Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Departamento de Semicondutores,
Instrumentos e Fotônica

SÍNTESE EVOLUTIVA DE SEGMENTOS SONOROS

José Eduardo Fornari Novo Júnior

Dissertação de Doutorado

Orientador **Furio Damiani**
Co-orientador **Jônatas Manzolli**

Comissão Julgadora:

Furio Damiani – FEEC/UNICAMP – Presidente
Adolfo Maia Júnior – IMECC/UNICAMP
Florivaldo Menezes Filho – UNESP
Peter Jürgen Tatsch – FEEC/UNICAMP
Raul Thomaz Oliveira do Valle – IA/UNICAMP

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

N859s

Novo Junior, José Eduardo Fornari
Síntese evolutiva de segmentos sonoros / José Eduardo Fornari Novo Junior.--Campinas, SP: [s.n.], 2003.

Orientadores: Furio Damiani e Jônatas Manzolli
Tese (Doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Inteligência artificial – Processamento de dados. 2. Processamento de som por computador. 3. Audição (Fisiologia). 4. Eletroacústica. 5. Algoritmos genéticos. I. Damiani, Furio. II. Manzolli, Jônatas. III Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Resumo

A síntese evolutiva de segmentos sonoros se baseia nos processos de reprodução e seleção de uma população de indivíduos similares aos que ocorrem na evolução biológica; os indivíduos aqui são segmentos sonoros. O processo de reprodução é feito com os operadores genéticos *crossover* e *mutação*. O processo de seleção utiliza uma medida de adequação do indivíduo: sua distância de Hausdorff ao conjunto alvo de indivíduos que condiciona a evolução. O indivíduo com menor distância de Hausdorff é escolhido para participar do processo de reprodução com os outros indivíduos na próxima geração. Ele será o resultado sonoro da síntese evolutiva na próxima geração. Um segundo modelo de síntese é apresentado, onde características psicoacústicas são extraídas de cada indivíduo, compondo o que denominamos de seu genótipo sonoro. Neste modelo os processos de reprodução e seleção atuam sobre o genótipo.

Abstract

The evolutionary sound synthesis method is presented. It is based on waveforms, gathered in a set called population. Its evolution is conditioned by a target set of waveforms. Two independent processes compound evolutionary synthesis: reproduction and selection. Reproduction is effected by two genetic operators: crossover and mutation. Selection is accomplished by waveform fitness evaluation, using the Hausdorff distance. The waveform with the smallest Hausdorff distance to the target set is chosen to spread its characteristics with all waveforms; it is the best waveform of its generation and therefore the resultant synthesized sound. The method also allows manipulating the psychoacoustics of the waveform, which are its genotype. This method creates a new non-deterministic field of sound synthesis techniques.

Agradecimentos:

Ao Prof. Dr. Furio Damiani, pelo apoio e orientação deste trabalho.

Ao Prof. Dr. Jônatas Manzolli, pelo apoio, inspiração inicial, e co-orientação deste trabalho.

Ao Prof. Dr. Adolfo Maia Júnior, pelo apoio e ajuda na formulação matemática.

Ao Prof. Dr. Peter Jürgen Tatsch, , pela participação na banca examinadora e pelos comentários para a melhoria do trabalho.

Ao Prof. Dr. Florivaldo Menezes Filho, pela participação na banca examinadora e pelos comentários para a melhoria do trabalho.

Ao Prof. Dr. Raul Thomaz Oliveira do Valle, pela participação na banca examinadora e apoio e apoio durante a preparação deste trabalho.

Ao Prof. Dr. Fernando Von Zuben, pelos comentários e sugestões para a melhorias deste trabalho.

Ao NICS – Núcleo Interdisciplinar de Comunicações Sonoras, pelo apoio institucional ao longo do desenvolvimento deste trabalho.

Ao CNPq, como órgão financiador da parte inicial da pesquisa. Bolsa Sanduíche durante o período de 05/02/1996 a 30/06/1997.

Índice Geral

Índice de figuras

1 Introdução

A maioria dos sons que escutamos em nosso dia-a-dia possui uma grande quantidade e variedade de informação sonora. Estes sons, também chamados de sons concretos, são aqueles que nos cercam e que nos trazem, a todo instante, informação [Schaeffer, 66]. O sistema auditivo humano é capaz de perceber e discriminar uma grande quantidade de padrões sonoros que compõem os sons. Chamamos de padrões sonoros às entidades informacionais que mesmo coexistindo em um estimulo sonoro podem ser percebidas separadamente pela percepção auditiva. A teoria da complexidade ótima de Daniel Berlyne afirma que a sensação de satisfação provocada por um estímulo está diretamente relacionada à um nível ótimo de informação nova contida nesse estímulo [Berlyne, 66]. Se o estímulo possui uma quantidade de informação menor ou maior que o ótimo, então este não será tão agradável à percepção. Do mesmo modo, dizemos que sons apresentando um determinado nível ótimo de padrões sonoros é agradável à percepção auditiva humana.

Muitos métodos de síntese sonora têm como objetivo proporcionar um ambiente para a geração conceitual de sons, ou seja, a criação de sons que se aproximem ao máximo de um objetivo estético almejado pelo usuário. Entendemos aqui que síntese sonora é um processo de geração de sons guiados por um objetivo estético. Desse modo, um bom método de síntese sonora deve ser controlável e propiciar a geração de sons que possuam riqueza de padrões sonoros do mesmo modo daqueles sons que escutamos em nosso cotidiano. No entanto, os métodos de síntese clássicos tem que lidar com o compromisso entre: 1) a definição de uma relação entre o controle dos padrões sonoros e as correspondentes qualidades estéticas do som sintetizado, e 2) a definição de um método que seja eficiente, que gere e controle sons com grande quantidade de padrões a um baixo custo operacional. Via de regra, é difícil estabelecer um método de síntese que resolva esse dilema. Existem três características que consideramos importantes constarem em um bom método de síntese sonora: 1) apresentar independência de controle dos padrões sonoros. 2) gerar sons com grande riqueza de padrões 3) estabelecer uma correspondência fidedigna entre os controles e os resultados perceptuais da síntese sonora. No entanto, criar uma síntese que preencha estas três características não é trivial. De fato, pelo nosso conhecimento ainda não existe um método de síntese que englobe essas três características.

A vantagem de sínteses com independência de controle dos padrões sonoros é que o seu aprendizado é otimizado pelo fato de sua manipulação ser intuitiva. Isto torna possível chegar a resultados sonoros esperados manipulando os controles da síntese simplesmente por tentativa e erro. Sínteses como a aditiva ou subtrativa possuem essa característica, porém tornam-se computacionalmente caras para gerar sons com grande quantidade de padrões. Do mesmo modo, sínteses não-lineares como a FM, geram facilmente sons ricos em padrões, porém sem que os mesmos possam ter os seus padrões sonoros controlados independentemente. Além disso, o resultado sonoro dessas sínteses é determinístico, ou seja, a variação do resultado da síntese é causa única e exclusiva da variação de um ou mais parâmetros que a controlam. Uma nova possibilidade seria, por exemplo, a adequação automática dos parâmetros de controle da síntese sonora de acordo com as premissas dadas pelo usuário.

Torna-se portanto interessante a criação de um método de síntese sonora independentemente controlável, que permita a geração de sons ricos em parâmetros e cujo controle não seja necessariamente causal. Uma síntese sonora deste tipo, cujo som de saída evolua ao longo do tempo para melhor se adequar aos resultados sonoros esperados pelo usuário, é o objetivo do trabalho aqui apresentado. Chamamos este método de síntese evolutiva de segmentos sonoros.

1.1 A criação de uma síntese evolutiva de segmentos sonoros

A síntese evolutiva de segmentos sonoros baseia-se na computação evolutiva, e como tal pode ser entendida como um método de aprendizado não supervisionado pelo usuário que gera e busca a melhor solução possível de um problema genérico, no caso, o melhor som sintetizado. A computação evolutiva baseia-se nos processos de seleção e reprodução da evolução biológica. Enquanto novas soluções são dinamicamente geradas, a melhor delas é selecionada através da

medida de sua adequação a um conjunto de critérios, fator condicionante da evolução na população de soluções possíveis.

A síntese evolutiva utiliza os princípios da computação evolutiva para a geração de som. As soluções possíveis são chamadas de *indivíduos*, que são *segmentos sonoros*. Os indivíduos fazem parte de um conjunto chamado conjunto *população*. O som sintetizado é o *melhor indivíduo* do conjunto população e os critérios de escolha do melhor indivíduo são dados pela medida da adequação (ou distância) de cada indivíduo da população, comparados a um conjunto de características esperadas pelo usuário. Estas características estão representadas sob a forma de outros indivíduos contidos num conjunto chamado *conjunto alvo*. A cada geração do conjunto população, novos indivíduos são criados pela reprodução dos indivíduos existentes na população. A seleção destes novos indivíduos é feita pela medida de sua adequação em comparação aos indivíduos do conjunto alvo. Uma vez que o conjunto alvo determina as características esperadas no som sintetizado, este passa a condicionar o rumo da evolução da população de sons ao longo de suas sucessivas gerações.

O controle de padrões da síntese evolutiva é feito através da manipulação dos indivíduos no conjunto alvo. Os indivíduos deste conjunto podem ser modificados independentemente e assim influenciar independentemente os critérios de condicionamento da seleção. Isto faz com que a síntese evolutiva tenha independência de controle e seu aprendizado seja intuitivo. Os seus resultados sonoros são gerados a partir da reprodução e seleção dos indivíduos da população. Uma vez que estes sejam amostras de sons quaisquer, o som sintetizado será descendente destes sons que, em geral, são ricos em padrões sonoros.

A síntese evolutiva pretende ser um método de síntese sonora que reúna pela primeira vez as três mais importantes características esperadas de um método de síntese sonora: a *independência de controle*, *aprendizado intuitivo* e *geração de sons ricos em padrões*.

1.2 A evolução histórica das sínteses sonoras tradicionais

De um modo geral, pode-se dizer que a síntese de sons é tão antiga quanto a própria humanidade. Sintetizar um som equivale a desenvolver um método para criar um novo som, ou imitar um som conhecido, porém, sempre com uma intenção estética preestabelecida. Assume-se que desde os primórdios da civilização o ser humano tem desenvolvido artefatos e técnicas corporais (como assobios ou bater palmas) no intuito de imitar ou criar novos sons. Para nós, o que distingue a geração de som da síntese sonora é que na síntese existe um objetivo perceptual almejado.

O som é para a humanidade um meio de aquisição e transmissão de conceitos. Conceitos podem ser definidos como módulos de informação interpretados, associados e armazenados pela mente humana. Os conceitos compõem a noção de realidade para o indivíduo. Ao perceber um som, o cérebro desenvolve associações com outros conceitos previamente conhecidos, armazenados em sua memória. Estas associações são feitas e refeitas dinamicamente. Desde o instante em que escutamos um som pela primeira vez, até quando este já nos é bem conhecido, novas associações deste som a outros conceitos são elaboradas. Conceitos determinam emoções que determinam ações. Muitos estudos no ramo das ciências cognitivas têm sido desenvolvidos sobre a relação entre conceitos e emoções, como em [Keltner,99] e [Lane,02].

Sons, de um modo geral, desencadeiam no ouvinte emoções relacionadas ao estímulo sonoro. Talvez por esta razão o ser humano sempre se interessou em imitar sons conhecidos, uma vez que o som imitado pode evocar no ouvinte os mesmos conceitos relacionados ao som original. O desenvolvimento natural da associação de conceitos à sons naturais deu origem às linguagens faladas. Através das muitas linguagens desenvolvidas pelos grupos humanos, tornou-se possível a transmissão de conceitos específicos, complexos e até mesmo intangíveis, como a noção de amor ou ódio, ainda que com diferenças de interpretação, dadas as diferenças cognitivas entres os conceitos transmitidos. Estabelece-se assim a diferença entre a cognição epistemológica e a cognição prática. A primeira trata daquilo que acreditamos e a segunda daquilo que devemos fazer [Pollock,97].

Um dos objetivos básicos da linguagem é a comunicação eficiente de conceitos associados a ações específicas. No entanto, antes do conceito determinar uma ação, este determina uma emoção. Pode-se dizer que o objetivo da arte, como meio de comunicação humana, é a transmissão de conceitos que determinem prioritariamente emoções. Por este fato, dos sistemas de comunicações sonoros, a música é o que mais diretamente comunica emoções e sentimentos humanos. Ao contrário

da linguagem em prosa, desenvolvida para comunicar conceitos objetivos, a música, como a poesia, procura transmitir aquilo que é subjetivo, embora a música seja, como linguagem artística, ainda mais universal e abstrata.

O som, enquanto meio de comunicação, apresenta dois níveis informacionais, separados entre si pelo limite de percepção de simultaneidade de eventos sonoros. São os níveis microscópico e macroscópico do som. Quando dois eventos sonoros ocorrem dentro de um intervalo de tempo tão pequeno que a percepção auditiva os interpreta como simultâneos, estes são eventos que pertencem ao nível informacional chamado de microscópico. Já, quando a percepção auditiva percebe a não-simultaneidade dos eventos sonoros, então estes passam a fazer parte do nível informacional chamado de macroscópico. Nota-se uma tendência natural ao longo da história da evolução social humana em afastar da música a manipulação do nível microscópico do som, atribuindo a este uma categoria mais técnica e menos artística, normalmente deixada ao encargo de fabricantes de aparelhos sonoros ou instrumentos musicais. Enquanto isso, a manipulação do nível informacional macroscópico vem sendo ostensivamente utilizada na composição musical de melodias, canções e ritmos.

