Apêndice: Os aspectos técnicos do som

1 Os aspectos objetivos do som

A definição de som deve considerar o seu aspecto objetivo e subjetivo. Entende-se por aspecto objetivo a sua geração e transmissão, ambos estudados pela acústica. Por aspecto subjetivo, entende-se a percepção do som pelo sistema auditivo e sua interpretação pelo cérebro. A percepção do som é estudada pela psicoacústica e a interpretação pelas ciências cognitivas.

Objetivamente, o som é definido como um fenômeno físico. Assim o som é o movimento organizado de moléculas causado pela vibração de um corpo em um meio material, tal como o ar ou a água [Stevens,80]. Sob o aspecto subjetivo, o som é definido pela psicoacústica como a sensação auditiva produzida pelo ouvido ocasionada pela alteração em pressão, deslocamento ou movimentação de partículas, que se propaga em um meio elástico [Olsen,67]. Pode-se dizer que a definição objetiva trata da causa do som, enquanto que a definição psicoacústica trata do seu efeito.

O som apresenta uma dependência do tempo e da freqüência, ou seja, a natureza do som é temporal e espectral. O som é constituído por uma seqüência de oscilações aproximadamente periódicas da pressão de um meio material, como o ar, que se propagam por esse meio na forma de compressões e expansões sucessivas (ondas longitudinais). A informação sonora está contida tanto na seqüência de padrões de oscilação propagados ao longo do tempo quanto nos padrões oscilatórios num determinado intervalo de tempo.

Iniciando pela análise da natureza temporal do som, a variação da pressão sonora é proporcional ao quadrado da variação da intensidade sonora. Matematicamente,

$$(P_1/P_2) = (I_1/I_2)^2. \qquad (1)$$

Define-se intensidade sonora como a taxa de energia transmitida por segundo, por unidade de área no meio onde o som está se propagando.

O aparelho auditivo humano é capaz perceber intensidades sonoras em ampla escala. Assim, é costume utilizar uma escala logarítmica para expressar essa variação. A unidade dessa escala logarítmica é o Bel, onde:

$$1\ \text{Bel} = \log_{10}(I_1/I_2). \qquad (2)$$

Como o Bel é uma unidade muito grande para descrever sons no contexto da audição humana, tornou-se comum usar uma fração do Bel como unidade de intensidade sonora. Utiliza-se 1/10 do Bel, ou seja, o decibel, cujo símbolo é dB. Tem-se assim que:

$$1\ \text{dB} = 10.\log_{10}(I_1/I_2) \qquad (3)$$

ou, em termos de pressão sonora:

$$1\ \text{dB} = 20.\log_{10}(P_1/P_2) \qquad (4)$$

Conclui-se da equação acima que dobrar a intensidade sonora eqüivale na escala decibel a um aumento de aproximadamente 6 dB, do mesmo modo que diminuir esta intensidade à metade eqüivale a um decréscimo de 6 dB.

O decibel é uma unidade relativa, ou seja, mede a variação de pressão ou intensidade sonora. Para se estabelecer uma escala absoluta de intensidade sonora em decibels deve-se primeiro estabelecer um padrão de referência para o zero dB. Este foi escolhido como sendo, na média, a menor intensidade percebida pelo ouvido humano, para um som senoidal de 1KHz. Esta referência de intensidade sonora é padronizada pela acústica como:

$$I_0 = 10^{-12}\ \text{W/m}^2 \qquad (5)$$

que eqüivale à pressão sonora de,

$$P_0 = 20.10^{-6}\ N/m^2 \text{ ou } 20 \text{ micropascal} \quad (6)$$

Uma intensidade sonora especificada com base nesta referência padrão é chamada de SPL (*sound pressure level*). Similarmente, para especificar o nível sonoro em termos da pressão padrão, usa-se a sigla RMS (*root-mean-square*).

Diz-se que o som é um fenômeno físico de natureza oscilatória, ou harmônica. Isto ocorre porque o som é gerado pela oscilação do deslocamento de um corpo material no ar. A oscilação do deslocamento de sua massa provoca as compressões e expansões do meio, que resultam em som

A natureza oscilatória do som pode ser comprovada tomando o seguinte exemplo. Dada uma barra de metal, aplica-se a esta um golpe com um outro objeto sólido, como uma outra barra de metal. A força exercida por essa colisão pode ser expressa por

$$F = -K.x \quad (7)$$

onde x é a deformação inicial na superfície da barra e k é uma constante elástica do metal. De acordo com a segunda lei de Newton,

$$F = m.a \quad (8)$$

onde m é massa e a é a aceleração. Assim tem-se que:

$$-K.x = m.a = m.(d^2x/dt^2) \quad (9)$$

ou seja:

$$d^2x/dt^2 = -(k/m)x \quad (10)$$

As funções do tipo $x(t)$ que satisfazem a equação anterior são as funções senoidais $\sin(wt)$ e $\cos(wt)$, onde $w = 2.\pi.f$. Uma vez que $\cos(wt) = \sin(wt+\pi/2)$, pode-se provar que a solução genérica é dada por

$$x(t) = I.\sin(2.\pi.f.t + \phi) \quad (11)$$

onde I é uma constante, f é a freqüência e ϕ é a fase.

A função x(t) descreve o som mais simples possível, cuja oscilação é periódica e senoidal ao longo do tempo, que é mostrado na figura ao lado. A este som simples dá-se o nome de componente sonora, ou apenas componente, que é um som de natureza senoidal, de intensidade I, dada em dB e freqüência f, dada em Hertz, inversamente proporcional ao período da oscilação da intensidade sonora

$$P = 1/f \quad (12)$$

Onde o período P é dado em segundos.

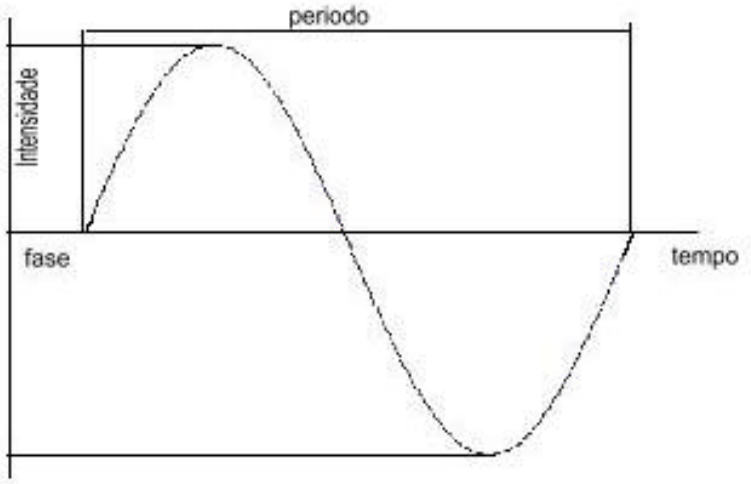


Figura A.1. A componente sonora no domínio do tempo.

A componente sonora é por si só também um som, na verdade é o som mais simples possível pois é constituído por apenas uma componente de oscilação de pressão sonora. Da mesma forma que qualquer sinal periódico no tempo, o som pode ser decomposto e representado por uma somatória de componentes. Como cada grandeza da componente pode variar continuamente no tempo. Atribui-se a cada componente sonora a formula dada abaixo:

$$h(A(t),f(t),\phi(t)) = A(t).\sin(2.\pi.f(t)+\phi(t)) \qquad (13)$$

onde $A(t)$ é a intensidade sonora da componente em dB, f é a sua freqüência em Hertz e $\phi(t)$ é a fase em radianos e. Nota-se que todas as variáveis são também funções contínuas do tempo.

As componentes h_0 também podem ser representadas em termos de fasores. Da fórmula de Euler tem-se a relação:

$$e^{jwt} = \cos(wt) + j.\sin(wt), \quad \text{onde } w = 2.\pi.f \qquad (14)$$

Assim, segue que:

$$h_i(t) = e^{j.wi.t}, \text{ onde } wi = 2.\pi.fi \qquad (15)$$

O som natural, $s(t)$, é formado por diversas componentes sonoras, do tipo:

$$s(t) = h_0 + h_1 + h_2 + \ldots h_N. = \sum_{i=0}^{N} h_i \qquad (16)$$

Nota-se que os componentes podem variar independentemente, em amplitude, freqüência, fase ao longo do tempo, bem como em número de componentes, N. São chamados de sons simples aqueles com apenas uma componente, e sons complexos aqueles compostos por diversas componentes. Os sons naturais são sons complexos.

