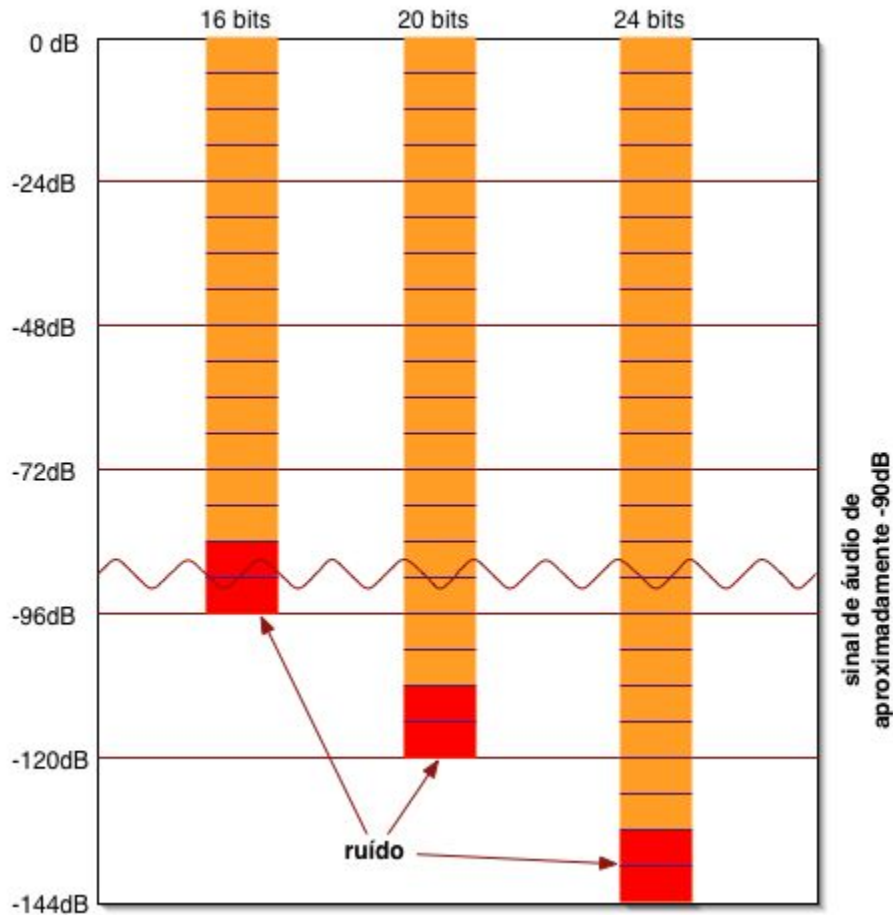
Faixa de Extensão Dinâmica

Cada bit acrescentado na resolução dobra o número de passos (ou valores) usados para representar a variação de amplitude da onda e com isso adiciona 6dB na escala de dinâmica que pode ser representada. Resoluções mais altas oferecem também maior relação sinal ruído.



Relação Sinal/Ruído

É a diferença, em dB, entre o nível máximo de amplitude que pode ser representado numa determinada resolução e o ruído do sistema. Quanto maior a resolução, ou seja, quanto mais bits são usados para representar a amplitude do som, maior será a diferença entre o nível mais alto de reprodução e o ruído.