Com o desenvolvimento da tecnologia eletrônica a partir da segunda metade do século XX alguns compositores musicais de vanguarda passaram a se interessar pela possibilidade da manipulação microscópica do som para fins de composição musical. A nova tecnologia eletrônica permitia representar o som como um sinal de áudio que podia ser armazenado em fitas magnéticas, manipulado por filtros ou mesmo gerado por osciladores. São citados como compositores precursores do uso dessa tecnologia, entre outros, Theremin, Martenot e John Cage, que em 1939 fez uso pioneiro desses meios na sua composição *Imaginary Landscapes*.

O primeiro compositor citado por divulgar, em sessões de radio difusão, a experimentação de recursos eletrônicos na música é *Pierre Schaeffer*, em Paris, no final dos anos 40. *Schaeffer* atribuiu o nome de *Musique Concrète* a este novo estilo musical. Na mesma época, outro grupo utilizando recursos eletrônicos se organizava na Alemanha sob o nome de *Elektronische Musik*, sob a liderança de *Herbert Eimert*. Apesar das diferenças e rivalidades entre os dois grupos, ambos contribuíram para o estilo musical posteriormente chamado de Música Eletroacústica [Bennet,02].

Com o advento da tecnologia digital e dos computadores, as possibilidades de manipulação de sons aumentaram ainda mais. A partir da década dos 70s, os computadores passaram a ter maior capacidade de memória e processamento. O som pôde então ser amostrado no tempo e armazenado em forma digital, como uma seqüência finita de números inteiros. A manipulação do som digital passou a ser muito mais fácil, uma vez que era feita através de algoritmos computacionais. Um dos primeiros trabalhos musicais com som manipulado por computador foi feito no centro de computação musical da *Stanford University*, CCRMA, por *Michael McNabb*, em sua peça entitulada *Dreamsong*.

O advento da tecnologia digital tornou possível a representação do som praticamente sem perdas perceptíveis para a audição humana. Programas de *software* editores de áudio são capazes de manipular o som de todas as formas imagináveis. Como o som digital é representado por uma seqüência de números inteiros, teoricamente é possível criar qualquer tipo de som que se queira pela criação de uma nova seqüência de números. No entanto, não existe uma correspondência intuitiva entre os números da seqüência que compõe o som digital e o som como é percebido pela audição e reconhecido pelo cérebro. O obstáculo para se criar uma síntese controlável e intuitiva deixou de ser técnico e passou a ser cognitivo [Chowning, 2000].

O ramo da engenharia elétrica que estuda os fenômenos sonoros é a engenharia de som. O som é representado como sinal de áudio, que pode ser contínuo (analógico) ou discreto (digital) no domínio do tempo. O estudo do som representado como sinal de áudio discreto no tempo faz parte do processamento digital de sinais, ou DSP (*digital signal processing*) mais especificamente chamado de ADSP (*audio digital signal processing*), conforme visto na figura 1.1. Pode-se dividir o ADSP em três ramos de técnicas: análise, transformação e síntese do som. Estas estão dispostas seqüencialmente uma vez que, na manipulação sonora, a análise precede a transformação que, por sua vez, precede a síntese sonora [Serra ,89]. As técnicas de análise tratam da investigação do material sonoro; da sua constituição em componentes e das suas características principais e específicas. As técnicas de transformação tratam da manipulação ou processamento sonoro bem como aquelas que são interessantes para a percepção auditiva. As técnicas de síntese tratam da geração do material sonoro, a partir de outros tipos de sinais, funções matemáticas ou a partir de outro material sonoro. A figura a seguir mostra alguns exemplos de técnicas conhecidas dos três ramos do ADSP.

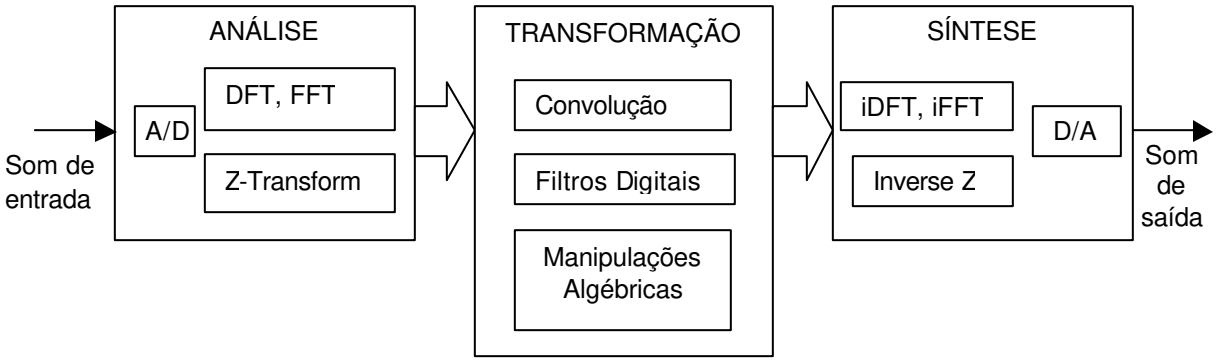


Figura 1.1 Diagrama com alguns exemplos dos três ramos do processamento digital de áudio, ADSP

Muitos métodos de síntese sonora digital foram desenvolvidos a partir dos avanços tecnológicos do ADSP. Citamos a seguir alguns que foram particularmente relevantes. Mais detalhes são encontrados no apêndice deste trabalho. Os primeiros métodos de síntese se basearam no princípio da teoria de Fourier, onde qualquer sinal periódico no domínio do tempo pode ser criado a partir da somatória de funções ortogonais, como senos e cosenos, que são a representação das mais simples componentes sonoras. Estas sínteses foram posteriormente catalogadas como pertencentes aos modelamentos espectrais do som.

O primeiro método desenvolvido, a síntese aditiva, é um método de síntese linear onde formas de onda geradas por osciladores elétricos são somados e o resultado é levado a uma saída de áudio e transformado em som. Os parâmetros da síntese aditiva são os controles dos parâmetros de cada oscilador, como: intensidade, freqüência e fase. A princípio a síntese aditiva possibilita a criação de qualquer som pela somatória de senóides. No entanto este método torna-se cada vez mais complexo computacionalmente na medida em que o som sintetizado torna-se mais elaborado, o que exige um maior número de osciladores. O resultado é que a síntese aditiva torna-se inviável para a geração e controle de sons próximos aos sons que escutamos em nosso dia-a-dia.

Já a síntese subtrativa se baseia no princípio oposto, ou seja, a partir uma fonte sonora complexa, como um ruído, as componentes em excesso são subtraídas por filtragem. O resultado sonoro é então a simplificação do som original. Esta síntese assemelha-se à maneira natural de geração de som em um instrumento acústico, como um saxofone, onde a palheta gera por vibração um amplo espectro de componentes sonoros e estas são filtradas pelo formato do corpo do instrumento. Apesar de gerar sons com grande riqueza de padrões, não existe muito controle para a modificação do som gerado que permanece restrito a um mesmo padrão timbrístico.

A síntese granular é um método linear de síntese sonora que se baseia no conceito de grão sonoro. Um grão é uma unidade sonora com duração de alguns milésimos de segundo, mas que pode ser bastante rica em componentes sonoras. Os grãos são armazenados em uma tabela e são acessados de diversas formas, repetindo-se ciclicamente ou misturados a outros grãos. Esta técnica de síntese gera sons elaborados e também permite a manipulação independente dos parâmetros sonoros pela modificação dos grãos que compõem o som sintetizado. No entanto, não existe a geração de um novo material sonoro, mas apenas aqueles que já estão representados pelos grãos, o que faz desta técnica, mais do que um processo de síntese sonora, um processo elaborado de edição ou colagem de sons.

O método não-linear de síntese sonora mais conhecido é o método por modulação de freqüência, ou síntese FM. O princípio deste método é que osciladores são modulados em freqüência por outros osciladores [Chowning, 73]. Criaram-se assim algoritmos com estruturas de conexão de osciladores modulando a freqüência e a fase de outros osciladores. Desta forma cada algoritmo foi mapeado a um tipo particular de resultado timbrístico. A síntese FM permite criar sons com grande riqueza timbrística a partir de poucos osciladores. Porém a relação entre a estrutura de cada algoritmo e seu timbre característico não é intuitiva, o que faz com que a síntese FM não tenha uma grande previsibilidade do som resultante.

Outro método conhecido de síntese não-linear é a síntese *waveshaping*. Como o nome sugere, um oscilador tem o espectro de freqüência do seu sinal de áudio alterado por uma função não-linear gerando assim novos componentes em freqüência que não existiam no som original.

Pode-se comparar este método à distorção que um amplificador a válvula insere no som de um instrumento musical ou voz, enriquecendo este som com novas componentes de freqüência. O resultado, via de regra, é um som perceptualmente mais interessante. No entanto, como todo método não-linear, o controle da síntese *waveshaping* não é intuitivo, ou seja, é difícil estabelecer métodos de parametrização de controles que cheguem a resultados sonoros previsíveis.

Todos os métodos de síntese descritos até aqui, sejam eles lineares ou não, se baseiam na criação de um som através do manipulação de seu espectro de freqüência. Outra família de síntese sonora usa o modelamento físico de uma fonte sonora, um instrumento musical virtual representado por equações dinâmicas no tempo. Esta abordagem é chamada de *physical modeling*. Pelo fato deste método criar um instrumento virtual, esta síntese tem grande possibilidade de controle de parâmetros. No entanto a modelagem de todo o processo físico de criação de um som por um instrumento musical é geralmente muito complexo e esta síntese torna-se computacionalmente muito complexa para emular sons de instrumentos acústicos.

Uma maneira computacionalmente barata de simular sons conhecidos é dada pela síntese *wavetable*. Esta é utilizada ostensivamente pela industria de multimídia. A síntese *wavetable* não se baseia no modelamento físico ou espectral do som. Ao invés, esta se baseia na amostragem do som. Nela, uma tabela armazena segmentos de sons originais, na sua representação digital. Estes segmentos são manipulados quanto ao *loudness* e *pitch* (ver definição no apêndice deste trabalho) para representar este som em uma dada intensidade sonora e nota da escala musical. Apesar do controle ser bastante intuitivo esta síntese não permite a criação de novos sons mas apenas a manipulação de parâmetros de controle dos segmentos sonoros previamente amostrados.

Observa-se que todos os métodos tradicionais de síntese sonora descritos acima tem uma característica em comum, todos são determinísticos. Os métodos de síntese sonora apresentam um som sintetizado de saída que é único, se a entrada e os parâmetros estão fixos. A saída varia somente se a entrada e os parâmetros variam. A síntese evolutiva introduz um novo ramo da síntese sonora, pois é o primeiro método de síntese que não é determinístico, mas sim evolutivo.

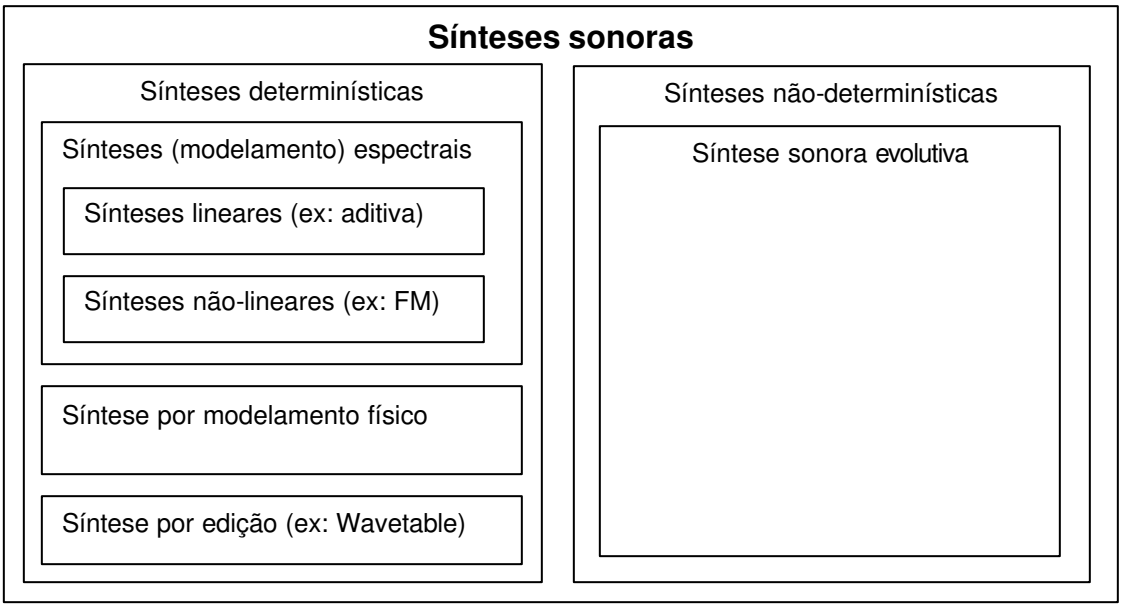


Figura 1.2 Diagrama de classificação das sínteses evolutivas.

1.3 A síntese evolutiva de segmentos sonoros

Conforme vimos, nenhum dos métodos de síntese descritos no item anterior, catalogados por modelagens espectrais ou físicos, lineares ou não-lineares, reúne ao mesmo tempo as mais importantes características esperadas de uma síntese sonora: controle intuitivo, geração de sons com riqueza de padrões similar aos sons escutados na natureza e baixo custo computacional. É sem dúvida um grande desafio criar um método de síntese sonora que reúna todas essas características, especialmente no que toca o compromisso entre a independência de controle e a riqueza sonora. Enquanto algumas sínteses são bastante controláveis e pouco ricas, outras são muito ricas mas pouco controláveis.