Qualquer som é formado por uma somatória de componentes sonoras. A representação do som em componentes sonoras vem da teoria desenvolvida pelo matemático francês *Jean Baptiste Joseph Fourier (1768 - 1830)*. A série de Fourier, como é conhecida, prova que qualquer sinal contínuo no domínio do tempo pode ser representado por uma somatória de funções ortogonais, como é o caso

das funções seno e cosseno. Se o sinal for periódico a somatória é finita, caso contrário a somatória é infinita.

Enquanto que no domínio do tempo as componentes h_i são senoides, no domínio da freqüência estas podem ser representadas simplesmente por pulsos:

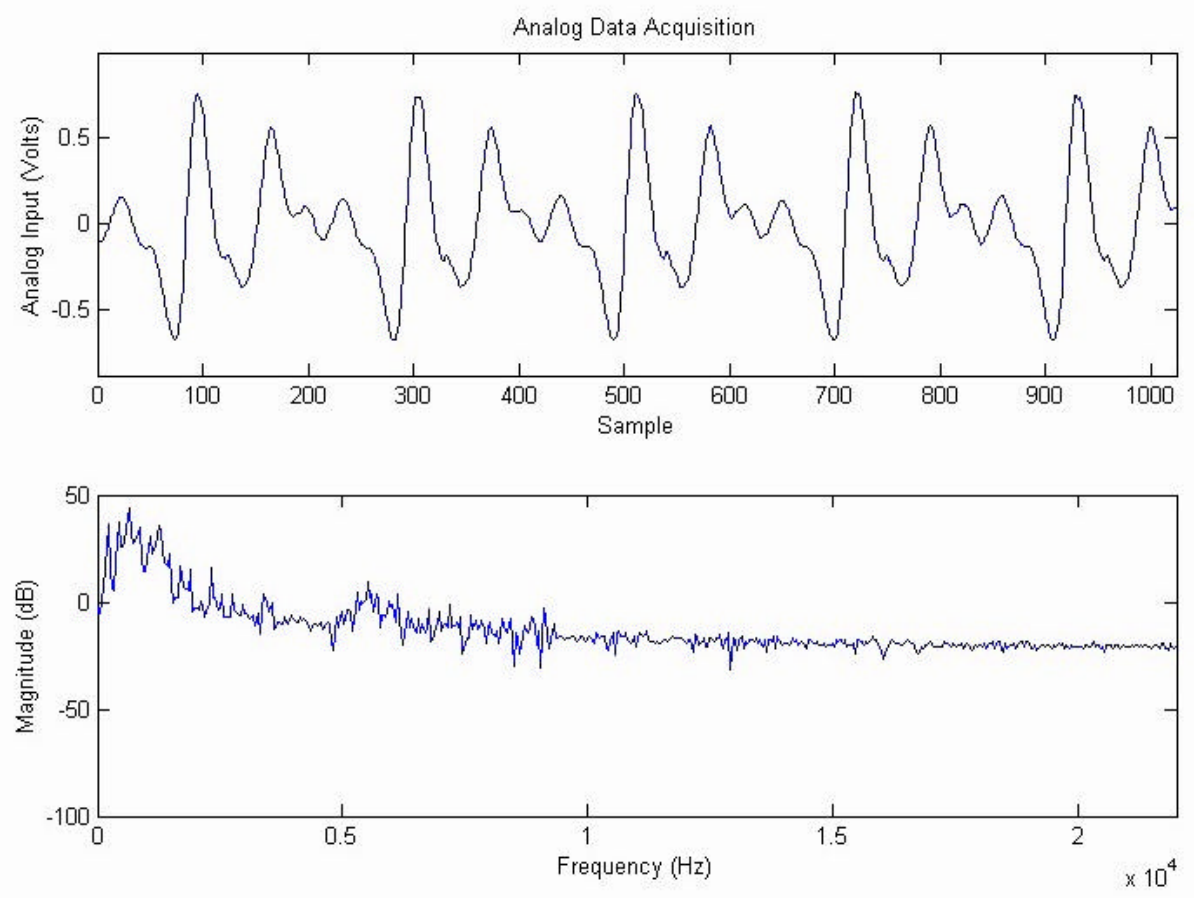


Figura A.2. Exemplo de som natural, no domínio do tempo (acima) e no domínio da freqüência (abaixo).

A representação de um sinal $s(t)$ no domínio da freqüência discrimina as suas componentes em pulsos individuais espalhados ao longo do eixo horizontal. Chama-se de espectro de freqüência ao conjunto formado por essas componentes. No caso do som natural, o espectro de freqüência varia ao longo do tempo. Para este o espectro de freqüência é similar a uma fotografia de um objeto em movimento, que registra um instante de seu deslocamento. É mais fácil analisar a composição de um instante de $s(t)$ no domínio da freqüência que no domínio do tempo. Para sons naturais que variem pouco ao longo do tempo, ou seja, aproximadamente periódicos, o espectro de freqüência se mantém aproximadamente constante ao longo do tempo. O espectro de frequencia é decorrente da serie de Fourier. A transformada de Fourier, vista abaixo, permite representar $s(t)$ no domínio da freqüência, $S(w)$, dado por:

$$S(w) = \int_{-\infty}^{+\infty} s(t).e^{-jwt}\, dt \qquad (17)$$

$S(w)$ é um número complexo do tipo $(a + i.b)$ que representa a componente do sinal em uma dada freqüência $f = w/2\pi$. A magnitude dessa componente é dada por $|S(w)| = (a^2 + b^2)^{1/2}$, e a fase $\phi = \tan^{-1}(b/a)$ [Oppenheim,75].

Até agora vimos a representação do som como sinal contínuo, nos domínios do tempo e da freqüência. No entanto o som também pode ser representado como sinal discreto, conhecido como som digital.

Pela teoria da amostragem é possível representar um sinal sonoro contínuo, do tipo $s(t)$ por uma seqüência de amostras discretas $s(n) = s(t)$, onde $t = n.Ts$ para $n = 1, 2, 3, \ldots, N$. Para amostrar adequadamente um sinal contínuo no tempo $s(t)$ em sinal discreto $s(n)$ é necessário que a taxa de amostragem $Fs = 1/Ts$ seja maior que o dobro da freqüência da última componente f_H, a chamada freqüência de Nyquist [DeFatta,88]. Caso $s(t)$ possua componentes com freqüência $f_H > Fs/2$, tem-se a ocorrência do ruído de *aliasing*, ou aliasamento.

Voltando a representação das componentes sonoras em fasores, tem-se que para sons contínuos no tempo:

$$s(t) = \sum_{i=0}^{N} h_i \qquad (18)$$

e

$$h_i(t) = e^{j.wi.t}, \qquad \text{onde } wi = 2.\pi.fi \qquad (19)$$

Para o som digital, $s(n)$, tem-se:

$$h_i(n) = e^{j.wi.n.Ts}, = e^{j.n.Fs.(wi + k.2.\pi/Ts)} \qquad (20)$$

A componente do som digital se repete em freqüência a cada período $wh = 2.\pi/Ts$ onde as componentes acima de f_H são representadas com freqüência $f_A - f_H$, o que irá gerar componentes que não existiam no som original [Steiglitz96].Tem-se então que $s(t)$ deve ser limitado abaixo da freqüência de Nyquist, que corresponde a ser filtrado por um filtro passa-baixas a uma freqüência de corte inferior a $Fs/2$.

O sinal discreto no tempo $s(n)$ pode ser analisado no domínio da freqüência de duas maneiras. A primeira é pela transformada-Z, dada por:

$$S(w) = \sum_{n=-\infty}^{\infty} s(n).e^{-j.w.n} \qquad (21)$$

A segunda maneira é pela transformada discreta de Fourier, ou DFT (*Discrete Fourier Transform*), dada por:

$$S(k) = \sum_{n=0}^{N-1} s(n).e^{-j.k.n.2.\pi/N} \qquad (22)$$

Ambas possuem transformadas inversas. A diferença entre elas é que a transformada-Z representa o sinal discreto, de extensão infinita, ou não-periódica, do domínio do tempo, para uma representação no domínio da freqüência que é contínua e periódica. A DFT por sua vez representa o sinal discreto e periódico do domínio do tempo na sua representação discreta e periódica no domínio da freqüência.