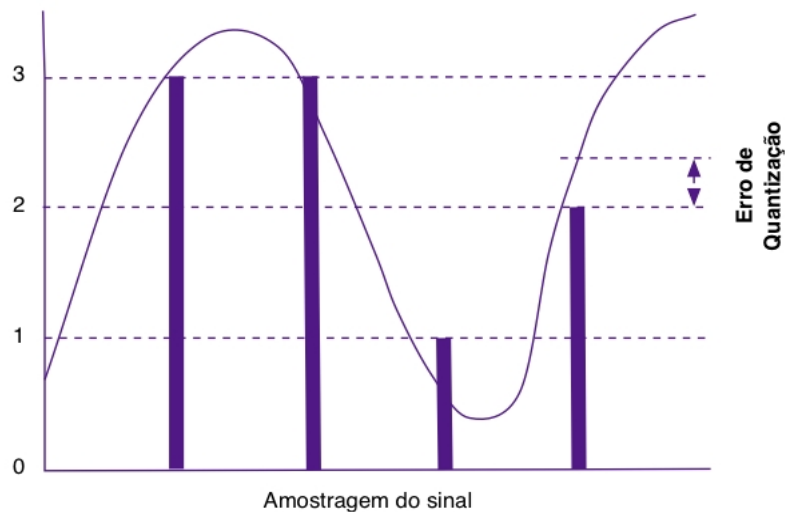


Embora sistemas com 16 bits sejam suficientes para representar áudio com boa qualidade, às vezes é desejável ter alguns bits extras. Na realidade o sistema nunca usa todos os bits para a representação da amplitude do sinal. Num conversor de 16 bits são gerados de 3 a 6 dB de ruído, o que já "rouba" 1 bit da resolução e diminui a faixa de dinâmica usável de 96 dB para 90dB. Se o material musical tem uma média de 78dB com picos ocasionais em 90dB, na maior parte do tempo o sinal não estará usando toda a faixa dinâmica possível, reduzindo em um ou dois bits (6 a 12dB) o outro extremo da escala. Na melhor das hipóteses, boa parte do tempo o sistema estará utilizando apenas 13 ou 14 bits de resolução disponível.

Deve-se notar também que quando o áudio é processado, são realizadas operações matemáticas em cada uma das amostras (*samples*) digitalizadas. Como os números que representam essas amostras são finitos, a cada operação é introduzido um pequeno erro. Quando o sinal passa por sucessivas transformações ou por transformações que envolvem operações complexas, esses erros vão se acumulando e passam a ser audíveis na forma de ruído. Quanto maior a resolução de amostragem, menores (e menos audíveis) serão esses erros.

Erro de quantização

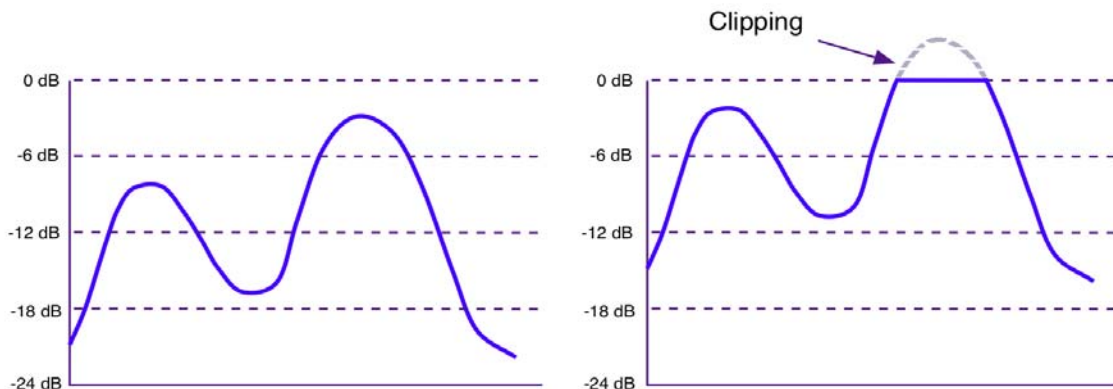
Quando é feita a amostragem do sinal, o valor medido é aproximado (quantizado) para o patamar mais próximo na escala de amplitude gerando pequenos desvios em relação ao valor do sinal original. Esses desvios, chamados erros de quantização modificam o sinal original introduzindo ruído nas frequências mais altas. Pode-se minimizar os erros de quantização com o aumento da resolução em bits.



Alguns sistemas introduzem um processo chamado *dithering* que é a adição de ruído aleatório ao sinal para distribuir os erros e minimizar os efeitos auditivos causados por eles.

Clipping

Uma vez que a extensão dinâmica do áudio digital é determinada pelo número de bits utilizados, não é possível representar valores acima de um determinado limite. O valor mais alto que pode ser representado geralmente é expresso como sendo 0 dB. Se a amplitude da onda ultrapassa esse valor, ocorre um corte (*clipping*) da crista da onda, mudando sua forma original e ocasionando uma distorção do som.