Outros métodos envolvendo síntese sonora e computação evolutiva foram propostos tais como [Masri,98], que introduz uma técnica de síntese sonora similar a *wavetable*, porém com um grande mapa de segmentos sonoros que são controlados por uma rede neural, e [Garcia,00] que apresenta um método de geração automática de modelos de síntese sonora através de algoritmos genéticos.

A síntese evolutiva, no entanto, é um método de síntese sonora que tem o objetivo de englobar as diversas vantagens das sínteses clássicas em uma síntese controlável que produza resultados sonoros com riqueza de padrões. A computação evolutiva é adequeada para este fim pois é um método de geração e busca de soluções para problemas não determinísticos, como é o caso da síntese sonora, onde se gera um novo som perceptualmente interessante a partir da manipulação de um conjunto de segmentos sonoros.

A computação evolutiva é um método desenvolvido para tratar com este tipo de problemática, ou seja, a partir da geração de uma série de soluções possíveis, procurar a que melhor se adapta para a solução do problema. Como tantos outros métodos ou invenções humanas, a computação evolutiva também se baseia na imitação da forma como a natureza resolve problemas, através de um método de busca paralelo-seqüencial, pois utiliza estratégias exploratórias por amostragem e comparação (em paralelo) e por evolução (seqüencial). A computação evolutiva surgiu em meados dos anos 50s, com as primeiras abordagens evolutivas para resolução de problemas da engenharia [Fraser,59]. Estes procuraram pela primeira vez emular os processos naturais de reprodução e seleção biológica. A reprodução é dada por **operadores genéticos**, algoritmos que simulam a ação dos processos naturais de *crossover* e mutação na divisão celular [Mrazek,96]. Pode-se dizer que a computação evolutiva é um método de busca de uma solução genérica para um problema específico, cujo candidato à solução faz parte de um conjunto de soluções possíveis, chamado de **população**. Os indivíduos da população estão sujeitos ao condicionamento de sua evolução pela adaptação a critérios que podem variar dinamicamente. Estes critérios são chamados de meio condicionante, ou **conjunto alvo**. Enquanto indivíduos são criados por reprodução, o indivíduo que mais satisfaz os critérios do conjunto alvo é escolhido como o mais adaptado ao meio e a melhor solução para o problema. A computação evolutiva faz parte do ramo da inteligência computacional, junto à lógica *fuzzy*, redes neurais e os agentes autônomos. É oportuno enfatizar aqui que um processo evolutivo arbitrário pode fazer uso de outros operadores que não o crossover e a mutação, os quais emulam processos biológicos. Claramente, existe na natureza uma enorme variedade de processos evolutivos não-biológicos tais como evolução estelar, crescimento de cristais, fenômenos meteorológicos, etc. Tais operadores poderiam ser emulados e incorporados no método de síntese evolutiva. Neste trabalho, no entanto, nos restringiremos aos operadores genéticos clássicos.

Encontram-se na bibliografia de música computacional vários exemplos de algoritmos evolutivos para a manipulação e geração de material sonoro. Um algoritmo evolutivo que simula um estudante aprendendo a improvisar solos de *jazz* sob a supervisão de um professor (o usuário) foi apresentado em [Biles,94]. Outro sistema evolutivo faz a distinção entre padrões rítmicos, gerando um grande número de variações interessantes [Horowitz,94]. Algoritmos evolutivos vem sendo utilizados nas artes visuais, em particular na computação gráfica, para a criação de cenas de animação [Foley,96]. Nestas aplicações o sistema aprende as regras pela interação com o usuário [Fogel,95].

Os algoritmos evolutivos se baseiam no processo da evolução natural e podem ser vistos como um procedimento computacional de aprendizado autônomo, não supervisionado pelo usuário. Algoritmos genéticos são um procedimento para a solução de um problema, independente do seu domínio, no qual o programa evolui para a melhor solução possível [Koza,97]. A definição do conceito

de algoritmo genético pode ser encontrada em diversos trabalhos da literatura de computação evolutiva, como em [Horowitz 94] e [Koza 97].

A motivação inicial para a criação de um método de síntese evolutiva veio do software Vox Populi, um sistema evolutivo para geração de seqüências musicais através de operadores genéticos. Pode-se definir o Vox Populi como um sistema híbrido entre um instrumento musical e um ambiente de composição musical [Moroni,00]. Desenvolvido no NICS, o Núcleo Interdisciplinar de Comunicação Sonora, da UNICAMP, este sistema utiliza seqüências de notas musicais como indivíduos da população que são manipulados interativamente por operadores genéticos. A melhor seqüência de notas é selecionada pelo sistema através da medida da distância de Hausdorff, entre cada indivíduo da população e os indivíduos de um conjunto de seqüências dadas pelo usuário [Manzolli,99].

A partir daí, consideramos como passo seguinte o desenvolvimento de um sistema para síntese sonora utilizando os mesmos processos de computação evolutiva que o Vox Populi utiliza para a geração de seqüências de controle MIDI. O novo sistema, chamado de síntese evolutiva, pode também ser visto como um ambiente de composição de timbres sonoros. O processo de percepção de timbres, no entanto, não é trivial. Propondo uma taxonomia, Schaeffer introduziu a idéia de classificação timbrística, distinguindo sons entre forma e matéria, no contexto da música concreta [Schaeffer,66]. Posteriormente, Risset associou o conceito de *forma* à curva de variação da percepção da amplitude, ou curva de *loudness*, e *matéria* à magnitude do espectro de freqüência do som [Risset,91]. Além da curva de loudness e espectro, um outro parâmetro que consideramos importante para a caracterização perceptual do som é seu harmônico fundamental que, em termos musicais, refere-se à altura, ou *pitch* do som. Iremos assim utilizar as funções da variação no domínio do tempo do *pitch* e do *loudness* bem como a distribuição do espectro de freqüência de um segmento sonoro como critério de adequação da síntese evolutiva. Estas são chamadas de curvas psicoacústicas. A definição formal de grandezas psicoacústicas tais como *loudness* e *pitch* podem ser encontradas em diversas referências, como em [Zwicker,98].

A síntese evolutiva de sons pode ser vista como um processo interativo que coordena uma série de regras para a geração e busca de segmentos sonoros através de operadores genéticos, com base em um modelo que mede a similaridade psicoacústica entre estes segmentos sonoros e aqueles em um conjunto alvo, contendo segmentos fornecido pelo usuário. Em nossa abordagem, segmentos sonoros são elementos, ou indivíduos, de conjuntos os quais denominamos *população* que é atualizada dinamicamente pelos operadores genéticos. Também definimos o conjunto de segmentos sonoros o qual denominamos de conjunto *alvo*. Este conjunto é fornecido pelo usuário e condiciona a geração de novos segmentos da população atual.

1.4 Organização da tese

Este trabalho tem cinco capítulos. Neste primeiro capítulo tratamos da problemática da síntese sonora e a motivação para a criação de um novo método que seja *controlável, elaborado, e evolutivo*.

No segundo capítulo detalhamos o primeiro modelo matemático da síntese evolutiva e apresentamos seus algoritmos genéticos *crossover* e *mutação* e o critério de seleção do resultado sonoro. Neste primeiro modelo os processos da síntese evolutiva ocorrem diretamente sobre o segmento sonoro.

No terceiro capítulo descrevemos o segundo modelo da síntese evolutiva, onde definimos o genótipo do indivíduo, formado pelas curvas psicoacústicas de *loudness*, *pitch* e espectro. Neste modelo as curvas psicoacústicas do genótipo são utilizadas pelos processos de reprodução e seleção, para a geração do som sintetizado.

O quarto capítulo é a parte experimental deste trabalho. Nele são discriminados todos os algoritmos desenvolvidos para a simulação de cada etapa dos dois modelos de síntese evolutiva. Estes algoritmos estão dados na forma de funções do software de simulação MATLAB.

No quinto capítulo apresentamos nossas conclusões e comentários. Fazemos uma abordagem crítica dos métodos utilizados no quarto capítulo e apresentamos algumas sugestões para o desenvolvimento de futuras pesquisas em síntese evolutiva.

No apêndice há material para o embasamento técnico necessário ao entendimento do estudo do som, realçando seus aspectos objetivos, ou acústicos e subjetivos, ou perceptuais. É também apresentada uma breve história comparativa dos métodos de síntese sonora no intuito de situar a síntese evolutiva entre os outros métodos de síntese sonora e desse modo realçar a razão para o seu desenvolvimento.

2 Descrição do método da síntese evolutiva

No capítulo anterior vimos que o método da síntese evolutiva de segmentos sonoros está baseado em dois processos independentes: reprodução e seleção. A reprodução é dada pela ação dos operadores genéticos *crossover* e mutação. A seleção é dada pela ação de uma função de adequação que mede a distância entre o indivíduo e o conjunto alvo.

Aqui neste capítulo formalizaremos o primeiro modelo da síntese evolutiva, que é também a base para o segundo modelo, descrito no próximo capítulo. O primeiro modelo de síntese evolutiva é baseado na ação dos processos de reprodução e seleção diretamente sobre o segmento sonoro.

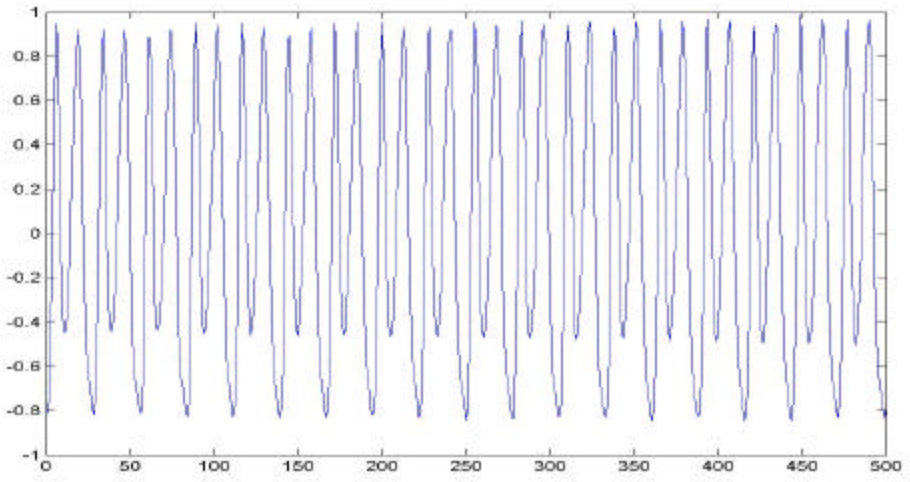
2.1 Introdução

Tem-se visto na literatura de música eletrônica que a computação evolutiva vem sendo utilizada para a geração de técnicas de propósitos musicais. [Homer,93] desenvolveu um método para a automatização da geração de parâmetros da síntese FM usando a computação evolutiva. [Garcia,00] utilizou algoritmos que simulam operações genéticas, GAs, na automação de projetos de novas técnicas de síntese sonora. [Manzolli,99] tem trabalhado no desenvolvimento de programas para composição interativa onde GAs manipulam seqüências MIDI para a automação da composição musical. [Johnson,99] desenvolveu um sistema envolvendo computação evolutiva (máquina) e ouvintes (seres humanos) para a exploração de novos parâmetros para a síntese sonora. [Roads,94] usou GAs na parametrização da síntese granular.

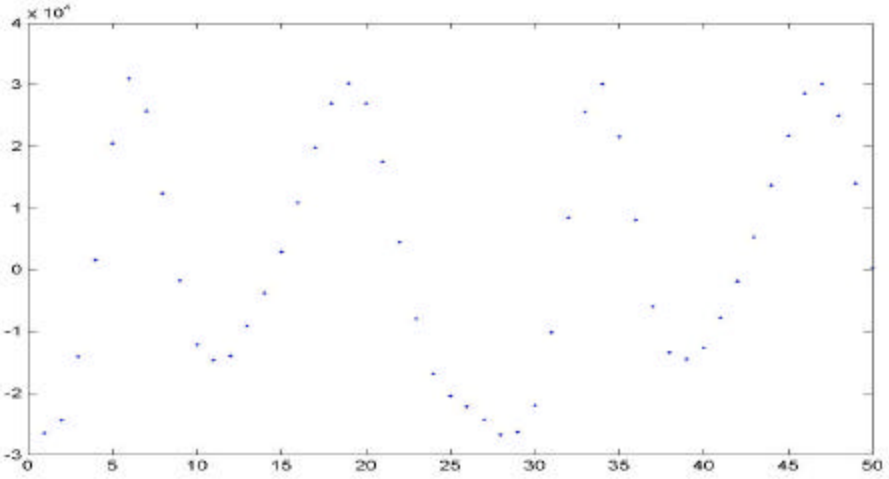
A síntese evolutiva, que é apresentada neste capítulo, é um método que utiliza ferramentas da computação evolutiva, como operadores genéticos e funções de adequação, para a geração de segmentos sonoros. Esta se baseia na evolução biológica das espécies, composta por dois processos básicos: a seleção e a reprodução, conforme descrita para teoria darwiniana.

Na síntese evolutiva os segmentos de som são tratados como indivíduos pertencentes a uma população, o conjunto população. A criação de novos indivíduos se dá pela aplicação dos operadores genéticos: *crossover* e mutação, sobre os indivíduos da população. A seleção do melhor indivíduo ocorre através da aplicação de funções de adequação que medem sua similaridade com os indivíduos do conjunto alvo, os quais possuem características sonoras importantes para o usuário.

Cada indivíduo da população é um segmento de som digitalizado, dado por uma seqüência finita de números inteiros, que representam a amostra de um segmento sonoro, numa determinada taxa de amostragem e resolução.