Ilustrando o que foi visto até agora de ADSP, tem-se o gráfico abaixo:

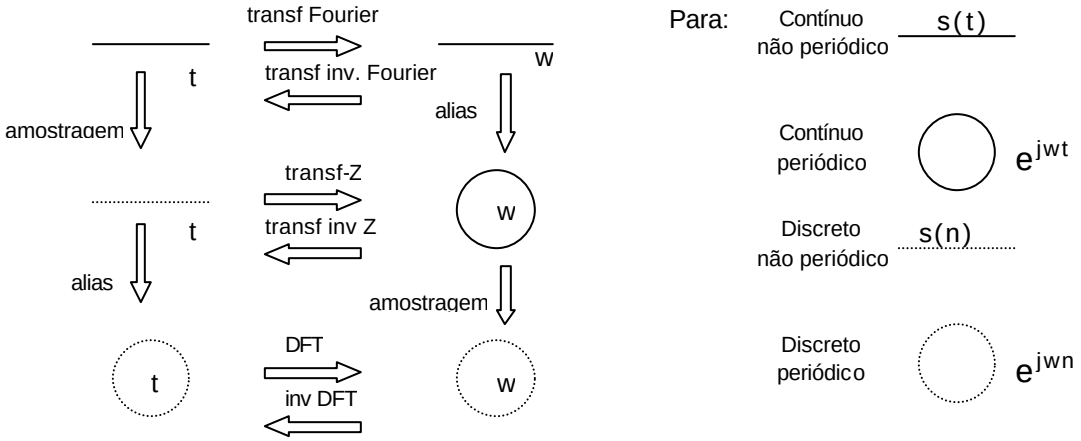


Figura A.3. Relação entre as transformações no domínio do tempo e da freqüência para sinais contínuos e discretos.

Também durante o processo de amostragem de $s(t)$ para $s(n)$, cada instante amostrado de $s(t)$ é representado por um valor inteiro dado por uma palavra binaria com um número finito e fixo de bits. Cada palavra binária possui b bits e permite representar 2^b níveis de intensidade sonora. Este processo é chamado de quantização. A relação sinal-ruído, ou SNR, do som digital quantizado em b bits é dada por:

$$SNR = 20.\log_{10}(2^b) \quad [dB] \qquad (23)$$

No padrão de amostragem utilizado nos CDs (*compact disk*) quantiza o som digital em 16bits, o que eqüivale aproximadamente a 96dB de SNR, e uma taxa de amostragem de 44100Hz, que permite representar sem aliasamento sons com componentes de freqüências até 22050Hz. Este padrão é suficiente para representar o som digital com boa qualidade sonora. No entanto, em processamento de sinais, devido aos arredondamentos das operações aritméticas feitas pelos algoritmos que manipulam o som digital quantizado, tem-se adotado padrões superiores de amostragem e quantização, tais como 24bits de resolução e 96000Hz de taxa de amostragem.

O sinal digital é via de regra mais fácil de ser analisado e processado que o sinal analógico pois a matemática envolvida nos algoritmos de representação e processamento sonoros se reduz a equações a diferenças, com operações de soma e multiplicação, ao contrário dos sinais contínuos no tempo, que são representados por equações diferencias. Também é mais fácil representar e armazenar sons digitais em computadores, o que facilita a implementação de algoritmos computacionais para a manipulação sonora.

2 Os aspectos subjetivos do som

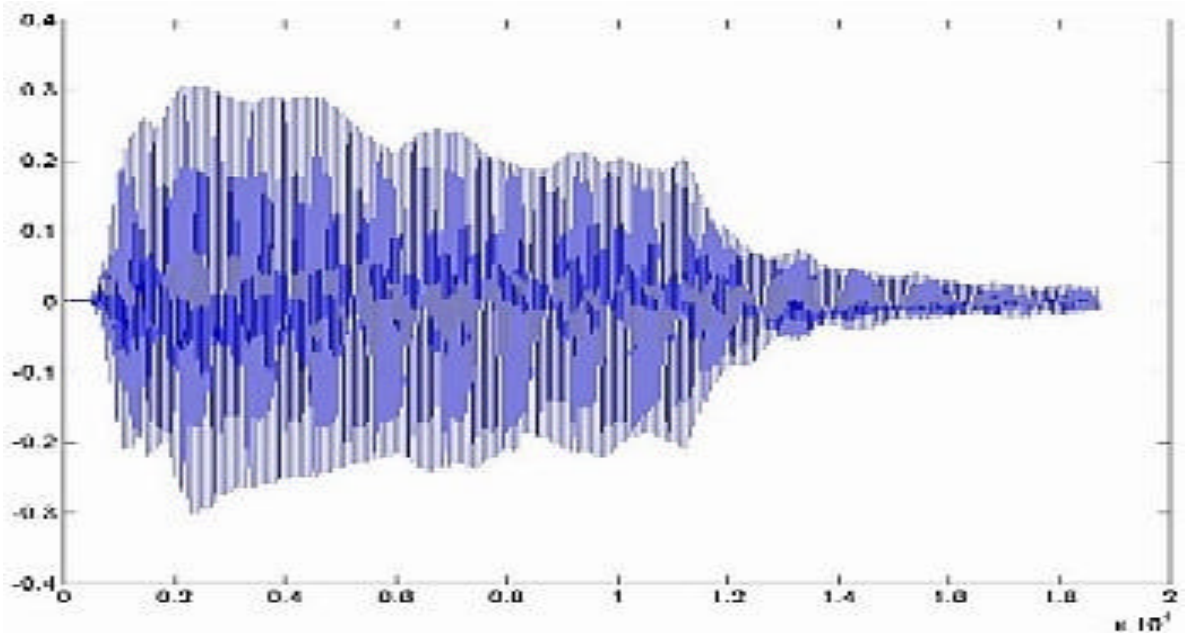
No processo natural de percepção sonora, o sistema auditivo capta a informação que está contida nas oscilações de pressão do ar e a converte de vibrações mecânicas a impulsos elétricos que são transportadas pelo nervo auditivo ao cérebro e interpretadas como a sensação fisiológica chamada de audição. Este processo não é linear. A audição privilegia a percepção de sons importantes para a sobrevivência e o relacionamento humano. Por exemplo, o ouvido é mais sensível a sons parecidos com a fala humana. O processo de audição também é limitado. Percebe-se sons dentro de uma escala de variação de amplitude e freqüência. Pode–se dizer que a percepção auditiva é uma interpretação da realidade acústica. Como tal, o estudo da percepção sonora trata dos aspectos subjetivos da natureza do som.

A percepção do som ocorre no domínio do tempo em dois níveis perceptuais. Por analogia com a visão, chamamos estes níveis de *macroscópico* e *microscópico*. A divisão entre estes se dá

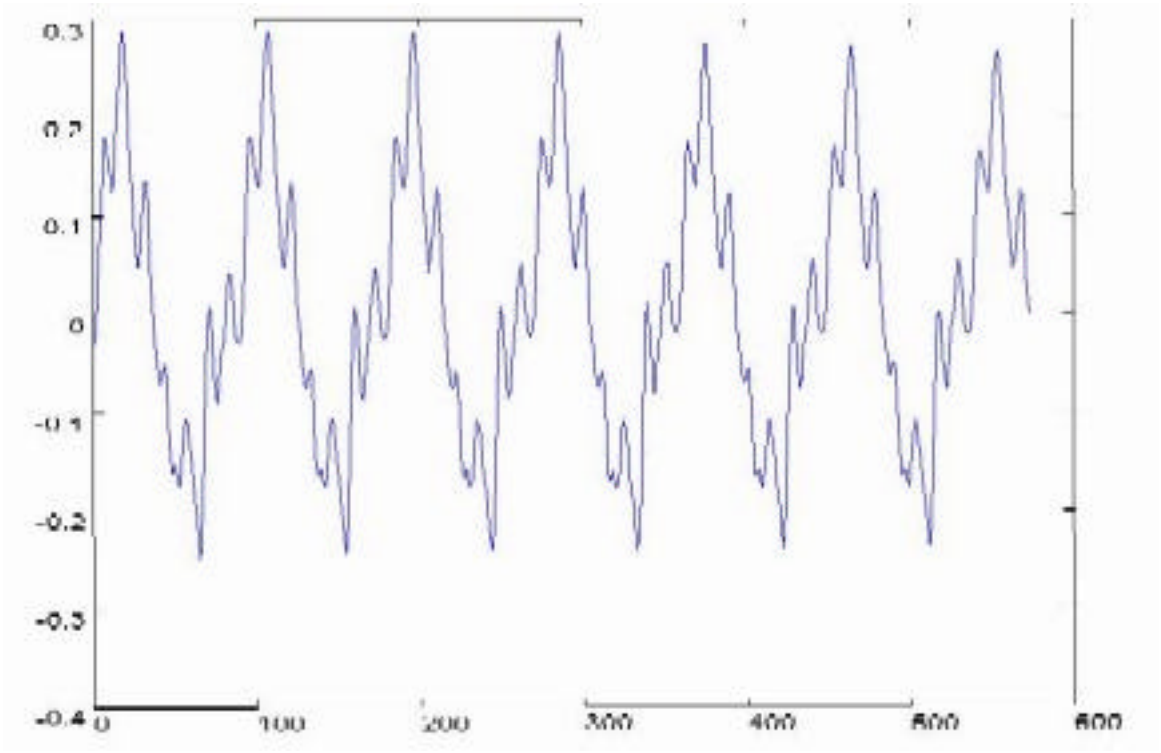
pela definição de um intervalo de tempo conhecido como persistência auditiva. Eventos sonoros que ocorram separados no tempo por intervalos menores que o da persistência auditiva são percebidos pela audição como se ocorressem simultaneamente. O intervalo da persistência auditiva médio é 30ms.