Tamanho de Arquivos

Resoluções e taxas de amostragem maiores implicam em arquivos maiores e que precisam de mais espaço para serem armazenados, mais tempo para serem transmitidos e mais poder de processamento para que sejam processados. Para se calcular o tamanho em bytes de um arquivo pode-se usar a seguinte fórmula:

$$TA * R/8 * C * t$$

Onde:

- **TA** = taxa de amostragem em Hz
- **R** = resolução em bits (como queremos o valor em bytes e cada byte tem 8 bits, é preciso dividir por 8)
- **C** = número de canais de áudio
- **t** = tempo em segundos

Assim, num CD em que o áudio é armazenado com 44,1 kHz/16 bits, em dois canais (estéreo), um minuto de música ocuparia aproximadamente 10Mb de espaço:

$$(44.100 \text{ Hz}) \times (16 \text{ bits} / 8) \times (2 \text{ canais}) \times (60 \text{ segundos}) = 10.584.000 \text{ bytes,}$$

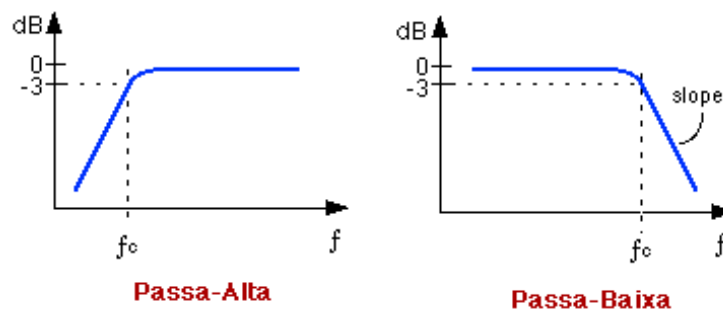
ou aproximadamente 10 Mb.

Filtros

A princípio, qualquer operador ou dispositivo que modifique um sinal de áudio pode ser considerado um filtro. De um modo mais explícito, um filtro atenua a quantidade de energia presente em certas frequências ou faixas de frequências de áudio. Desse modo, superfícies ou quaisquer obstáculos presentes no meio de propagação de uma onda sonora podem atuar como filtros mecânicos, uma vez que, ao proporcionarem a reflexão ou absorção de certas faixas de frequência, alteram as características das ondas sonoras. Do mesmo modo, os botões que controlam a quantidade de "graves" e "agudos" presentes em aparelhos de som são filtros elétricos. Sistemas mais sofisticados são implementados em equalizadores nos quais se pode controlar com maior precisão as faixas de frequências que serão afetadas na filtragem. Filtros digitais podem também ser implementados na forma de algoritmos em computadores e outros aparelhos digitais.

Existem muitos tipos de filtros, com características e aplicações bastante diferentes. De modo geral, esses filtros podem se enquadrar num dos tipos a seguir:

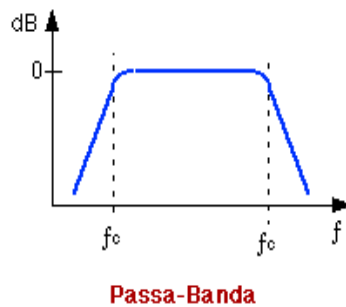
Passa Alta e Passa Baixa: Como os próprios nomes indicam, são filtros que permitem a passagem do sinal de áudio acima (Passa Alta, *High Pass*, ou *Low Cut*) ou abaixo (Passa Baixa, *Low Pass*, ou *High Cut*) de uma determinada frequência, chamada **frequência de corte** ou **frequência de sintonia** (de fato, a frequência de corte situa-se no ponto onde há uma atenuação de 3dB em relação ao sinal que não está sendo filtrado). Esses filtros são comuns em alguns consoles de mixagem usados em estúdios de gravação, em processadores de efeito e diversos outros dispositivos de áudio.