(a)



(b)

Figura 2.1. Exemplo de indivíduo sonoro. (a) Segmento completo, (b) detalhe mostrando que o indivíduo é composto por uma seqüência finita e discreta de números inteiros.

O conjunto alvo da síntese evolutiva contêm indivíduos com as características sonoras selecionadas pelo usuário, que condicionam o processo evolutivo. O indivíduo com características sonoras mais próximas dos indivíduos do conjunto alvo é o som sintetizado. Este é determinado por uma função de adequação, dada aqui pela distância de Hausdorff entre cada indivíduo da população e o conjunto alvo.

Sob o ponto de vista algorítmico, a síntese evolutiva é um sistema de auto-aprendizado interativo, mas que é guiada pela interferência do usuário através da inclusão e exclusão de indivíduos no conjunto alvo que assim direciona o processo evolutivo desta síntese.

2.2 A manipulação evolutiva dos segmentos sonoros

A representação matemática de timbre vem sendo um dos problemas mais desafiadores na música computacional. È uma tarefa bastante complexa estabelecer uma taxonomia que classifique timbres sob ambos os aspectos qualitativo e quantitativo. [Schaeffer,66] introduziu através do conceito de música concreta a distinção entre forma e matéria. Conforme é explicado por [Risset,91], o conceito de Schaeffer estabelece como *forma* o envelope de amplitude do som e *matéria* o conteúdo do espectro em freqüência. Esta talvez tenha sido a primeira tentativa de se descrever a natureza timbrística do som. Hoje em dia sabe-se que o espectro em freqüência do som varia dinamicamente ao longo do tempo, assim este não pode ser adequadamente definido por um conceito estático como o de matéria. As mudanças dinâmicas do espectro sonoro carregam em si importante informação sonora. [Smalley,90] declarou que a informação sonora expressa no espectro de freqüência não pode ser separada do domínio do tempo uma vez que o espectro é percebido ao longo do tempo do mesmo modo que o tempo, na percepção sonora, é percebido pela mudança do espectro. [Risset,91] declarou que variantes sonoras produzidas por mudanças nos parâmetros de controle da síntese são intrigantes no sentido em que não costuma existir uma relação intuitiva entre controle e mudança sonora. Pequenas mudanças paramétricas podem gerar grandes mudanças sonoras, e vice-versa. Partindo desses dois princípios, imaginaram-se variantes sonoras produzidas por GAs numa população de segmentos sonoros ao longo do tempo (isto é, curvas no espaço cartesiano bidimensional *amplitude x tempo*), criando assim uma evolução dinâmica do timbre.

Em analogia com a genética, em computação evolutiva também pode-se associar a um indivíduo o seu respectivo genótipo. A definição propriamente dita de genótipo tem um caráter arbitrário, isto é, existem diversas possibilidades para a escolha do genótipo do som. Em geral o segmento sonoro, como uma forma de onda discretizada no tempo, é considerado um elemento primário. No primeiro modelo do método de síntese evolutiva vamos considerar genótipos como os próprios segmentos sonoros. Esta identificação é possível porque formas de ondas são passíveis, como veremos, de manipulações por operadores genéticos. No capítulo 3 veremos uma escolha diferente para a descrição de genótipos da que foi apresentada aqui.

Do mesmo modo que Risset interpretou os conceitos de Schaeffer, podemos relacionar o segmento de som, ou genótipo, com a forma e o seu timbre, ou o fenótipo correspondente, com a matéria. Neste contexto, esses dois elementos são usados pela Síntese Evolutiva de modo similar ao processo biológico da evolução que usa a informação genética para gerar novos indivíduos.

A síntese evolutiva é aqui definida como um método de interação entre algoritmo e usuário. Inicialmente o usuário especifica os indivíduos pertencentes ao conjunto alvo. A seguir o algoritmo cria novos indivíduos tendo o conjunto alvo como critério. O usuário pode então mudar os indivíduos do conjunto alvo a qualquer momento que deseje. Quando isto acontece, um novo rumo para a evolução da população é estabelecido.

Existem três estruturas principais na síntese evolutiva.

- $B^{(n)}$, a n-ésima geração do conjunto população. O conjunto população inicial, antes da primeira geração, é denotado por $B^{(0)}$.
- T, o conjunto alvo.
- f, a função de adequação (*fitness*), usada para avaliar o melhor indivíduo da n-ésima geração, denotado por: w_*^n.

Define-se como melhor indivíduo w_*^n aquele indivíduo pertencente a $B^{(n)}$ mais próximo de T. A cada nova geração pode-se ter um novo w_*^n que é enviado para a saída do sistema como o som sintetizado.

Inicialmente utilizamos como indivíduos segmentos de som fixos em 1024 pontos. Cada ponto representado por uma palavra binária de 16 bits, que define 2^{16} níveis distintos, num intervalo de [-32768, +32768]. Os indivíduos que são elementos do conjunto T possuem as mesmas características e são dados pelo usuário.

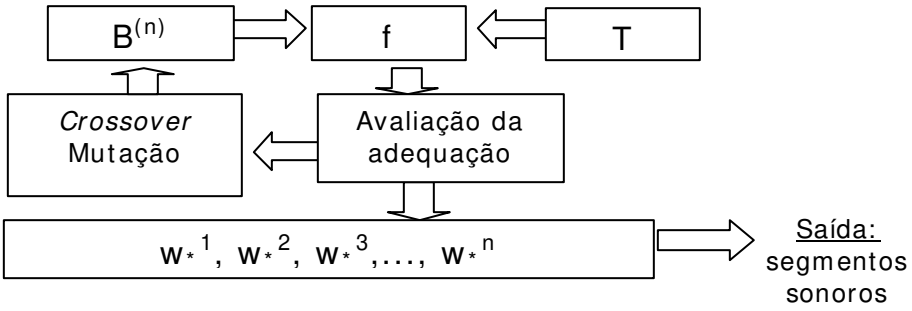


Figura 2.2 Diagrama do primeiro modelo da síntese evolutiva. evolutiva.

2.3 Os operadores genéticos

Os indivíduos da população de segmentos sonoros são manipulados pelos operadores genéticos: *crossover* e mutação durante o processo de reprodução. Estes permitem a interessante característica da síntese evolutiva de gerar dinamicamente uma seqüência de indivíduos ao longo do tempo com características similares entre si porém que evolui, ou converge, para um dado referencial estético. Do mesmo modo que a evolução biológica produz uma grande diversidade de seres na natureza, a síntese evolutiva pode também criar e manipular uma vasta quantidade de segmentos sonoros bastante variados entre si, sob o aspecto cognitivo. O operador *crossover* incrementa a variância do indivíduo pela transmissão de características cognitivas dos predecessores destes indivíduos, enquanto o operador mutação aumenta a variação genotípica da população como um todo. O algoritmo destes dois operadores é dado a seguir:

2.3.1 Operador *Crossover*

Inicia-se escolhendo a taxa do *crossover*, dada por um número real entre $0 \leq \alpha \leq 1$. Esta é a taxa de operação do *crossover*, que é de variação contínua e dinâmica. O melhor indivíduo da n-ésima geração w_*^n é então o progenitor na população $B^{(n)}$, $w_*^n = (a_1, a_2, a_3, a_4, \ldots a_{1024})$. Os outros indivíduos em $B^{(n)}$ são denotados por w_r^n onde $0 \leq r \leq M$ e M é o número de indivíduos na população. A operação de *crossover* na n-ésima geração é definida nos seguintes passos:

1. Tenha um gerador de números aleatórios inteiros pertencentes ao intervalo [1, N].
2. Gere dois números aleatórios $k_1^{(n)}$ e $k_2^{(n)}$, onde $k_1^{(n)} < k_2^{(n)}$.
3. A partir de $k_1^{(n)}$ e $k_2^{(n)}$ selecione um segmento de w_*^n como s_*^n onde $s_*^n = (a_{k1}, \ldots, a_{k2})$.
4. Combine o segmento S_*^n com o correspondente segmento S_r^n em $B^{(n)}$, conforme segue abaixo:

$$s_r^{(n+1)} = \alpha . s_*^n + (1-\alpha) . s_r^n \qquad (2.1)$$

onde:
$0 \leq i \leq M$
$s_r^n = (b_{k1}, \ldots, b_{k2})$

O novo segmento gerado após a operação de *crossover* é: $s_r^{(n+1)} = (b'_{k1}, \ldots, b'_{k2})$

5. A operação de *crossover* substitui cada $S^{(r,n)}$ fazendo com que $W_r^n = (b_1, b_2, .., b'_{k1} .. b'_{k2}, .. b_{1024})$
6. Repetir os passos (4) e (5) para todos os indivíduos $W^{(r,n)}$ em $B^{(n)}$ de modo que $W_i^n \neq W_*^n$

No passo (4) aplica-se a combinação convexa como operação interna de *crossover*, o que é equivalente a misturar canais de som usando um *mixer* de som.

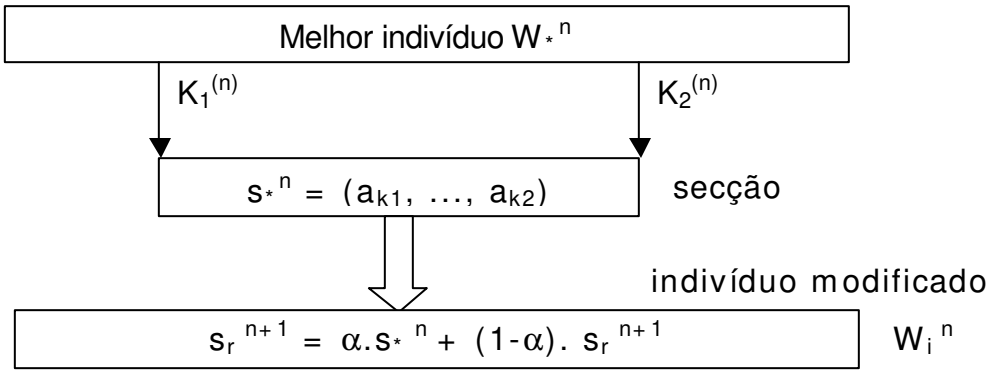


Figura 2.3 Diagrama da operação de *crossover.*

Na biologia, a operação de *crossover* ocorre durante o processo de formação dos gametas, ou células sexuais, pela divisão dos cromossomos por meiose. Neste processo, pares de cromossomos trocam segmentos de material genético entre si, chamados de genes. Analogicamente, a operação de *crossover* na síntese evolutiva troca partes do segmento sonoro de cada indivíduo com cada indivíduo particular, chamado de melhor indivíduo w_*^n. Após a operação de *crossover* cada um dos M indivíduos da população $B^{(n)}$ passa a possuir partes de segmento sonoro do melhor indivíduo. Afim de se obter indivíduos mais bem adaptados, todos os indivíduos da população trocam segmentos com o melhor indivíduo a cada geração da população.

2.3.2 Mutação

Nos organismos vivos a mutação pode ser encarada como um processo que constantemente provoca variações no genótipo da população. A mutação é geralmente causada por fatores externos. Esta forma de perturbação no processo de reprodução aumenta a variabilidade dos indivíduos e assim contribui para a sobrevivência da população aumentando sua possibilidade de adaptação à novas condições impostas pelo meio. Nós utilizamos este conceito para definir um algoritmo de mutação, conforme é visto a seguir.

Inicia-se pela definição do coeficiente de mutação $0 \leq \beta \leq 1$ que define o grau de perturbação que será aplicado em $B^{(n)}$. Uma vez que os segmentos sonoros pertencem ao espaço vetorial $W = R^N$ um vetor de mutação é gerado com N elementos escalares de valor aleatoriamente espalhados dentro do intervalo $[1-\beta, 1]$, chamado de intervalo de perturbação. Assim o operador mutação é definido na n-ésima geração pelos passos abaixo:

1. Gere um vetor mutação $m = [m_1, m_2, m_3, \ldots, m_{1024}]$ onde cada m_j pertence ao intervalo $[1-\beta, 1]$.
2. Aplique a perturbação $w_j^{n+1} = w_j^n.m$ para todos os indivíduos elementos de $B^{(n)}$, de modo que:

$$B^{(n)} = \{ w^{(1,n)}, w^{(2,n)}, \ldots, w^{(M,n)} \} \qquad (2.2)$$

3. Repita os passos (1) e (2) para cada geração de $B^{(n)}$.

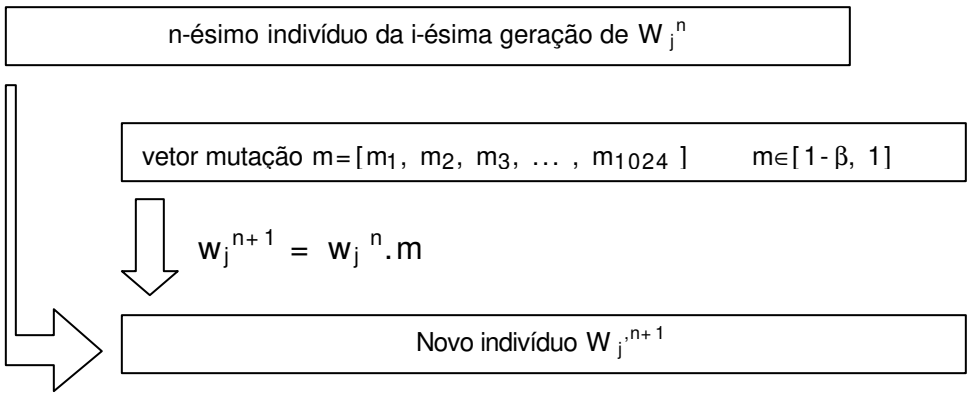


Figura 2.4 Diagrama da operação de mutação.