A percepção macroscópica leva em conta a organização temporal ou rítmica do som. È neste nível de percepção que a audição reconhece ritmos, melodias, silabas e palavras, A percepção macroscópica não leva em conta o timbre do instrumento ou da voz, que é definido adiante, desse modo é possível para a audição reconhecer silabas e palavras pronunciadas por diferentes indivíduos, com timbres de voz diferentes.

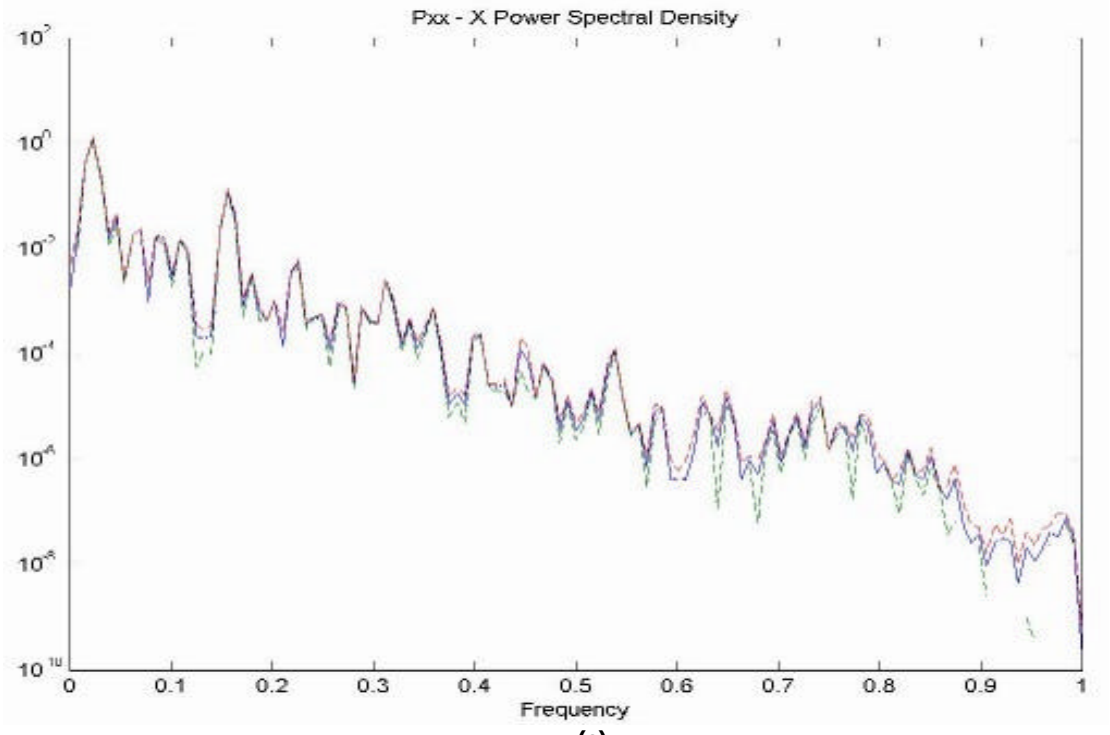
Já no nível de percepção microscópico, a audição reconhece o timbre, ou seja, as características estruturais do som, como seu ataque e a sua composição espectral. Através da percepção microscópica a audição reconhece, por exemplo, a diferença das vozes de indivíduos ou a diferença entre o som de instrumentos musicais, mesmo tocando a mesma frase ou nota musical.



(a)



(b)



(c)

Figura A.4. Exemplo dos níveis (a) macroscópico (b) microscópico e (c) espectro de freqüência do som de uma nota emitida por um violoncelo.

A psicoacústica é a ciência que estuda a percepção do som pela audição humana, levando em consideração seus limites e não-linearidades. A percepção das grandezas acústicas é estudada pela psicoacústica de modo a fornecer um mapeamento de cada grandeza em relação à sua percepção

subjetiva. Deste mapeamento surgem as grandezas psicoacústicas. Para a percepção da intensidade sonora, tem-se o *loudness*. Para a percepção da freqüência, tem-se o *pitch*. Para a percepção das componentes em freqüência o tem-se a distribuição espectral. Além destas grandezas perceptuais, o sistema auditivo é composto pelos dois ouvidos, percepção também chamada de bi-audição. Esta permite reconhecer a localização espacial de uma fonte sonora, pela diferença de tempo de chegada do som a cada ouvido, bem como por outros detalhes, como ecos, reverberações e reflexos na estrutura da orelha e no ombro do ouvinte [Begault,94].

Cada um dos dois ouvidos é um sistema composto por três partes, a saber: ouvido externo, médio e interno. O ouvido externo é composto pelo pavilhão ou orelha e conduto auditivo. Alêm de proteger as camadas internas do ouvido, este apresenta a propriedade de filtrar o som de modo a realçar as freqüências mais importantes para o reconhecimento da voz humana e ajudar na localização da posição da fonte sonora no espaço. O ouvido médio é composto pelo tímpano, uma membrana que capta o som e o transforma de oscilações de pressão do ar para vibrações mecânicas. O tímpano está conectado a um conjunto de minúsculos ossos associados à músculos, respectivamente de fora para dentro: o martelo, a bigorna e o estribo. Estes acomodam (atenuam ou amplificam) a vibração mecânica e a transportam para o ouvido interno através de uma abertura chamada: janela oval. O ouvido interno é composto pela cóclea e pelos canais semicirculares. A cóclea é responsável pela tradução das vibrações mecânicas vindas do ouvido médio em impulsos elétricos. Dentro da cóclea está o órgão de *Corti*, no formato de uma cunha, conectado à milhares de células cilhadas. Estas são neurônios especializados que respondem com potenciais elétricas à estímulos mecânicos. Na ocorrência de som, o órgão de *Corti* entra em vibração. De acordo com as componentes sonoras presentes no som, partes distintas deste órgão entram em ressonância. As células cilhadas conectadas à região que vibra, respondem gerando impulsos elétricos que são transportados pelo nervo auditivo aos lobos temporais do cérebro e interpretados como percepção sonora.

Do mesmo modo que para outros sentidos da percepção humana, a audição apresenta limites de percepção. Escutamos os sons que ocorrem dentro de uma faixa de intensidade, freqüência e tempo. O limite da percepção de intensidade sonora é dado pelo nível mínimo de percepção sonora, onde o ouvido percebe a existência do som, até o limiar da dor, onde a intensidade sonora é tão grande que provoca sensação de desconforto ou dor no ouvinte. A percepção da intensidade está relacionada a freqüência das componentes do som. Para sons simples, com apenas uma componente sonora, a percepção da intensidade sonora varia aproximadamente entre 0dB para o limiar da percepção até 120dB para o limiar da dor. A grandeza da percepção da intensidade sonora é chamada de *loudness*. Experimentos realizados por Fletcher e Munson, [Fletcher,33] demonstraram que para sons senoidais, ou seja, com apenas uma componente sonora, a percepção da intensidade é dependente da freqüência da componente.

A unidade de *loudness* é chamada de *phon* e as curvas cujo *loudness* se mantêm constante são as curvas isofônicas. Estes experimentos foram realizados dentro dos limites de percepção de intensidade e freqüência, ou seja, sons senoidais de intensidades variando entre 0 e 120 dB e freqüência entre 20 e 20000 Hertz. A partir desses dados empíricos estabeleceu-se o que é conhecido hoje como curvas isofônicas de Fletcher e Munson, vistas a seguir.