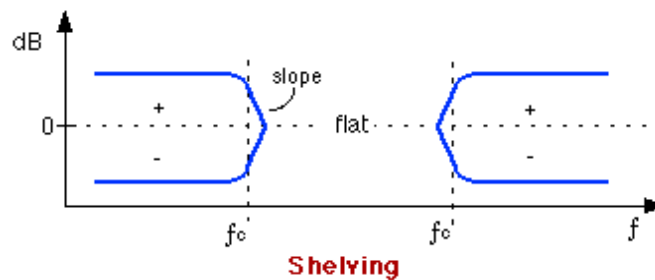
SLOPE: Note que nesses filtros não ocorre um corte abrupto, mas sim progressivo a partir da frequência de corte. A inclinação da curva, a partir da frequência de corte, define o parâmetro chamado *slope*. Quanto mais abrupta a queda de energia a partir da frequência de corte, maior o *slope*; quanto mais suave, menor o *slope*. Os valores do *slope* são dados geralmente em dB/8va, ou seja, a quantidade de decibéis atenuados a cada oitava além da frequência de corte. Às vezes o *slope* é referido também por seu número de "ordem", sendo que cada ordem equivale a 6 dB/8va, conforme a tabela abaixo. Como regra geral, quanto maior a ordem do filtro, mais preciso o controle que ele oferece. Entretanto, filtros de ordem alta são mais caros e difíceis de serem implementados.

slope = dB/8va
1ª ordem = 6 dB/8va
2ª ordem = 12 dB/8va
3ª ordem = 18 dB/8va
4ª ordem = 24 dB/8va

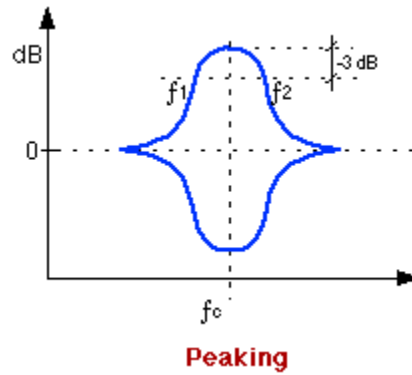
Passa Banda: É uma combinação dos filtros Passa Alta e Passa Baixa em que se pode definir uma frequência de corte inferior e outra superior. É utilizado nos divisores de frequências (*crossovers*) de caixas acústicas e sistemas de som.



Shelving: Este filtro opera em todas as frequências a partir da frequência de corte de maneira plana. Ou seja, ele cria uma transição entre uma região extrema (no extremo grave ou no extremo agudo) do espectro de áudio e a região central, conforme o gráfico abaixo. Esse é o tipo de filtro que se encontra em aparelhos de som caseiros nos controles de graves e agudos.



Peak: Geralmente encontrado nos equalizadores gráficos e consoles de mixagem, permitem um maior controle das faixas de frequências a serem filtradas. Três parâmetros definem sua atuação: frequência de corte; quantidade de atenuação; **Q** (*quality*), as vezes referido como ressonância do filtro. Filtros que permitem o controle desses três parâmetros são chamados paramétricos; aqueles que possuem as frequências de corte fixas são chamados semi-paramétricos.



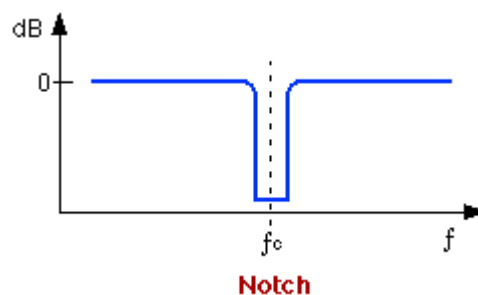
O **fator Q** pode ser entendido como a ressonância do filtro. Quando Q é alto, a curva de ressonância do filtro é estreita atuando numa faixa limitada de frequências. Quando Q é baixo, a curva de ressonância é mais larga e o filtro atua numa faixa maior do espectro de frequências. Q pode ser definido como a razão entre a frequência de corte e a largura da banda da curva de ressonância (essa largura é definida pelos dois pontos que estão 3 dB abaixo da frequência de corte):

Fator Paramétrico Q

$$Q = \frac{f_c}{f_2 - f_1}$$

$$f_c = \sqrt{f_2 \times f_1}$$

Notch: São filtros capazes de rejeitar uma faixa bastante estreita de frequências. Sua utilização é recomendada quando o sinal a ser atenuado é bem definido. Pelo fato de atuar em faixas reduzidas de frequências, filtros *notch* interferem pouco na qualidade do sinal.