A força da mutação é controlada pelo parâmetro β dentro do intervalo real [0,1]. Quanto mais próximo β estiver de 0 mais fraca será a mutação. A medida que β se aproxima de 1 a mutação se torna mais forte. Note que o operador de mutação neste modelo de síntese evolutiva pode ser visto como um processo similar ao utilizado pela síntese *waveshaping*, onde o segmento sonoro é modificado por outro segmento aleatório, de acordo com uma proporção dada pelo coeficiente de mutação.

2.4 A medida da adequação do som

Inicialmente define-se uma métrica auxiliar, chamada de função de distância. Nosso modelo matemático encara os indivíduos, que são segmentos de som digital, como vetores em um espaço vetorial real $W = R^N$ onde cada vetor do espaço tem N componentes.

2.4.1 Distância vetorial

Dados dois vetores v e w no espaço W, define-se a métrica Euclidiana entre eles como se segue:

$$d_2(w, v) = [\ \Sigma_{i = 1, \ldots, 1024}\ (w_i - v_i)^2\ ]^{1/2} \qquad (2.3)$$

Esta métrica induz à norma $|w| = [\ \Sigma_{i = 1, \ldots, 1024}\ (w_i)^2\ ]^{1/2}$ e dá a energia total do som resultante. Contudo, outras métricas podem ser usadas e testadas para este fim.

2.4.2 Distância de Hausdorff

Agora nós definimos a distância entre os dois conjuntos. Seja: $T = \{t_1, t_{2,..}, t_L\}$ o conjunto alvo de L indivíduos e $B^{(n)} = \{w_1, w_{2,....}, w_M\}$ a n-ésima geração do conjunto população de M indivíduos. Uma vez que estes são sub-conjuntos do espaço vetorial W pode-se definir a distância entre eles como se segue:

$$d(T, B^{(n)}) = \min \{d_2(t_j, w_k)\} \qquad (2.4)$$

com $j = 1,..,L$ e $k = 1,..,M$, onde L é o número de indivíduos no conjunto alvo T e M o número de indivíduos no conjunto população $B^{(n)}$.

Uma vez que T e $B^{(n)}$ são conjuntos finitos, o mínimo da equação (2.4) é alcançado pelo menos por um vetor em $B^{(n)}$, que nós denotamos por $w_*{}^{(n)}$. Este vetor é chamado de melhor indivíduo na n-ésima geração de $B^{(n)}$ comparada com o conjunto alvo T usando a métrica definida na equação (2.3).

Agora é possível definir a função de adequação da n-ésima geração
f: T x $B^{(n)}$ → $B^{(n)}$ como:

$$f(T, B^{(n)}) = w_*{}^n \qquad (2.5)$$

Métricas como a definida na equação (2.5) tem sido usadas por [Polansky,96] e colaboradores, em aplicações para composições. Estas foram chamadas de métricas morfológicas (*morphological metrics*) e foram usadas para medir similaridades entre linhas melódicas, acordes e outras estruturas musicais. O modelo de síntese evolutiva introduzido aqui estende a abordagem das métricas morfológicas para a síntese sonora que usa a computação evolutiva como ferramenta para a criação de novos sons.

3 Utilização de curvas psicoacústicas na síntese evolutiva

No capítulo anterior foi dada a introdução formal ao método da síntese evolutiva, através da apresentação do primeiro modelo, onde os processos de reprodução e seleção ocorrem sobre o segmento sonoro representado pelo indivíduo.

Aqui neste capítulo, apresentamos o segundo modelo do método de síntese evolutiva de segmentos sonoros. Os processos de reprodução e seleção deste segundo modelo são quase os mesmos definidos no capítulo 2 com a diferença que agora estes ocorrem sobre o genótipo do indivíduo, formado por curvas psicoacústicas.

3.1 Introdução

No capítulo anterior, o genótipo do indivíduo da síntese evolutiva era o próprio r-ésimo segmento sonoro pertencente à n-ésima geração do conjunto população $B^{(n)}$. Este era constituído por um único segmento de som discretizado no tempo que determinava todas as suas características sonoras. Apesar dessa abordagem conseguir englobar todas as características sonoras do indivíduo na representação de seu genótipo (uma vez que este era identificado com o próprio segmento sonoro), ela não permite a manipulação independente dessas características. É necessário, então, estabelecer um método que permita a manipulação independente das características sonoras do indivíduo que são pertinentes à percepção auditiva. Essa nova abordagem permite uma flexibilidade muito maior no controle do processo de síntese levando a resultados esteticamente mais próximo aos desejados pelo usuário.

Este próximo passo é tratado neste capítulo, que implica uma nova representação para o genótipo do som. Este novo genótipo do indivíduo é formado por três curvas psicoacústicas extraídas do segmento sonoro, a saber: 1) percepção da intensidade sonora, loudness, $L(t)$; 2) percepção da freqüência, pitch, $P(t)$; 3) percepção das componentes de freqüência, (espectro de freqüência ou simplesmente de espectro) $S(f)$.

3.2 Do segmento sonoro às curvas psicoacústicas

Neste capítulo, do mesmo modo que no capitulo anterior, vamos considerar o segmento sonoro denotado por w_r^n, o r-ésimo indivíduo da n-ésima geração da população. Este segmento sonoro possui todas as características do som relativas à percepção auditiva. No entanto, estas características não estão apresentadas separadamente. Afim de se fazer uma análise qualitativa e quantitativa destas características, tomamos como nosso ponto de partida conceitos da psicoacústica, ou seja, a ciência que estuda a percepção sonora dos fenômenos acústicos. As grandezas psicoacústicas procuram analisar separadamente cada uma das características que definem a percepção sonora. No apêndice deste trabalho, definimos as grandezas psicoacústicas mais relevantes, que são as que caracterizam a percepção da intensidade sonora, a percepção da freqüência e a percepção da composição espectral, ou simplesmente loudness, pitch e espectro. Algumas definições de timbre associam-no à magnitude da distribuição do espectro de freqüência das componentes sonoras, enquanto outras definições de timbre consideram-no como também dependente de outras grandezas psicoacústicas, como *loudness* e *pitch*. Para sons bem comportados, como os sons periódicos, existe uma forte relação entre a curva de espectro do som e seu timbre, enquanto para sons não-periódicos ou ruidosos o espectro não se mostra suficiente para descrever o timbre sonoro. Assim decidimos utilizar as três curvas psicoacústicas para caracterizar o novo genótipo do som. Definimos então o genótipo do r-ésimo indivíduo da n-ésima geração como sendo a tripla de funções:

$$g_r^n = \left(l_r^n(t), p_r^n(t), s_r^n(f)\right) \qquad (3.1)$$

onde a variável temporal t pertence a um intervalo [0,T] definido pelo tamanho do segmento sonoro e a variável de freqüência f pertence a um intervalo $[0, f_{MAX}]$. Do ponto de vista prático a freqüência máxima f_{MAX} é dada pela metade da taxa de amostragem F_s (vide apêndice).

As triplas g_r^n podem ser vistas como elementos de um espaço vetorial G o qual denominamos espaço das curvas psicoacústicas. Claramente, G é um produto cartesiano de 3 espaços de funções contínuas, a saber,

$$G = L \times P \times S$$

onde L é o espaço das funções de loudness, P é o espaço das funções de pitch e S é o espaço dos espectros. Os espaços L, P, S são espaços vetoriais sob as operações usuais de soma de funções e produto por escalar.

Todo o processo de síntese evolutiva pode ser visto como "trajetórias finitas e discretas" de um conjunto inicial de M genótipos em G, ou ainda, como deformações deste conjunto inicial ao longo do tempo. Note que a variável tempo aqui considerada é extrínseca às curvas psicoacústicas sendo nele que ocorrem as diversas gerações oriundas da população inicial. Podemos chamar este tempo de "tempo genético". Deste ponto de vista o conjunto alvo, que também é um subconjunto finito de G, funciona como um conjunto "atrator" das trajetórias em G.

Como o segmento sonoro que representa o indivíduo é digital, ou seja, discretizado no domínio do tempo à uma taxa de amostragem F_s, as curvas ***l(t), p(t), s(f)*** são também discretizadas. A representação discreta do segmento sonoro é dada por w_r^n onde: k = 1, ...,N, corresponde a uma discretização da variável t no intervalo [0,T]. Incidentalmente iremos utilizar a mesma variável k para fazer também a representação discreta da variável freqüência f por questão de simplicidade. Pela teoria da amostragem o intervalo de tempo representado por este segmento sonoro é dado por $t_k = (N+1) / F_s$ e a máxima freqüência nele contida é $f_k = (N+1) / 2.F_s$ [Oppenheim,75].

Abaixo mostramos um esquema da decomposição do genótipo g_r^n de um segmento sonoro w_r^n:

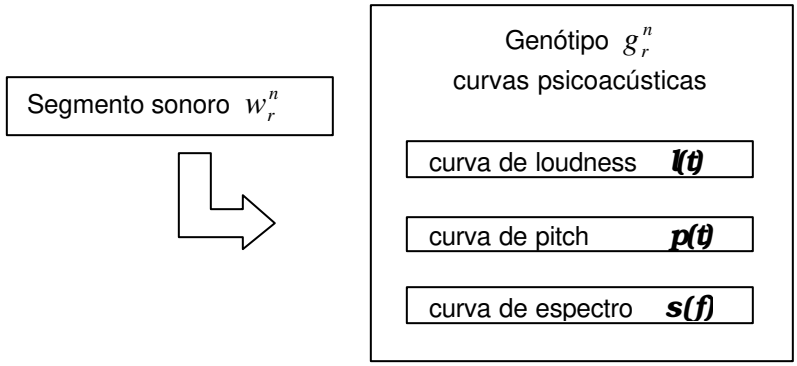


Figura 3.1 Diagrama da representação de genótipo do indivíduo sonoro.

3.3 Extração das curvas psicoacústicas dos indivíduos

Sabe-se que o processo da percepção auditiva não é linear. Apesar do ouvido humano ser sensível à percepção de freqüências aproximadamente entre 20 e 20000Hz, o ouvido privilegia a percepção de sons nas freqüências relacionadas à fala humana. Por isso o ouvido é mais sensível à variação das baixas freqüências sonoras, até aproximadamente 1000Hz, onde se concentram a grande parte das freqüências dos sons formantes das linguagens humanas.

O limiar da percepção sonora determina o grau mínimo de intensidade (em SPL dB) que a audição passa a perceber o som, em uma dada freqüência. Conforme visto no apêndice deste trabalho, os experimentos elaborados por *Fletcher* e *Munson* tiveram como objetivo determinar a relação entre a percepção da intensidade e a freqüência sonora em sons simples, ou senoidais (contendo apenas uma componente em freqüência). Estes foram medidos empiricamente utilizando um público diverso e osciladores senoidais para gerar sons com uma única componente em freqüência (senóides). O resultado gráfico do limite aproximado da percepção auditiva humana é dado a seguir:

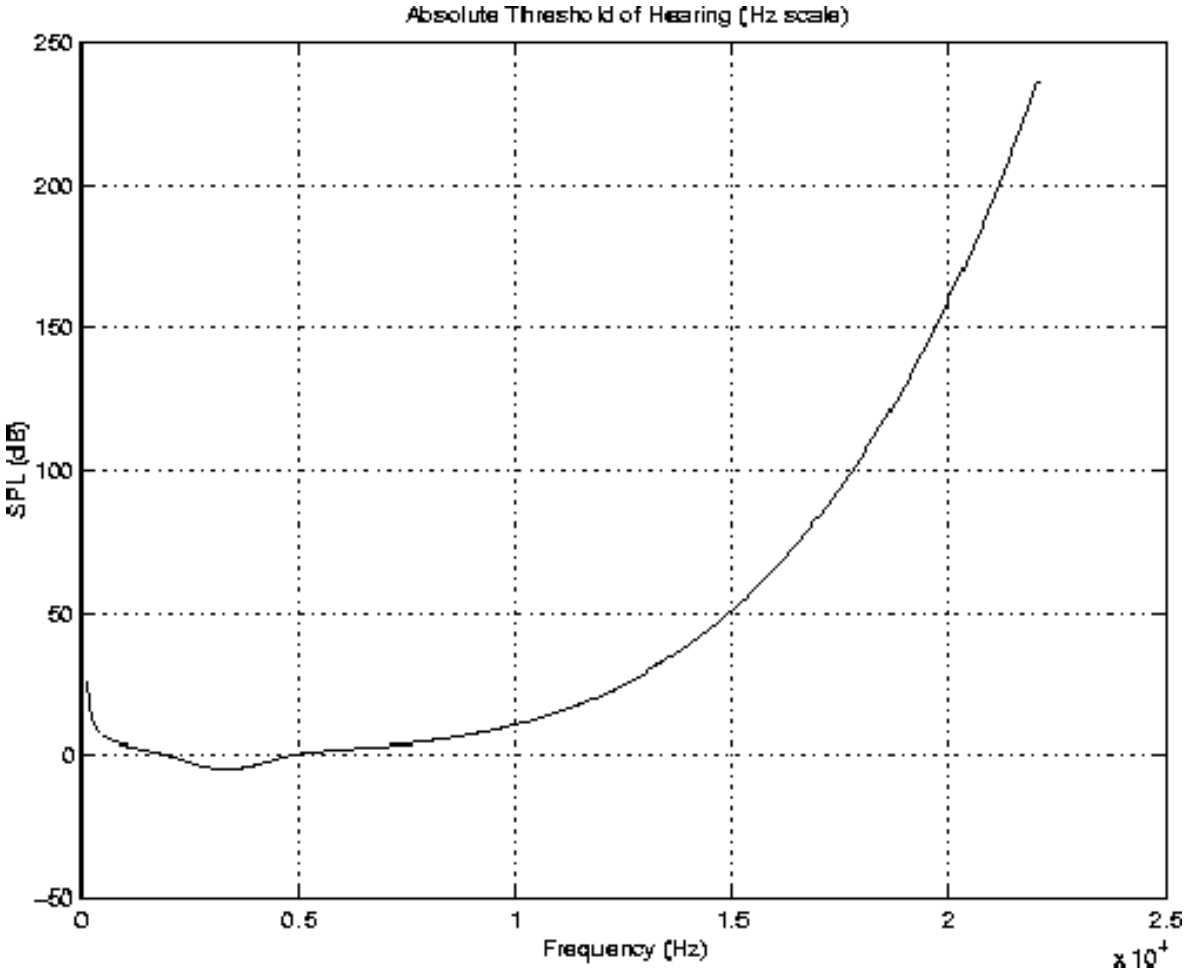


Figura 3.2 Limiar da percepção nas escalas de freqüência.