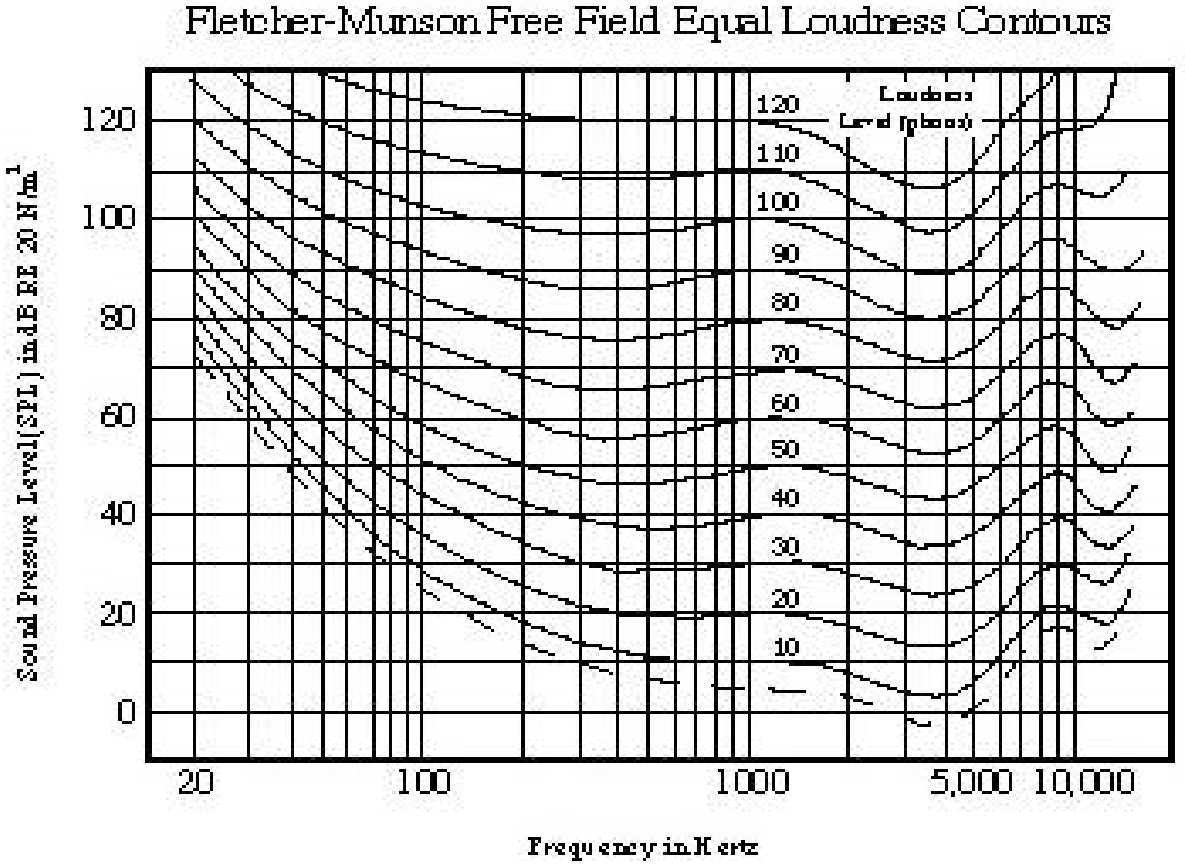


Figura A.5. Curvas Isofônicas de *Fletcher* e *Munson.*

Analisando as figura ao lado pode-se verificar que o ouvido é mais sensível para intensidades sonoras com freqüências medianas, que estão próximas da fala humana. É importante realçar que tais experimentos foram realizados para sons senoidais, que possuem um único componente em freqüência. Na realidade a quase totalidade dos sons que escutamos são sons complexos, compostos por muitas componentes sonoras que variam dinamicamente ao longo do tempo. Assume-se assim que os formatos das curvas isofônicas devem vir a se modificar de acordo com a composição do espectro de freqüência de cada som complexos.

O limite da percepção da freqüência sonora é relacionado ao formato em cunha do órgão de Corti, dentro da cóclea, que é sensível à freqüências sonoras aproximadamente entre 20Hz e 20.000Hz. Para efeito de comparação, as freqüências fundamentais das notas do piano, um dos instrumentos com maior extensão de escala musical, vão de 27,5 Hz para a primeira nota, o A_0, até 4.186 Hz, para a última nota, o C_8. A voz humana varia a freqüência fundamental entre 80 Hz para baixos até 1.000Hz para sopranos.

A percepção da freqüência sonora se reduz com a idade do indivíduo. Entre indivíduos de audição normal, crianças podem escutar até acima de 20 KHz., adolescentes e jovens adultos até 16 KHz. pessoas muito idosas, consideradas com audição normal, podem apresentar esta percepção diminuída para o máximo de 5000Hz [Rosemberg,82].

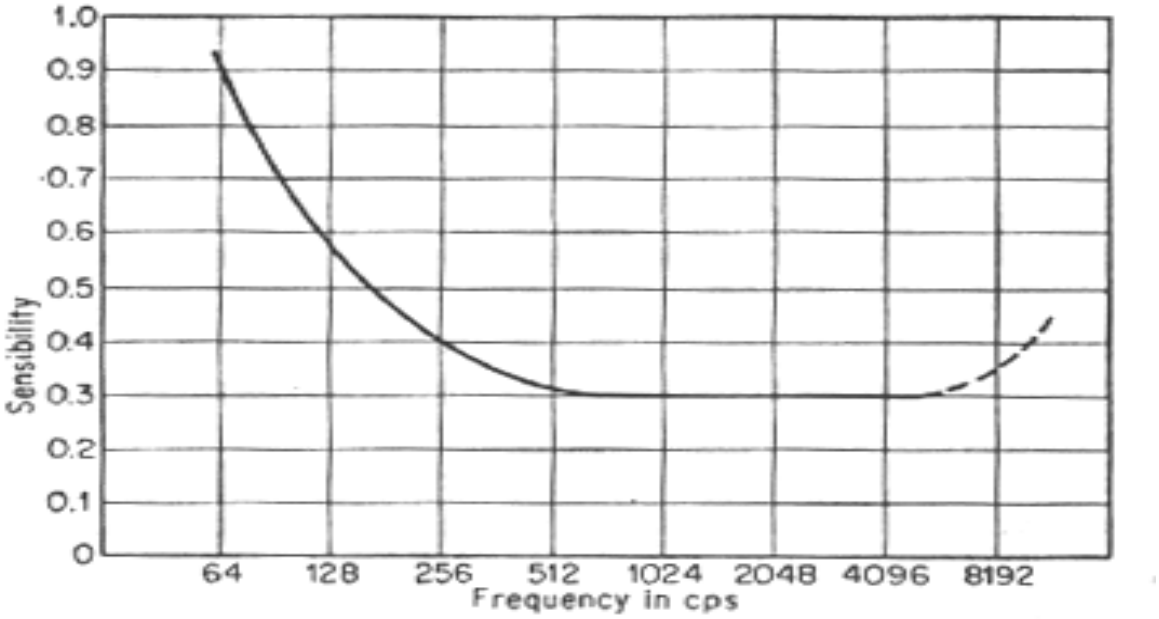


Figura A.6. Percepção da variação de freqüência. [Culver, 68].

Para representar a percepção auditiva da variação de freqüência sonora foi criada a escala Bark de freqüência. Ao invés da escala linear de freqüência em Hertz, a escala Bark apresenta maior resolução para baixas freqüências e menor resolução a medida que a freqüência aumenta . O gráfico da relação entre Bark e Hertz é dada abaixo:

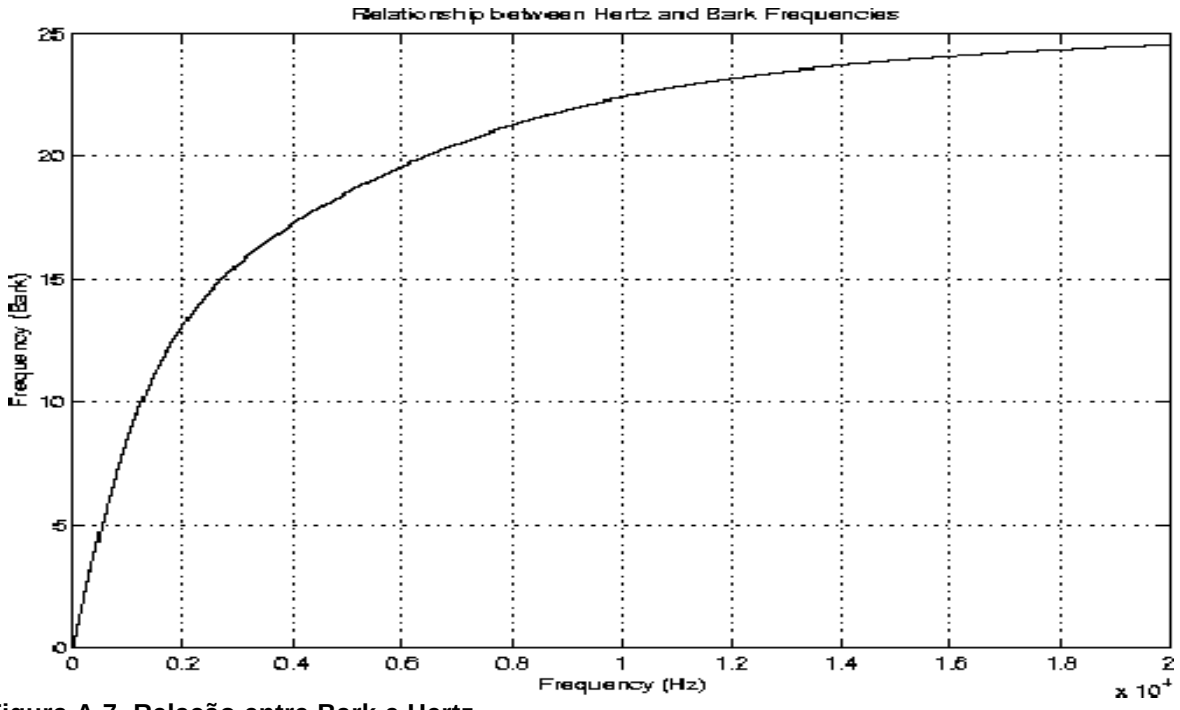


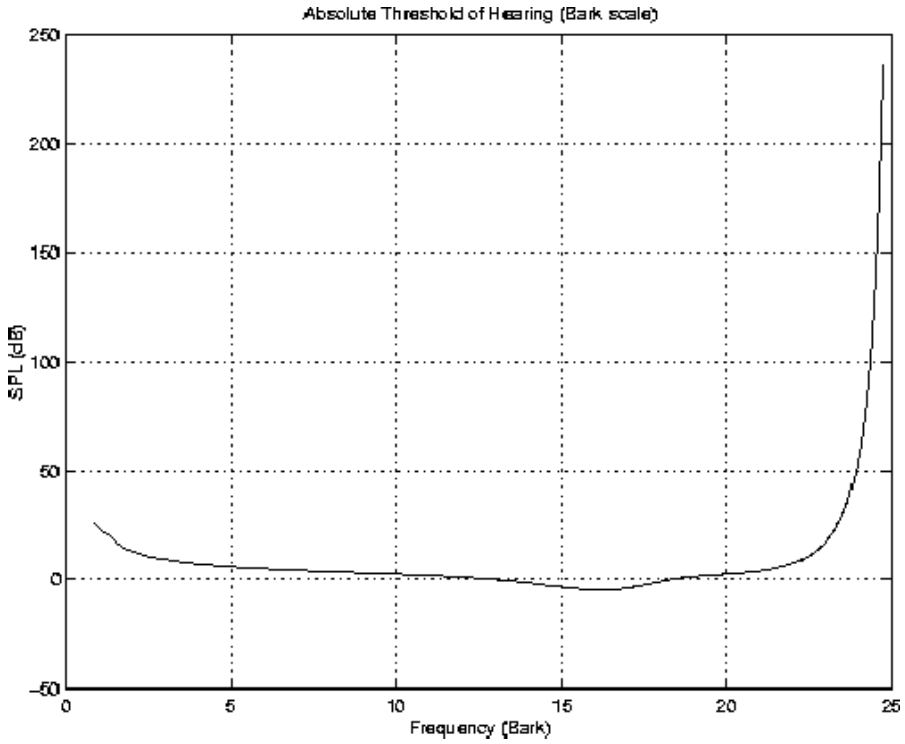
Figura A.7. Relação entre Bark e Hertz.

Diversas aproximações para relação entre Bark e Hertz foram elaboradas. Algumas são dadas pela tabela abaixo:

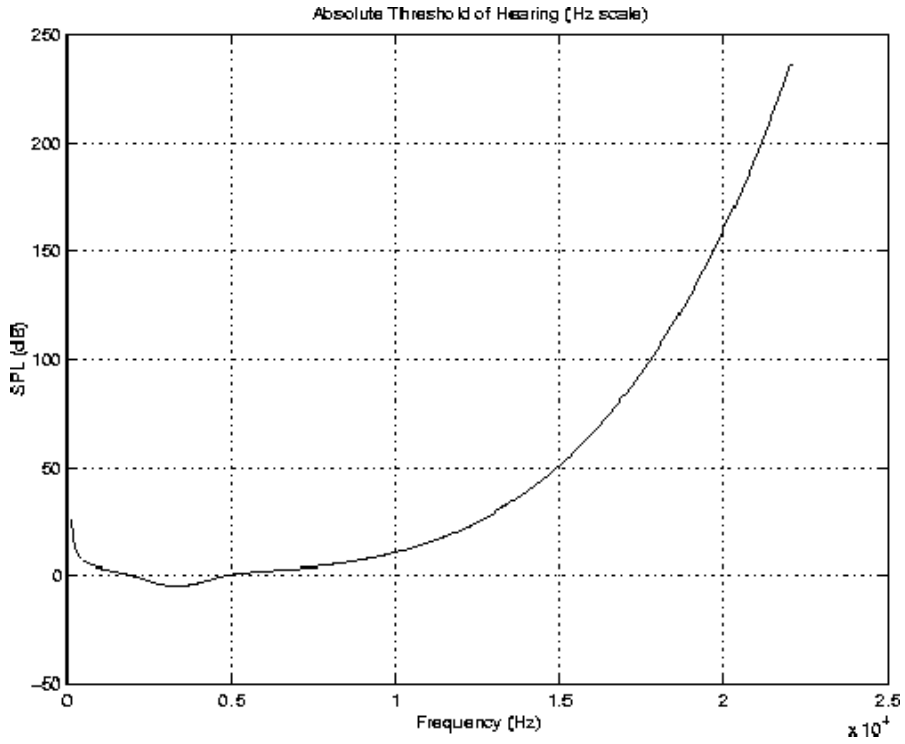
Zwicker & Terhardt (1980)	$B = 13.\tan^{-1}(0{,}76.f/1000) + 3{,}5.\tan^{-1}(f/7500)^2$ $B = 8{,}7 + 14{,}2.\log10(f/1000)$
Terhardt (1979)	$B = 13{,}3\tan^{-1}(0{,}75f/1000)$ $B = 12{,}82\tan^{-1}(0{,}78.f/1000) + 0{,}17(f/1000)^{1.4}$
Wang, Sekey & Gersho (1992)	$B = 6.\sinh^{-1}(f/600)$
Schroeder (1977)	$B = 7.\sinh^{-1}(f/650)$
Traunmüller (1990)	$B = 26{,}81/(1+(1960/f)) - 0{,}53$

onde f é a freqüência dada em Hertz.

A figura abaixo mostra o limiar da percepção auditiva determinado pelos experimentos de Fletcher e Munson, comparativamente em Hertz e na escala de Bark.



(a)



(b)

Figura A.8. Limiar da percepção nas escalas de freqüência (a) Bark e (b) Hertz.

O ouvido humano possui uma grande sensibilidade à variação na freqüência do som. Foram realizadas medidas da sensibilidade à variação da freqüência com público não treinado [Culver, 68]. A

sensibilidade à variação de freqüência, $\Delta f/f$ atinge um máximo de 0,3%, aproximadamente 1/20 de semitom, entre 500 e 4.000 Hz. Essa sensibilidade é essencial para o entendimento da fala humana, por este motivo, ela é maior na região do espectro correspondente a melhor percepção à variação da intensidade.

A grandeza psicoacústica relacionada à percepção da freqüência sonora é chamada de *pitch*. Alguns dicionários definem *pitch* como sendo "o atributo da audição que permite catalogar o som ouvido em uma escala musical". Pollack investigou a habilidade de ouvintes discriminarem o pitch atribuindo notas musicais a sons melódicos (de instrumentos musicais). Chegou-se a conclusão que se pode fazer isso até o máximo de 5 ou 6 notas simultâneas [Pollack,52].

Apesar de estarem intimamente relacionados, o *pitch* e a freqüência sonora não são sinônimos. Como foi dito anteriormente, o som é composto por uma séria de componentes, cada qual com a sua freqüência particular. Uma classe importante dessas componentes é chamada, na terminologia musical, de harmônicos. Diz-se que o som de um instrumento musical melódico é composto por harmônicos, que são, por assim dizer, as componentes "principais" do som. Os harmônicos correspondem às freqüências naturais de ressonância de um som melódico, como aquele emitido por uma corda tencionada ou por um tubo em ressonância. O primeiro harmônico é chamado de "fundamental" e os harmônicos seguintes são chamados por sua ordenação numérica (o segundo, o terceiro, o quarto harmônico, e assim por diante). É atribuído ao harmônico fundamental a freqüência equivalente ao *pitch* do som melódico, embora existam exceções.