Efeitos

Podem ser divididos em 7 categorias principais:

1 - delay;

2 - phase;

3 - flange;

4 - chorus;

5 - reverb;

6 - pitch shifting/harmony;

7 - amplitude

Todos eles, com exceção de efeitos de *pitch shift* e amplitude, são resultado de variações temporais (*delays*). A figura abaixo mostra os efeitos gerados para cada tempo de delay.

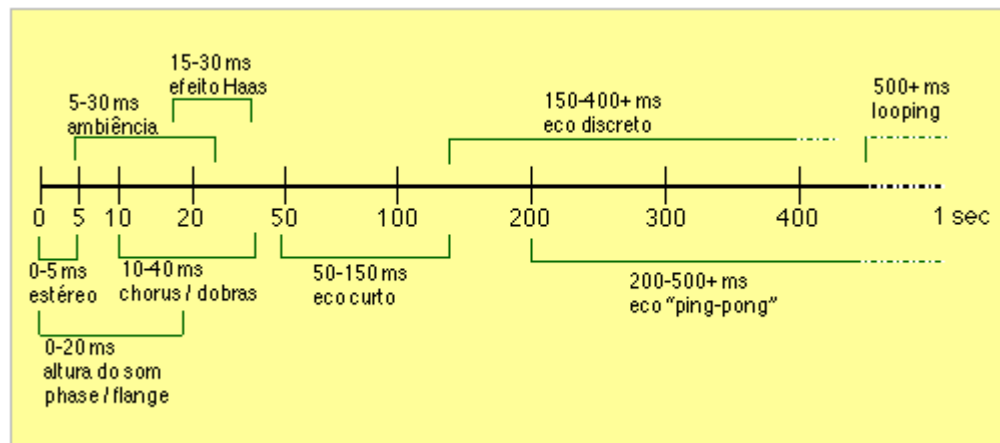


Figura: Tempos de Delay

1 - Delay

Descrição:

Geralmente gerado pelo armazenamento do sinal de áudio em um buffer eletrônico por um certo período de tempo para depois ser reenviado para a saída de áudio. O efeito mais simples é conseguido pela soma do sinal original com o sinal atrasado. Delays múltiplos podem ser gerados pela reinserção repetida do sinal atrasado. Multitap delays são gerados a partir de um único e longo delay que é repetido em intervalos diferentes, gerando múltiplas repetições. Ping-pong delays são obtidos pelo direcionamento alternado de cada repetição para os canais esquerdo e direito da saída de áudio.

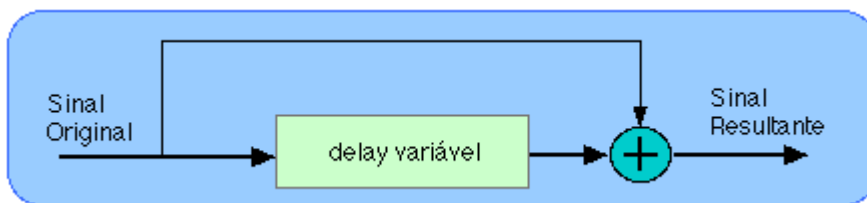


Figura: Diagrama de Efeito Delay

Parâmetros:

Delay time: controla quanto tempo o buffer vai atrasar o som, ou seja, quanto tempo vai decorrer entre o sinal original e as repetições.

Feedback : controla a quantidade de sinal atrasado que vai ser reinjetada na entrada do efeito. Aumentar o feedback significa aumentar o número de repetições e a o tempo de decaimento do efeito.

Filtro passa-baixa : Em ambientes acústicos reais, as frequências mais altas são atenuadas nos sons atrasados, e essa atenuação aumenta proporcionalmente ao número de repetições. Para simular esse efeito usa-se um filtro passa-baixa a cada repetição do sinal.

Tap-tempo: alguns aparelhos oferecem um botão onde se pode "clicar" em um determinado andamento para programar o tempo de delay.

Aplicações:

- transformar um som mono em estéreo, tornando-o mais "cheio"
- looping
- "dobra" de vozes (20 a 40 ms)
- ecos

2 - Phase

Os períodos das oscilações em ondas sonoras na faixa audível (20Hz - 20kHz) varia entre 50ms e 0,05ms. Portanto, defasagens nessa faixa de tempo irão interferir nas oscilações de frequências periódicas (cancelamento de fase). Esse atraso relacionado às frequências sonoras é a base para 3 tipos de efeito: phase, flange, chorus (a diferença entre eles está ligada ao tempo de delay).