O modelamento do limiar da percepção auditiva (LPA) é dado abaixo.

$$\mathrm{LPA(f)} = 3{,}64\,.\,(\frac{f}{1000})^{-0{,}8} - 6{,}5.\exp^{-0{,}6.((\frac{f}{1000})-3{,}3)^2} + 10^{-3}.(\frac{f}{1000})^4 \quad (\mathrm{dB\ SPL}) \qquad (3.2)$$

Esta equação modela a primeira curva das curvas de Fletcher-Munson, também chamadas de curvas de *equal-loudness*, ou isofônicas. As curvas isofônicas se estendem do limiar da percepção, definido em 0dB, dado pela equação acima, até o chamado limiar da dor, que é próximo de 120dB para freqüências de 1000Hz. Uma curva isofônica que pode ser utilizada para o nosso modelamento da percepção da intensidade sonora (a curva de loudness, ou L(t)) é o seu modelamento para 40dB, dado abaixo:

$$I_{40dB}(f) = \left(\frac{0{,}05f + 4000}{f} \right) 10^{-\left(\frac{0{,}05f + 4000}{f} \right)} \qquad (3.3)$$

A percepção da freqüência sonora, dada pela grandeza psicoacústica pitch, depende da complexidade do som. Sons periódicos costumam apresentar um pitch enquanto sons ruidosos ou não-periódicos não definem um pitch à percepção auditiva, apesar de ambos serem formados por diversas componentes em freqüência. A sensação de pitch está intimamente relacionada à relevância da componente fundamental do som, em relação a suas outras componentes. Um som senoidal, que apresenta apenas uma componente em freqüência, define claramente um pitch (como o som de um diapasão de metal, ou "*tunning fork*"). Assim a freqüência da fundamental do som é diretamente relacionada ao seu pitch A curva de pitch pode ser calculada através de algoritmos de detecção instantânea da freqüência, similar àqueles utilizados em afinadores de instrumentos musicais. Muitos desses programas são conhecidos como "*pitch tuners*". O algoritmo que utilizamos detecta a componente em freqüência de maior magnitude presente no segmento sonoro que, via de regra, é o harmônico fundamental do som. Desse modo, ao isolarmos a fundamental deste som, calculamos a sua curva de pitch.

É importante destacar que a curva de loudness é também calculada a partir da intensidade de sua fundamental. Uma vez que conseguimos detectar e isolar a fundamental do restante das componentes em freqüência de um som, podemos então medir simultaneamente as curvas de loudness e pitch. O algoritmo que desenvolvemos para a detecção dessas duas curvas psicoacústicas é explicado a seguir:

Dado um segmento sonoro, representado por uma seqüência de N pontos, amostrados em uma taxa de Fs pontos/segundo e com resolução em ponto-fixo, de b bits.

1. Aplica-se um filtro digital passa-faixa, com entre 20 a 5000Hz, o qual é considerado como sendo o intervalo de freqüência da componente fundamental que é mais importante para a percepção de pitch.
2. Normaliza-se a magnitude dos pontos da seqüência entre $[2^{(b-1)}, 2^{(b-1)}]$ que são mapeados para representação em ponto-flutuante entre [-1, 1]
3. Aplica um algoritmo de detecção de picos. Os picos de maior magnitude são associados à fundamental do som.
4. A variação de freqüência da fundamental é, conseqüentemente, a curva de pitch, e a variação de sua intensidade é a curva de loudness.

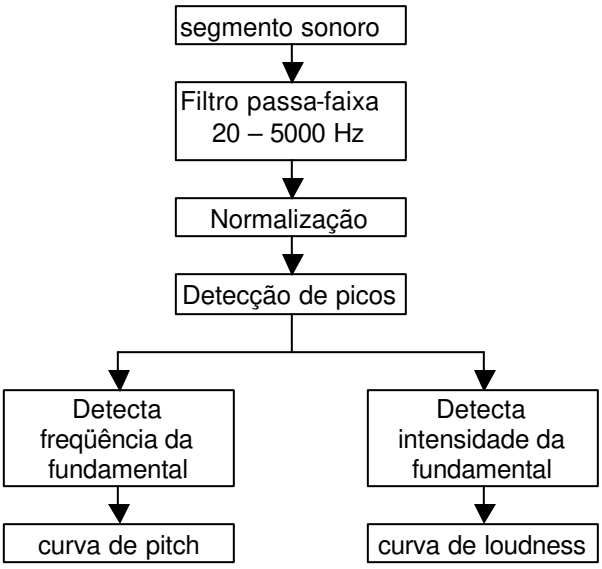


Figura 3.3 Cálculo das curvas de *loudness* e *pitch*.

A curva de espectro é calculada pela Transformada de Fourier. Como o segmento sonoro é digital, utiliza-se a transformada discreta de Fourier, ou DFT, para transformar o segmento sonoro do domínio discreto do tempo para o domínio discreto da freqüência. A DFT é dada pela formula abaixo [Steiglitz,96]:

$$Y(k) = \frac{1}{N}\sum_{n=1}^{N} X(n).\exp^{-i.\frac{2\pi(n-1)k}{N}} \qquad (3.3)$$

onde a inversa da transformada discreta de Fourier é dada por:

$$X(k) = \sum_{n=1}^{N} Y(n).\exp^{i.\frac{2\pi(n-1)k}{N}} \qquad (3.4)$$

Um algoritmo que calcula eficientemente a DFT é conhecido como FFT (*Fast Fourier Transform*). Pela natureza de seu algoritmo, a FFT é uma transformada que só aceita como entrada seqüências de N pontos, onde N é múltiplo de dois. A saída, ou resposta, da FFT é complexa e simétrica. Uma seqüência $X(n)$ de N pontos reais, onde $n = 1, \ldots N$, é transformada pela FFT em uma seqüência simétrica de pontos complexos, do tipo $Y(k) = a(k) + j.b(k)$, onde $Y(N-k) = Y^{*}(k)$. Exemplificando, para uma entrada de 8 pontos, cujos índices são: $n = \{1, 2, ,3, 4, 5, 6, 7, 8\}$ a saída da FFT tem os índices w que se repetem, do tipo $k = \{1, 2, 3, 4, 5, -4, -3, -2\}$. Se X é uma seqüência de N pontos reais que representa uma sinal discreto no domínio do tempo, a sua transformada $Y(k)$ no domínio da freqüência é dada tem sua magnitude dada por:

$$R(k) = |Y(k)| = \sqrt{a(k)^2 + b(k)^2} \qquad (3.4)$$

onde $k = 1, 2, \ldots, N/2$. A freqüência mais alta é representada pelo último valor desta seqüência $R(N/2)$ que é a freqüência de Nyquist, dada por $Fs/2$ Hz, onde Fs é a taxa de amostragem do sinal discretizado. Cada ponto k representa um degrau de freqüência $\Delta f = (Fs/N)$ Hz. A fórmula

3.5 deveria nos fornecer o espectro com uma amostragem de N pontos. No entanto devido à simetria acima mencionada temos apenas um a amostra de N/2 pontos. Uma vez que amostramos o loudness e o pitch com N pontos, é desejável fazer o mesmo com a curva do espectro.

Para se obter uma seqüência que represente o espectro do som com N pontos, utiliza-se a técnica conhecida como "*zero-padding*". A seqüência X é substituída por uma seqüência X' onde:

$$X'(k) = \begin{cases} 0 & se \quad 1 \le k < \frac{N}{2} \\ X(k - \frac{N}{2} + 1) & se \quad \frac{N}{2} \le k \le N \\ 0 & se \quad N < k \le 2N \end{cases} \qquad (3.5)$$

Graficamente a função pode ser visualizada como:

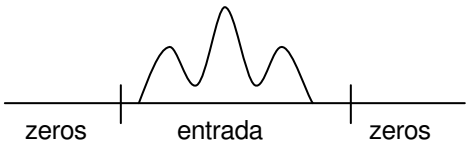


Figura 3.4 Diagrama da técnica de *zero-padding*.

A nova seqüência X' tem o dobro do tamanho de X. Assim a curva do espectro é dada pela magnitude de cada ponto da saída da FFT, ou seja:

$$S(k) = |FFT(X')| \qquad (3.6)$$

onde k= 1, ... N, e S(k) é uma seqüência de números reais, que representam as amplitudes das componentes até a freqüência de Nyquist.

A figura abaixo mostra o esquema a seqüência de operações para se obter o espectro digitalizado.

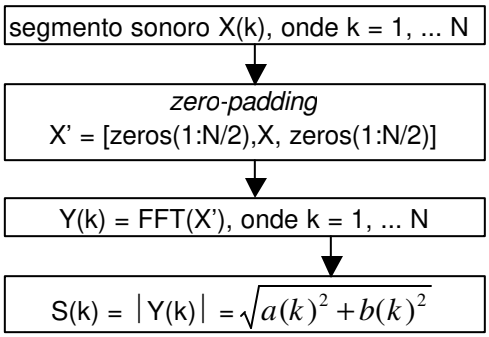


Figura 3.5 Cálculo da curva de espectro.

3.4 A medida da adequação do som através das curvas psicoacústicas

Vimos anteriormente que L,P,S são espaços vetoriais. É possível introduzir nesses espaços uma infinidade de funções que medem distâncias entre elementos de um mesmo espaço. Estas funções são também denominadas métricas. Nada impede que L,P,S tenham métricas diferentes entre si. No entanto, por simplicidade, neste trabalho utilizaremos a mesma métrica nos três espaços, a saber, a bem conhecida métrica euclidiana.

Dadas c_a e c_b duas curvas psicoacústicas que podem ser (simultaneamente) curvas de loudness, pitch ou espectro, associadas respectivamente aos segmentos sonoros w_a e w_b, definimos a distância entre duas curvas genéricas c_a c_b como:

$$d_c(c_a, c_b) = \sqrt{\sum_{k=1}^{N} \left| c_a(k) - c_b(k) \right|^2} \tag{3.7}$$

Estas funções distâncias acima são denominadas distâncias (ou métricas) euclidianas. É ainda possível definir métricas com funções peso. Por exemplo para se ρ_k, com $1 \leq k \leq N$ é um conjunto de N pesos, podemos definir a seguinte métrica.

$$d_c(c_a, c_b) = \sqrt{\sum_{k=1}^{N} \rho_k \left| c_a(k) - c_b(k) \right|^2} \tag{3.8}$$

Uma vez que a informação mais significativa está no inicio da curva, pode-se definir um peso decrescente, do tipo:

$$d_c(c_a, c_b) = \sqrt{\sum_{k=1}^{N} (N-k) \left| c_a(k) - c_b(k) \right|^2} \tag{3.9}$$

Uma outra possibilidade é definir a distância pela derivada das curvas psicoacústicas. Como estas são seqüências discretas, a equação a diferenças, do tipo:

$c'(k) = c(k) - c(k-1)$ para k= 1, 2, …, N e $c'(1) = 0$

Assim a distância da derivada é:

$$d_c(c_a, c_b) = \sqrt{\sum_{k=1}^{N} \left| c'_a(k) - c'_b(k) \right|^2} \tag{3.10}$$

Uma vez que temos as três métricas nos espaços L,P,S podemos então definir uma métrica no espaço G= LxPxS como uma média aritmética. Mais detalhadamente, se $g_a = (l_a, p_a, s_a)$ e $g_b = (l_b, p_b, s_b)$ são dois elementos quaisquer de G, definimos a distância entre eles como:

$$D(g_a, g_b) = \frac{1}{3}\left[d_L(l_a, l_b) + d_P(p_a, p_b) + d_S(s_a, s_b)\right] \qquad (3.11)$$

É importante enfatizar que outras escolhas para a métrica D são possíveis e na realidade existe uma infinidade delas. A nossa escolha como dito anteriormente pautou pela simplicidade. Utilizamos estas funções descritas acima para determinar a distância entre o genótipo de cada indivíduo da população e o conjunto de genótipos dos indivíduos do conjunto alvo.

3.4.1 Distância de Hausdorff

Do mesmo modo que no capítulo 2, para implementar um processo de seleção precisamos de um critério de adequação (*fitness*) para que se encarregue da escolha do melhor indivíduo, aquele que é mais adaptado em uma dada geração. Para esse propósito, utilizamos a distância de Hausdorff, desta vez entre dois subconjuntos de G, como segue.