Os harmônicos apresentam uma relação em freqüência entre si, do tipo, f, $2.f$, $3.f$, $4.f$, onde f é a freqüência do harmônico fundamental. Esta relação de freqüências é chamada de séria harmônica. A partir dela organizou-se a escala musical. O gráfico abaixo mostra esta relação:

Nota	C0		C1		G1		C2		E2		G2		Bb2		C3	...
Freqüência	f		2.f		3.f		4.f		5.f		6.f		7.f		8.f	...
Intervalo		VIII		V		IV		III		IIIm		IIIm		II		

Os sons que apresentam componentes em freqüência organizados dessa maneira definem um pitch, apesarem de serem sons complexos (como seria o caso, por exemplo, do som da chuva, que não define pitch). Em termos cognitivos, estes são chamados de sons Shepard, que é uma expressão idealizada dos sons de instrumentos melódicos.

Na música o *pitch* é representado pela escala musical. A escala comumente utilizada para instrumentos de teclas (piano, sintetizadores eletrônicos) é a escala temperada cromática. Ela é dividida em 12 semitons, cada semitom eqüivalendo a um intervalo de freqüência de $2^{1/12}.\mathbf{f} \cong 1{,}059463.\mathbf{f}$, ou aproximadamente 6% da freqüência **f** da nota anterior. Cada intervalo é um semitom. O conjunto de 12 semitons perfaz uma oitava na escala musical, que equivale ao intervalo em freqüência de $2^{12/12}\ \mathbf{f} = 2\mathbf{f}$.

Na região de maior sensibilidade, a audição humana pode discriminar variações de freqüência da ordem de 1/20 de semitom, o que corresponderia a uma escala temperada onde cada oitava teria (12*20)=240 notas.

Figura A.9. O *pitch* correspondente às notas de três oitavas da escala musical cromática temperada, de A_4 (440 Hz) até A_7 (3520 Hz).

Timbre é formalmente definido pela *International Standards Organization & American National Standards Institute*, ANSI S1.1-1960(R1976)-12.9*:* como: *"... o atributo da sensação auditiva que permite o ouvinte poder julgar se dois sons similarmente apresentados, com mesmo loudness e pitch, são dissimilares".*

Quando o som possui a variação de sua intensidade aproximadamente periódica no domínio do tempo, este é chamado de "comportado". Exemplos de sons comportados são os sons de instrumentos musicais acústicos melódicos, como a flauta ou o violino. Para essa classe de sons é possível discriminar pela audição o seu harmônico fundamental e assim atribuir a este som uma determinada freqüência, ou nota musical. Já para sons menos harmônicos, como o som de um instrumento percussivo não existe um parcial dominante que possa ser associado a um harmônico fundamental e assim não há como o ouvido perceber e atribuir um *pitch*.

A distribuição espectral representa as componentes que compõem o som. Uma vez que as componentes são variáveis no tempo, o espectro do som é igualmente variável. Assim a distribuição espectral dá a composição aproximada de componentes sonoras em um dado momento de menor variação, ou seja, em um período de tempo onde as suas componentes se apresentam aproximadamente constates. Para sons musicais o instante inicial do som, conhecido como ataque, é normalmente considerado como o momento de maior variação espectral. Em seguida, tem-se um momento onde as componentes se apresentam constantes por um longo período de tempo, isto considerando que não haja mudança de pitch, ou seja, que o instrumento ou voz mantenha a mesma nota musical. A distribuição espectral é como uma fotografia deste momento de constância espectral. Este momento é chamado de "estacionário" e a distribuição espectral é colhida dentro deste período, em uma janela de aproximadamente 50ms, intervalo próximo da persistência auditiva e suficiente para abranger freqüências desde 20Hz até a freqüência de Nyquist, vista anteriormente como sendo a metade da taxa de amostragem do segmento sonoro. Um dos métodos mais utilizados para a obtenção da janela de distribuição espectral é a transformada rápida de Fourier em tempo pequeno, ou STFT.

Alguns pesquisadores definem timbre como uma variável multidimensional, ao contrario do pitch e *loudness*, variáveis unidimensionais, que permitem a classificação seqüencial de sons. Para estes, timbre é: "... aquele atributo da sensação auditiva que permite ao ouvinte diferenciar dois sons complexos que tenham o mesmo *loudness*, *pitch* e duração" [Plomp,70].

3 Métodos de processamento e síntese sonora

Existe uma grande variedade de técnicas de processamento e síntese sonora. Nos concentraremos aqui aos métodos que manipulam o som digital, que floresceram a partir da década dos 70s com os avanços da computação. Estes métodos podem ser classificados em duas categorias: (a) temporais ou espectrais, (b) lineares e não-lineares. Trataremos inicialmente dos métodos de processamento sonoro, que visam manipular o som emitido por uma fonte sonora. Posteriormente trataremos dos métodos de síntese sonora, que geram novo material sonoro. Maiores detalhes sobre estes e outros métodos de síntese podem ser obtidos em diversas referencias, tais como [Miranda,2002].

Processamentos temporais

São os processamentos que manipulam o som no domínio do tempo.

- *Delay* : insere um atraso entre o som de entrada e a saída.
- *Reverber*: simula o efeito de reverberação sonora, ou seja, as múltiplas reflexões geradas por uma fonte sonora em um ambiente específico, como uma sala de concertos, uma catedral ou uma caverna.
- Câmara de Eco: simula o efeito do eco que é a reflexão sonora em um intervalo de tempo grande o suficiente para ser percebido pela audição.
- *Time Stretch*: Modifica a duração do som sem alterar a sua freqüência.

Processamentos espectrais

São os processamentos que manipulam o som no domínio da freqüência.

- Filtros digitais: Manipulam o espectro em freqüência do som de modo a eliminar componentes sonoras indesejáveis ou intensificar componentes que estejam muito atenuadas.
- *Pitch-Shift*: desloca o espectro de freqüência do som no eixo da frequencia, o que corresponde a tornar o som mais grave ou mais agudo, sem alterar sua duração no domínio do tempo.
- *Chorus*: Cria o efeito de coro (várias vozes em uníssono) para um som de entrada. A sensação das vozes em uníssono é dada pelas pequenas diferenças em tempo e frequencia existentes entre cada voz.
- Equalizadores: São constituídos por um banco de filtros passa-faixa que manipulação regiões específicas do espectro de acordo com seus parâmetros.
- Compressão: Limita a variação da intensidade sonora em um intervalo específico. Utilizado normalmente em emissões radiofônicas de gravações sonoras.
- Distorção: Enriquecimento do som de saída através da amplificação não-linear do som de entrada. Este processo gera novas componentes sonoras que não eram presentes no som original.

Síntese Aditiva

A síntese aditiva é uma técnica linear de síntese de sons que se baseia na teoria de *Fourier* e afirma ser possível gerar qualquer função periódica pela somatória de funções senoidais. Na síntese aditiva o som $s(t)$ é gerado através da adição dos sons gerados por osciladores. Para o caso de osciladores senoidais, cada um deles gera uma componente sonora, do tipo visto na equação (11). O som gerado pela síntese aditiva pode ser expresso pela função abaixo:

$$s(t) = \sum_{i=1}^{N} A_i(t).\sin(2.\pi.f_i(t)+\phi_i(t)). \qquad (24)$$

A vantagem da síntese aditiva é permitir controlar independentemente os parâmetros de intensidade, freqüência e fase de cada componente sonora que constitui o som gerado. No entanto, como processo linear, o número de componentes sonoras do som de saída é igual ao número de componentes sonoras geradas pelos osciladores. Como, no geral, os sons de instrumentos acústicos possuem uma quantidade muito grande de componentes sonoras, torna-se computacionalmente caro para a síntese aditiva gerar sons parecidos com os sons naturais.

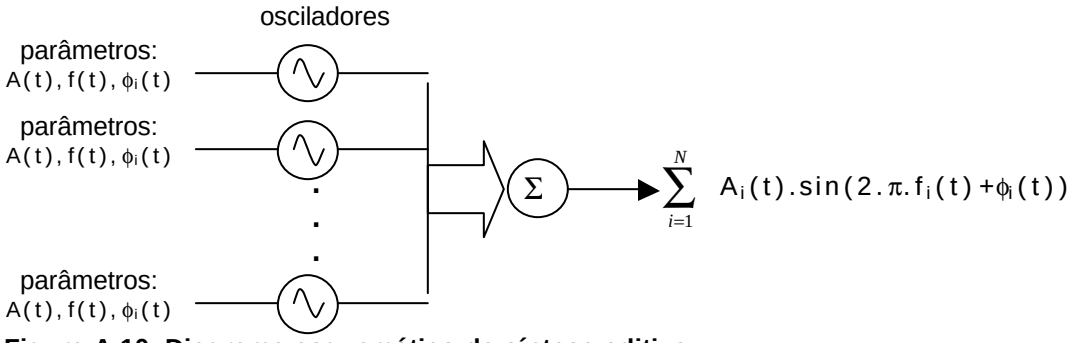


Figura A.10. Diagrama esquemático da síntese aditiva.