Descrição:

O efeito de phase emprega atrasos muito curtos na faixa de 1 a 10 ms. Quando o sinal original é atrasado em relação ao sinal repetido ocorre um efeito conhecido por comb filter no qual as frequências cujos períodos estão diretamente relacionados ao tempo de atraso são atenuadas e reforçadas devido ao cancelamento de fase. Efeitos de phase utilizam um determinado número de filtros para gerar o efeito comb. Usando um modulador (LFO) para mover esse filtro dentro de uma determinada região do espectro causa um cancelamento de fases variável dependente das frequências usadas. Esse efeito é conhecido como phase.

Parâmetros:

Rate (ou *speed*): determina a velocidade com o que o modulador irá varrer ciclicamente a faixa de espectro determinada.

Range: determina essa faixa do espectro a ser varrida pelo modulador.

Outros: filtros, feedback loop.

3 - Flange:

Descrição:

Esse efeito é semelhante ao phase e foi usado pela primeira vez em uma gravação pelo inovador guitarrista Les Paul. O efeito era alcançado com dois gravadores magnéticos contendo o mesmo material sonoro fazendo com que um dos gravadores diminuísse ocasionalmente a rotação para gerar uma diferença de fase entre os sinais. Nos sistemas digitais, o flanger é obtido de modo semelhante ao phaser, com atrasos de 1 a 20ms e um modulador que varia o atraso (regular ou aleatoriamente).

A diferença entre phase e flange é que neste último a atenuação e o reforço das frequências ocorrem em intervalos regulares enquanto que no phase isso depende da disposição dos filtros. Além disso, no phase o espaçamento, a largura e a intensidade (depth) podem ser variáveis. Em geral, flange tem um efeito no campo das alturas mais pronunciado que o phase.

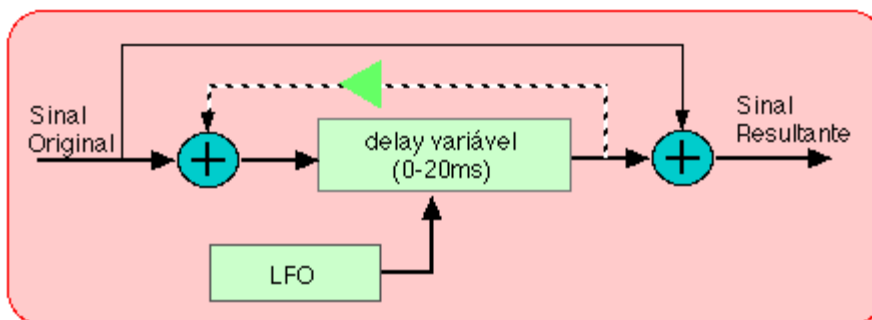


Figura: Diagrama de Efeito Flanger

Parâmetros:

Delay: Controla o tempo de delay

Feedback: Controla a quantidade de sinal processado que é reinjetada no efeito. Alguns permitem determinar se o feedback é positivo (em fase, acentua harmônicos pares, som mais metálico) ou negativo (fora de fase, acentua harmônicos ímpares, som mais "quente").

Rate : Controla a velocidade com que o modulador varia a o delay. Por exemplo, Rate= 0.1 Hz significa que o efeito fará uma varredura de um ciclo a cada 10 segundos.

Depth: Em geral expresso como uma razão, especifica a relação entre o delay mínimo e máximo. Por exemplo, 6:1 pode gerar uma varredura de 1 a 6 ms ou de 3 a 18ms.

Outros : tipo de onda moduladora

4 - Chorus:

Descrição:

Atua introduzindo pequenas variações de afinação no sinal através de um delay gerando o efeito de "dobra" dos sons. Geralmente são produzidos em estéreo, utilizando delays mais longos que o flanger (10 a 30ms) e muitas vezes sem feedback (o que introduz um caráter artificial no som). Existem várias implementações de chorus. Geralmente, são empregados dois delays variáveis modulados pelo mesmo oscilador, mas a saída de um oscilador é invertida antes de ir para um dos delays o que elimina mudanças mais acentuadas de afinação.