Seja: $G^{(n)} = \{ g_1^n, g_2^n, \ldots, g_M^n \}$ a n-ésima geração de genótipos associada ao conjunto população $B^{(n)}$ com M indivíduos e $\overline{G} = \{ \overline{g}_1, \overline{g}_2, \ldots, \overline{g}_Q \}$ o conjunto dos genótipos de um conjunto alvo de indivíduos T fornecido pelo usuário. A distância de Hausdorff entre os conjuntos $G^{(n)}$ e $\overline{G}$ é definida como:

$$D_H(G^{(n)}, \overline{G}) = \min_{\substack{1 \le a \le M \\ 1 \le j \le Q}} D(g_a^n, \overline{g}_j) \qquad (3.12)$$

Como $G^{(n)}$ e $\overline{G}$ são conjuntos finitos, o mínimo da equação acima é atingido por pelo menos um vetor em $G^{(n)}$. Este elemento ótimo de $G^{(n)}$ corresponde ao melhor indivíduo da n-ésima geração $B^{(n)}$, aquele que mais se assemelha a pelo menos um elemento do conjunto alvo T, usando a métrica de Hausdorff. É este elemento ótimo que, além de passar à geração seguinte $G^{(n+1)}$ invariante, transfere para todos os outros elementos de $G^{(n)}$ seu material genético através da operação de crossover (definida a seguir). O ciclo então se repete na geração seguinte $G^{(n+1)}$.

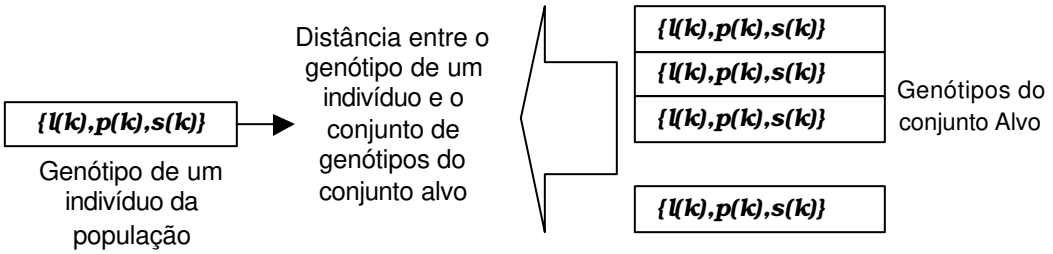


Figura 3.6 Distância vetorial entre genótipo de um indivíduo e os genótipos do conjunto alvo.

3.5 As operações genéticas sobre as curvas psicoacústicas

No capítulo 2 as operações genéticas de crossover e mutação eram feitas diretamente sobre o segmento sonoro w_r^n. Neste capítulo estas operações genéticas manipulam as três curvas psicoacústicas { ***l(k), p(k), s(k)***} que compõem o genótipo do segmento sonoro. As operações genéticas de crossover e mutação sobre as curvas psicoacústicas são definidas a seguir.

3.5.1 Operador Crossover

Seja $B^{(n)}$ uma população de segmentos sonoros. Definimos a operação de crossover como segue: dado o genótipo g_r^n de um indivíduo arbitrário da população $B^{(n)}$ e o genótipo g_*^n do melhor indivíduo como definido na secção anterior, o crossover de g_r^n com g_*^n é definido como:

$$C(g_r^n, g_*^n) = \alpha . g_r^n + (1-\alpha).g_*^n = g_r^{(n+1)} \tag{3.13}$$

onde α é um parâmetro de controle que denominamos **taxa de crossover**, e que pertence ao intervalo real [0,1].
Claramente, as curvas psicoacústicas do novo indivíduo, após a operação de crossover, são dadas por:

$$\begin{aligned} l_r^{n+1}(k) &= \alpha . l_r^n(k) + (1-\alpha) l_*^n(k) \\ p_r^{n+1}(k) &= \alpha . p_r^n(k) + (1-\alpha) . p_*^n(k) \\ s_r^{n+1}(k) &= \alpha . s_r^n(k) + (1-\alpha) . s_*^n(k) \end{aligned} \tag{3.14}$$

A operação crossover cria o genótipo de um novo indivíduo da geração seguinte. Se esta seqüência convergir ao longo da sucessão de gerações, ou seja:

$$\lim_{n\to\infty} D(g_r^{(n+1)}, g_r^n) \xrightarrow[n\to\infty]{} 0 \tag{3.15}$$

teremos que o r-ésimo indivíduo da (i+1)-ésima geração da população é aproximadamente um clone em G. Se a seqüência não convergir dizemos que esse indivíduo é um clado em G.

Podemos ainda refinar essa noção da seguinte maneira. Dado ε, um número arbitrariamente pequeno, dizemos que um elemento da população inicial g_r^0 é um ε-clone em G se $D(g_r^{(n+1)}, g_r^n) \le \varepsilon$, para n suficientemente grande.

3.5.2 Operador Mutação

A operação genética mutação insere variabilidade no genótipo da população. No caso deste capítulo, onde o genótipo é constituído de três curvas psicoacústicas, a mutação modifica as curvas psicoacústicas do genótipo do r-ésimo indivíduo da n-ésima geração, $w_r^n(k)$ do conjunto população, $B^{(n)}$. Seja o vetor mutação R onde cada elemento é um valor aleatório dentro do intervalo $R(k) = [1-\beta, 1]$ para $k = 1, 2, \ldots N$ onde β é o parâmetro de controle da taxa de mutação, definido para o intervalo $0 \le \beta \le 1$. Tem-se que a operação mutação sobre o genótipo do r-ésimo indivíduo da n-ésima geração da população como:

$$M(g_r{}^n(k)) = g_r{}^n(k).R(k) \tag{3.16}$$

A operação de mutação aplicada para cada curva psicoacústica e´dada por:

$$\begin{aligned} l_r^n(k) &= l_r^n(k).R_l(k) \\ p_r^n(k) &= p_r^n(k).R_p(k) \\ s_r^n(k) &= s_r^n(k).R_s(k) \end{aligned} \tag{3.17}$$

onde k = 1, 2, 3, ... N para N sendo o último ponto do segmento sonoro.

3.6 Construção do novo indivíduo a partir do genótipo modificado

A construção do indivíduo a partir do seu genótipo modificado só é necessária se este é escolhido como o melhor indivíduo da sua geração, ou seja, se este genótipo modificado realiza a distância de Hausdorff com o conjunto alvo. Neste caso, o indivíduo é modificado com base nas três curvas psicoacústicas do seu genótipo. As transformações são **destrutivas**, ou seja, não preservam cópia do antigo indivíduo. O novo indivíduo resultante ocupará o lugar do indivíduo anterior no conjunto população.

Seja $w(t)$ um indivíduo (segmento sonoro) arbitrário de uma dada geração. Seja $l(t)$ a curva de loudness original do r-ésimo indivíduo, e $\bar{l}(t)$ a curva modificada pelos operadores genéticos. Definimos o operador de loudness sobre o indivíduo $w(t)$ como:

$$(Lw)(t) = \frac{\bar{l}(t)}{l(t)}.w(t) \tag{3.18}$$

Identicamente definimos o operador pitch como:

$$(Pw)(t) = \frac{\bar{p}(t)}{p(t)}.w(t) \tag{3.19}$$

e o operador espectro como:

$$(Sw)(t) = \Im^{-1}\left[\frac{\bar{s}(f)}{s(f)}.\Im\left[w(t)\right]\right] \tag{3.20}$$

De posse desses três operadores definimos finalmente a nossa transformação. Dado w_r^n um segmento sonoro da n-ésima geração, o seu sucessor na (n+1)-ésima geração é o segmento sonoro definido por:

$$\bar{w}(t) = SPL\left[w(t)\right] \tag{3.21}$$

Observe que, a princípio, os operadores SPL não comutam entre si. Assim existem seis possibilidade de comutação para se obter diferentes gerações sonoras: SPL, SLP, PSL, PLS, LPS e LSP.

3.7 A síntese evolutiva baseada na manipulação de curvas psicoacústicas

A utilização de curvas psicoacústicas como objeto de medida de adequação dos indivíduos na síntese evolutiva torna este processo mais sofisticado e mais coerente com os propósitos de uma síntese sonora. As curvas psicoacústicos procuram mapear quantitativamente a percepção da informação sonora. Assim, os dois processos que compõem a síntese evolutiva (reprodução e seleção) são melhorados. A reprodução, dada pelos operadores genéticos crossover e mutação, manipula as curvas psicoacústicas dos segmentos sonoros, o que implica na manipulação de características perceptualmente significativas deste. Do mesmo modo, a seleção, dada pela medida de distância do indivíduo ao conjunto alvo, calcula agora a distância entre as características perceptuais do indivíduo na população e os indivíduos que compõem o conjunto alvo. Abaixo temos a seqüência de passos dada pelo algoritmo da síntese evolutiva utilizando curvas psicoacústicas:

1	$B^{(n)} = \{w_1^n, w_2^n, ...w_r^n, ...w_M^n\}$ $T = \{t_1, t_2, ...t_j, ...t_Q\}$ Se n = 1 escolhe w_*^n	**Dado um conjunto população de M indivíduos, em sua primeira geração (n=1), e um conjunto alvo com Q indivíduos Se esta é a primeira geração (n=1) então escolhe-se aleatoriamente um indivíduo como o melhor indivíduo da população.**
2	$g_r^n(k) = \{l_r^n(k), p_r^n(k), s_r^n(k)\}$ onde: k = 1, 2, ..., N	**Calcula-se o genótipo de todos indivíduo da população B e alvo T pelo cálculo das curvas psicoacústicas de loudness pitch e espectro.**
3	$C(g_r^n) = \alpha.g_r^n + (1-\alpha).g_*^n$	**Aplica-se a operação de crossover em todos os genótipos dos indivíduos da população com o melhor indivíduo calculado na geração anterior. O operador crossover escolhe aleatoriamente uma das curvas L, P, S para realizar a operação. As outras duas curvas permanecem inalteradas.**
4	$M(g_r^n) = \begin{cases} l_r^n(k) = l_r^n(k).R_l(k) \\ p_r^n(k) = p_r^n(k).R_p(k) = g_r^{n+1} \\ s_r^n(k) = s_r^n(k).R_s(k) \end{cases}$	**Na seqüência, aplica-se a operação de mutação sobre o genótipo modificado pelo crossover. R é uma seqüência de N pontos aleatórios.**
5	$D_H(G^{(n)}, \overline{G})$	**Para cada indivíduo da população, calcula-se a distância de Hausdorff, que é a menor distância entre o genótipo do r-ésimo indivíduo e todos os genótipos do conjunto alvo**
6	$g_*^{n+1} = \{l_r^{n+1}, p_r^{n+1}, s_r^{n+1}\}$	**Acha-se o genótipo do melhor indivíduo, aquele com a menor distância de Hausdorff, resultado sonoro da n-ésima geração da síntese evolutiva.**
7	$L(w_r^{n+1}) = \frac{\overline{l_r^{n+1}}}{l_r^{n+1}}.w_r^{n+1}$	**Reconstrói o loudness do melhor indivíduo pela operação L**
8	$P(w_r^{n+1}) = \frac{\overline{p_r^{n+1}}}{p_r^{n+1}}.w_r^{n+1}$	**Reconstrói o pitch do melhor indivíduo pela operação P**
9	$S(w_r^{n+1}) = \Im^{-1}\left[\frac{\overline{s(f)}}{s(f)}.\Im\left[w_r^{n+1}\right]\right]$	**Reconstrói o espectro do melhor indivíduo pela operação S**
10	$SPL, SLP, LPS, LSP, PLS, PSL$	**As operações L,P,S não são necessariamente comutativas. Pode-se reconstruir o melhor indivíduo de seis maneiras diferentes.**
11	w_*^n	**O melhor indivíduo da n-ésima geração da população é o som sintetizado.**
12		**Repete os passos 1 a 11**

As gerações do conjunto população são contadas pelo incremento da variável n. A cada geração os genótipos dos indivíduos são modificados pelos operadores genéticos, o que faz com que

a população se modifique mas permaneça com o mesmo número de indivíduos. Isto pode ser interpretado como um conceito de morte, ou eliminação do indivíduo por substituição deste pelo seu descendente, após a reprodução, exceção feita para o melhor indivíduo que passa para a próxima geração inalterado.

Por razões de eficiência computacional, os indivíduos da população não são reconstruídos a cada geração da população. A construção do novo indivíduo a partir de seu genótipo modificado só é necessária se o genótipo deste indivíduo for escolhida pela função de adequação como sendo melhor indivíduo, o que significa que este é o resultado da síntese evolutiva e portanto deve ser transformado em som. Para os outros indivíduos da população as operações genéticas a cada geração continuam ocorrendo apenas sobre o seu genótipo, que é mantido numa tabela associada a este segmento sonoro. Assim, para cada geração, apenas os genótipos são modificados pelos operadores genéticos. Dessa maneira a quantidade de cálculos da síntese evolutiva é otimizada.

A síntese evolutiva possui dois processos básicos: a reprodução e a seleção. O processo de reprodução modifica os genótipos através dos operadores genéticos crossover e mutação. O processo de seleção mede a distância entre os genótipos dos indivíduos da população e o conjunto de genótipos do conjunto alvo. O conjunto alvo pode ser modificado a qualquer momento pelo usuário, porém a cada modificação de um elemento do conjunto alvo, o correspondente genótipo deve ser recalculado e este só terá efeito no processo evolutivo na geração seguinte.

Tem-se a seguir duas figuras que ilustram respectivamente os processos de reprodução e seleção da síntese evolutiva utilizando curvas psicoacústicas.