Síntese Subtrativa

A síntese subtrativa é uma variação da síntese aditiva. Esta é também uma técnica linear porém que se baseia na subtração de um material sonoro inicial, muito rico em componentes, geralmente através de um banco de filtros passa-faixa. O som inicial pode ser, por exemplo, gerado por um gerador de ruído branco. O resultado sonoro da síntese subtrativa sempre será um som com menos componentes que o som inicial, e portanto mais comportado e reconhecível. Pode-se comparar a síntese subtrativa com a escultura, que remove o material excessivo de um bloco afim de modelar um objeto desejado.

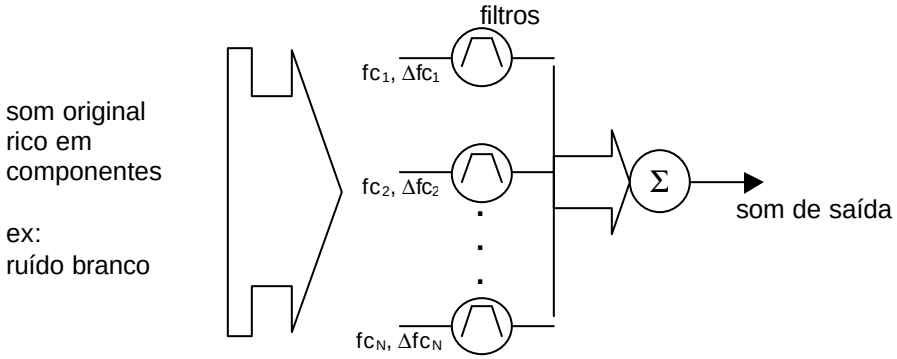


Figura A.11. Diagrama esquemático da síntese subtrativa.

Síntese FM

A síntese por modulação de freqüência, ou FM (*frequency modulation*) é uma síntese tipicamente não-linear que se baseia no controle do parâmetro da freqüência, ou modulação, de um oscilador por outro oscilador. Pode-se ter algoritmos com diversas formas de conexão de osciladores para a geração de sons. Como a freqüência do som resultante não é fixa, mas varia, ou é modulada por outro oscilador, a síntese FM tem como propriedade gerar sons com espectro variante no tempo. A grande vantagem da síntese FM é gerar sons com mais componentes em freqüência que as componentes geradas pelos osciladores. Isto permite gerar sons que se aproximem de sons naturais, porém, como é um processo não-linear, sem a independência de controle de parâmetros, como na síntese aditiva. Um exemplo básico de algoritmo para síntese FM é visto abaixo:

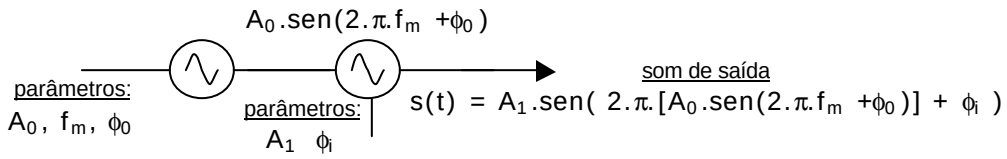


Figura A.12. Exemplo de um algoritmo de síntese FM com dois osciladores.

Síntese Granular

A síntese granular se baseia no conceito de grão sonoro, de algum modo similar ao conceito de *wavelets*, que deriva da teoria de Fourier. Enquanto a teoria de Fourier prova ser possível representar um sinal periódico no tempo através de funções ortogonais, como a família de funções trigonométricas, a teoria wavelet procura representar sinais através de segmentos de sinais finitos no tempo, chamados de wavelets. Do mesmo modo que a síntese aditiva gera sons através da somatória de senoides (que representam as componentes sonoras) a síntese granular gera sons através de segmentos sonoros de poucos milésimos de segundos de duração e muito ricos em componentes sonoras, chamados de grãos. Os grãos sonoros são armazenados em uma tabela do tipo look-up e são utilizados para criar o som de saída

Síntese Wavetable

Utiliza um conceito similar ao da tabela de sons da síntese granular. No entanto o som é armazenado em períodos mais extensos que determinam um som conhecido, normalmente o som de um instrumento musical acústico. A tabela de trechos sonoros é chamada de wavetable, de onde vem o nome deste tipo de síntese. Existem dois tipos de armazenamento do som, em one-shot: para sons não-periódicos e attack-cycle, para sons melódicos ou quase-periódicos. Através de um sistema simples a síntese wavetable consegue simular com grande fidelidade sons conhecidos, razão pela qual é tão utilizada hoje em dia.

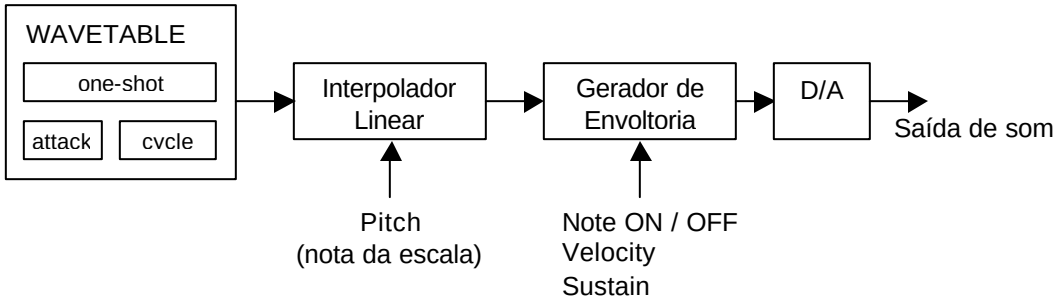


Figura A.13. Diagrama simplificado da síntese *Wavetable*.

A síntese wavetable funciona como um processo de edição sonora em tempo-real. De acordo com os parâmetros de controles o sistema monta o som desejado para uma dada nota musical a uma dada intensidade. Os parâmetros de controle da síntese wavetable são dados em MIDI, a linguagem padrão de comunicação entre instrumentos musicais digitais.

Síntese por modelamento físico (*physical modeling*)

Inicialmente desenvolvida por Julius Smith [Smith,91] a síntese por modelamento físico simula através de equações dinâmicas, ou *wave-guides*, o comportamento de uma fonte sonora, normalmente um instrumento musical acústico. O som gerado é o resultado das equações que modelam o comportamento físico do instrumento. Esta abordagem possibilita uma grande controlabilidade do som gerado porêm torna-se rapidamente complexa uma vez que cada característica do instrumento modelado deve ser descrita por uma equação dinâmica. Enquanto a síntese wavetable gera sons com grande similaridade ao som original mas com pouca controlabilidade (apenas aquelas dadas pelos parâmetros MIDI), a síntese por modelamento físico permite uma enorme controlabilidade sonora porêm a um alto custo computacional. O resultado é que a síntese por modelamento físico, até o momento, ainda é pouco utilizada pela industria.

Síntese por transformações sonoras

A síntese por transformações sonoras foi desenvolvida em nosso trabalho de tese de mestrado [Fornari,95]. Esta utiliza operadores espectrais que modificam o plano espectral do som, que é a magnitude do espectro em frequencia do som em relação ao tempo. Modificações da topografia do plano espectral correspondem a transformações sonoras. A dificuldade deste método consiste em se encontrar famílias de operadores espectrais que provoquem modificações sonoras que sejam interessantes a percepção auditiva, em outras palavras, psicoacusticamente interessantes.

Síntese por decomposição estocástico-determinista

Desenvolvida por Xavier Serra, [Serra,89] esta síntese parte de uma análise inicial do som que o divide em duas categorias: a parte estocástica e a parte determinística. A parte estocástica é composta pelos componentes não-periódicos, ruidosos do som enquanto a parte determinística pelos componentes quase-periódicos que são reduzidos as componentes senoidais principais. Apesar de bastante engenhoso, este método não permite a síntese sonora em tempo real e a redução da parte determinística em senoides implica perda de informação sonora.

Síntese Evolutiva

Conforme foi visto ao longo deste trabalho, a síntese evolutiva reúne o controle intuitivo e a riqueza sonora, antes encontrados separadamente em outros métodos de sínteses. O aprendizado não-supervisionado dado pela computação evolutiva permite que este método de síntese chegue a resultados inusitados, que não eram esperados pelo usuário, e que sempre tendem a evoluir de acordo com os parâmetros sonoros ditados pelo conjunto alvo. Pode-se dizer que a síntese evolutiva delega um pouco da decisão criadora à máquina, o que antes era deixada totalmente ao encargo do usuário. Este passa a exercer a sua criatividade em um nível de abstração mais alto, determinando os parâmetros condicionantes da evolução sonora, onde o sistema da síntese evolutiva irá gerar novos sons.