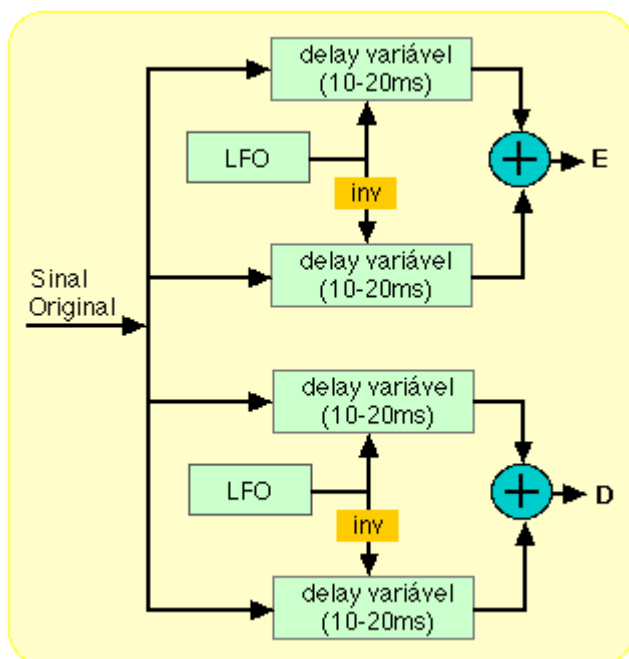


Figura: Diagrama de Efeito Chorus

Parâmetros:

Rate : o mesmo que em flanger.

Depth: o mesmo que em flanger.

5 - Reverb

Descrição:

Sem dúvida o tipo de efeito mais utilizado em processamento de áudio, o reverb simula o espaço acústico no qual o som é produzido. Em um ambiente qualquer, as ondas sonoras são refletidas ao encontrarem uma superfície refletora. Essas primeiras reflexões (early reflections) são seguidas de outras reflexões menos intensas e mais atrasadas em relação ao sinal inicial. A soma de todas essas componentes cria o efeito de reverberação.

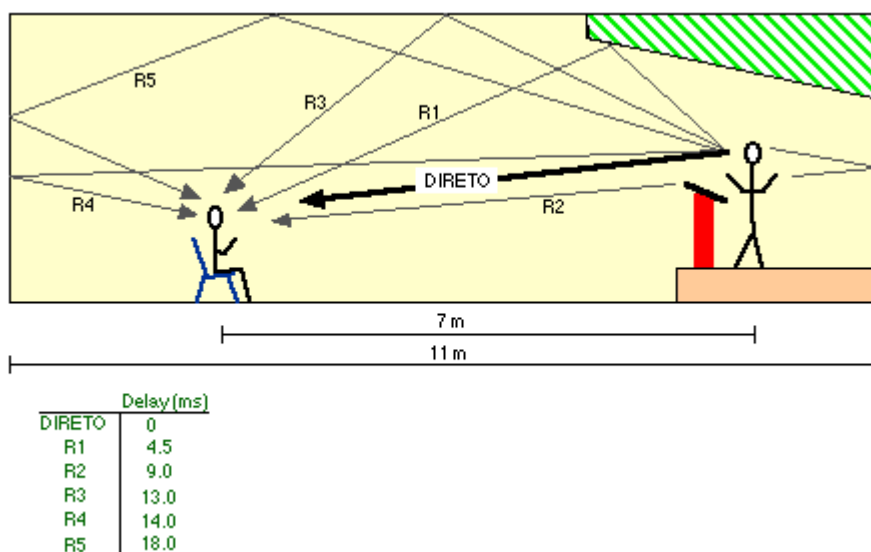


Figura: Early Reflections

Efeitos de reverb são alcançados pela utilização de uma série complexa de delays de um mesmo sinal que diminuem em amplitude e clareza de modo a simular o comportamento acústico de um espaço real.

Geradores de reverb são classificados em relação ao tipo de espaço simulado (room type). Os mais comuns são room, hall, plate e spring.

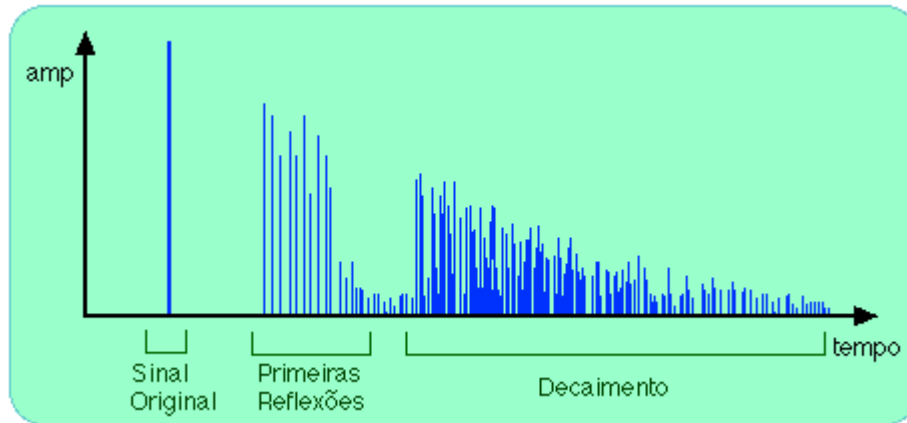


Figura: Esquema de reverberação

Parâmetros:

Size : Determina o tamanho da sala que está sendo simulada pelo efeito, usualmente dado em volume cúbico.

Predelay : Regula um dos parâmetros mais importantes do reverb: o tempo que decorre entre o sinal original e as primeiras reflexões. Isso é muito importante para criar um ambiente natural, já que numa sala real as primeiras reflexões chegam depois do sinal original. O tempo de predelay (em geral, abaixo de 50ms) ajuda a determinar o tamanho da sala. Quer dizer, um predelay curto dá impressão de um ambiente menor, e um predelay longo dá a impressão de um ambiente maior.

Densidade: Trata da quantidade de reflexões e está ligada a quantidade de superfícies difusoras da sala. Quanto maior a irregularidade dessas superfícies, maior o número de reflexões e, portanto, maior a densidade da reverberação.

Difusão: Usado em conjunto (e muitas vezes confundido) com o parâmetro densidade, a difusão trata do modo de decaimento das reflexões, estando ligada portanto às propriedades acústicas das superfícies da sala. Diz respeito aos tempos de reflexão: salas com grande difusão apresentam reflexões em intervalos muito irregulares, enquanto que em salas de baixa difusão os intervalos tendem a ser mais regulares.

6 - Pitch Shifting/Harmony

Descrição:

Entre os efeitos (especialmente os realizados em tempo real), esses são os que exigem os algoritmos mais sofisticados, e até recentemente, os resultados não eram convincentes. Eles funcionam comprimindo ou expandindo o sinal que está sendo processado. Para transpor um som para cima, o sinal é tocado mais rápido, o que o torna mais curto. Então é preciso copiar segmentos do sinal processado e adicioná-lo ao sinal resultante para eliminar essa diferença temporal. Para tornar um som mais grave, o sinal é reproduzido mais lentamente, o que requer o corte de algumas seções do sinal para diminuir sua duração. Ou seja, pitch shifters estão constantemente cortando ou colando pequenas porções do áudio a ser processado. Delays e feedbacks são frequentemente adicionados para criar uma defasagem em relação ao sinal original e não deixar o som muito artificial e uniforme.

Parâmetros:

Transposição: Esse é o parâmetro básico. Em geral existem dois controles: a) um harmônico, que permite transposições em passos de um semitom; b) um ajuste fino, que permite um ajuste em passos menores (geralmente, centésimos de tom).

Outros: Muitas vezes os efeitos de pitch shifting são usados em combinação com outros efeitos, exigindo outros controles como feedback, delay e modulação.

7 - Efeitos de Amplitude

Descrição:

Modificam a amplitude do sinal criando efeitos como tremulo e panning. É uma das poucas categorias de efeito que não empregam algoritmos baseados em transformações temporais. Geralmente um modulador é aplicado à amplitude do sinal que é direcionado para o(s) canal(ais) de saída.

Parâmetros:

Taxa de modulação : Determina a frequência da modulação.

Depth: Determina o quanto o sinal vai ser modulado.