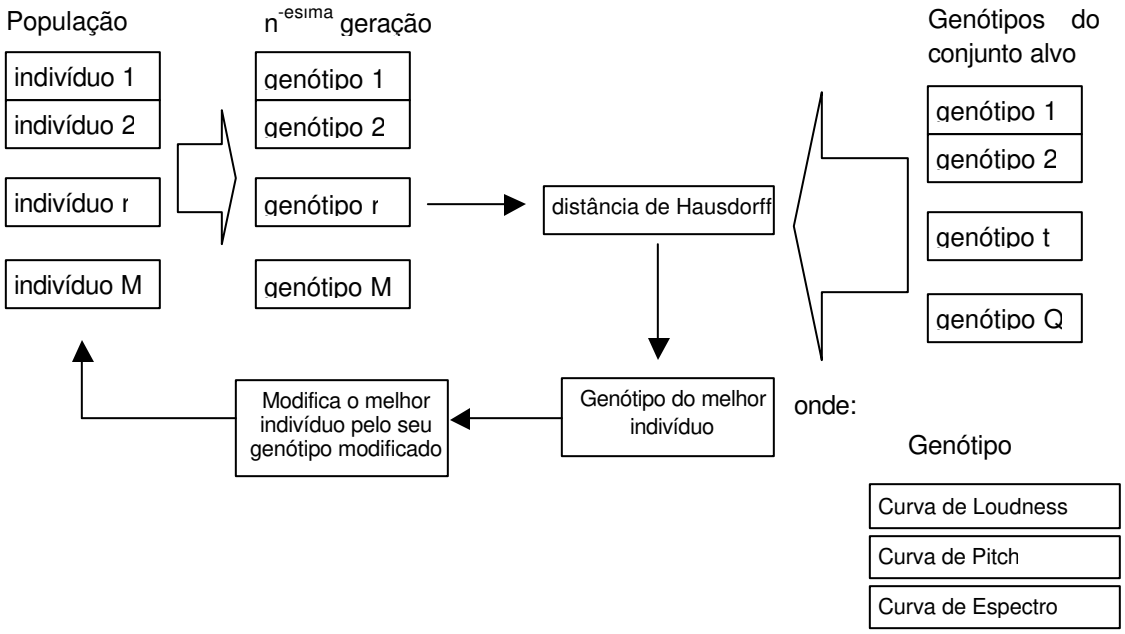


Figura 3.7 Diagrama do processo da seleção, que mede a distância entre o genótipo de cada indivíduo da população com o conjunto de genótipo dos indivíduos do conjunto alvo e seleciona o mais próximo, ou seja, o melhor indivíduo.

Note que o cálculo do genótipo deve ser feito toda vez que se modifica um indivíduo (segmento sonoro). Apesar de possível, por coerência com a evolução biológica, optou-se por não modificar o conjunto população durante a operação da síntese evolutiva. Já o conjunto alvo pode e deve ser modificado a qualquer momento. Cada modificação do conjunto alvo eqüivale à modificação do meio ambiente que, na evolução biológica, condiciona o processo de seleção dos indivíduos.

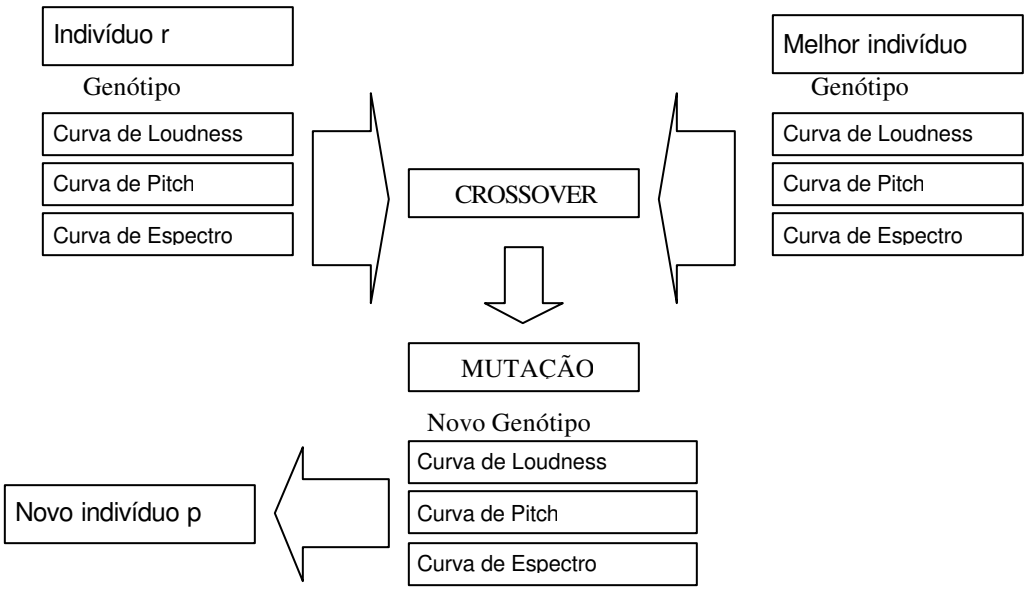


Figura 3.8 Diagrama do processo da reprodução, que aplica os operadores crossover e mutação no genótipo de cada indivíduo na população.

4 Simulação dos modelos de síntese evolutiva

No capítulo 2 apresentamos o primeiro modelo para o método da síntese evolutiva. Este modelo é baseado nos processos de seleção e reprodução de indivíduos que são segmentos sonoros. A seleção destes indivíduos é dada pela medida da distância entre cada indivíduo da população e o conjunto alvo. A reprodução é dada pela ação dos operadores genéticos crossover e mutação. No capítulo 3 apresentamos um segundo modelo da síntese evolutiva. Neste modelo expandimos o modelo anterior da síntese evolutiva apresentado no capítulo 2 com a inclusão de curvas psicoacústicas como genótipo do indivíduo. Utilizamos três curvas psicoacústicas extraídas dos indivíduos para caracterizar o seu genótipo: as curvas de loudness, pitch e espectro. Neste novo modelo os processos de seleção e reprodução apresentados no capítulo 2 passam a ser feitos sobre o genótipo do som.

Apresentaremos a seguir o resultado experimental das simulações desses dois modelos de síntese evolutiva. As simulações foram feitas utilizando o software MATLAB 6.5.

4.1 Cálculo do indivíduo como segmento sonoro

Para a síntese evolutiva, o indivíduo é um segmento sonoro, elemento de um conjunto chamado de população, onde os processos de reprodução e seleção ocorrem. A evolução da população ocorre em gerações sucessivas e é condicionada por um conjunto de indivíduos fornecidos pelo usuário, o conjunto alvo. No capitulo 2 os processos de reprodução e seleção da síntese evolutiva ocorrem diretamente sobre o segmento sonoro. No capítulo 3 estes processos ocorrem sobre as 3 curvas psicoacústicas do segmento sonoro: loudness, pitch e espectro, que compõem o genótipo do indivíduo.

Pela nomenclatura introduzida no capítulo 2, cada indivíduo é representado por $w_r^n = (a_1, a_2, \ldots a_N)$ que é uma seqüência de N pontos fixos (números inteiros) com resolução de 16 bits (entre -32768 a 32768). Este é o r-ésimo indivíduo da n-ésima geração da população $B^{(n)}$, que possui M indivíduos. O conjunto alvo possui Q indivíduos, similares aos do conjunto população, mas que não sofrem ação dos processos de reprodução e seleção da síntese evolutiva. Cada w_r^n é uma seqüência correspondente a discretização, ou amostragem, de um segmento de som. O processo de amostragem é feito por um conversor analógico-digital, AD, à uma dada taxa de amostragem fs amostras/s ou Hertz. A resolução de cada amostra é b=16 bits. O tempo de duração de um segmento sonoro é N/Fs segundos, e a sua relação sinal-ruído é $20.\log_{10}(2^b)$ dB. Pela teoria da amostragem, o segmento sonoro, ao ser discretizado, deve conter componentes em freqüências iguais ou menores a $fs/2$ Hz (freqüência de Nyquist), caso contrário o processo de conversão AD irá gerar no segmento sonoro um ruído de baixa freqüência conhecido por *aliasing noise*. Por esta razão, antes de ser discretizado, o segmento de som deve ser filtrado com freqüência de corte em fc <fs/2 Hz.

Os valores utilizados em nossas simulações para a taxa de amostragem e resolução do indivíduo sonoro são respectivamente: fs= 11025Hz e b= 16 bits. No capítulo 2 o valor de N é fixo em 1024 pontos, onde os indivíduos tem a duração de 1024/11025= 92,87mseg e relação sinal-ruído $20.\log_{10}(2^{16}) = 96{,}32$dB. No capítulo 3 o valor de N não é constante e assim os indivíduos da população podem ter tamanhos distintos, porém, o valor da taxa de amostragem fs permanece em 11025KHz. Por motivos de eficiência no processamento das simulações, optou-se, sempre que possível, em adotar segmentos sonoros de pequeno tamanho. Para intervalos de duração muito pequenos, existe a necessidade de se repetir o segmento sonoro sintetizado várias vezes para que este seja adequadamente escutado pelo usuário.

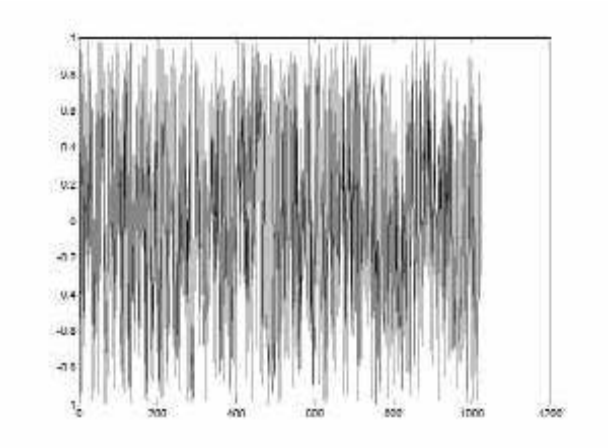
Analisamos duas técnicas de repetição de segmentos sonoros de curta duração. A primeira técnica é chamada de *overlap-and-add*. Nela o segmento sonoro de N pontos é "janelado" (multiplicado por uma função janela de N pontos que força as extremidades do segmento para amplitude próxima de zero) e somado a sua própria cópia "janelada" à uma taxa de entrelaçamento (*overlap*) no tempo que vai de zero (sem *overlap*) a N pontos (100% de *overlap*). Um valor típico para a taxa de *overlap* é 50%, ou N/2 pontos.

Criamos um *script* do MATLAB que gera um segmento aleatório de 1024 pontos entre [-1,1], que em termos sonoros representa o ruído branco. Janelamos este segmento por uma função janela

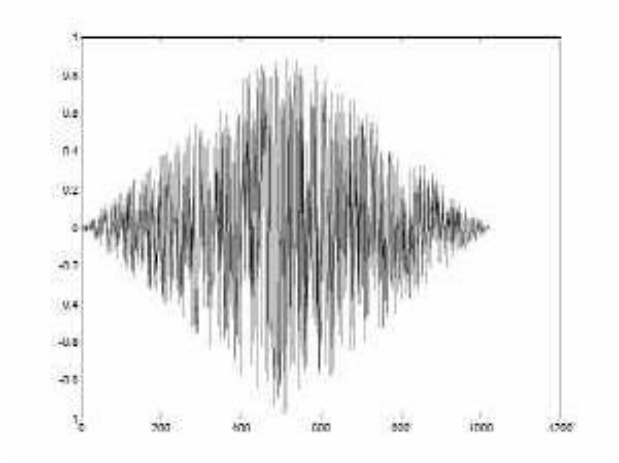
triangular. O resultado gráfico do *overlap-and-add* deste segmento à uma taxa de 50% é dado a seguir.

```
% oaa.m
% Overlap-and-Add
%
clear;
N=1024;     % numero de elementos do segmento sonoro
Fs=11025;   % taxa de amostragem
o=round(0.5*N);      % taxa de overlap, de zero a 50% de overlap
% segmento sonoro
ss(1:N)=(rand(1,N).*2)-1;     % ruído branco
jn=triang(N)';    % janela triangular
ssj=ss.*jn;     % janelamento do segmento sonoro ss
% Overlap-and-Add
      s1 = ssj(1:N-o);               % primeiro segmento
      s2 = ssj(N-o+1:N)+ssj(1:o);    % segundo segmento
      s3 = ssj(o+1:N-o);               % terceiro segmento
    ssjo = [s1,s2,s3];      % primeiro o-a-a
 while length(ssjo) < 3*Fs,     % faz o-a-a enquanto ssjo < 3 segundos
    ssjo =[ssjo,s2,s3];  % overlap-and-add e´ a junçao dos 3 segmentos
end;
```

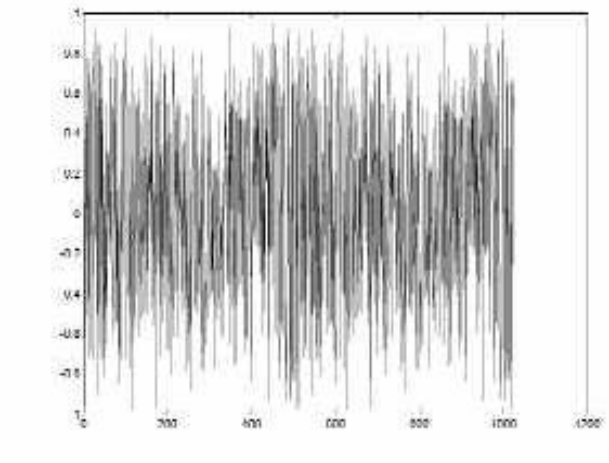
O resultado gráfico é visto na figura a seguir:



(a)



(b)

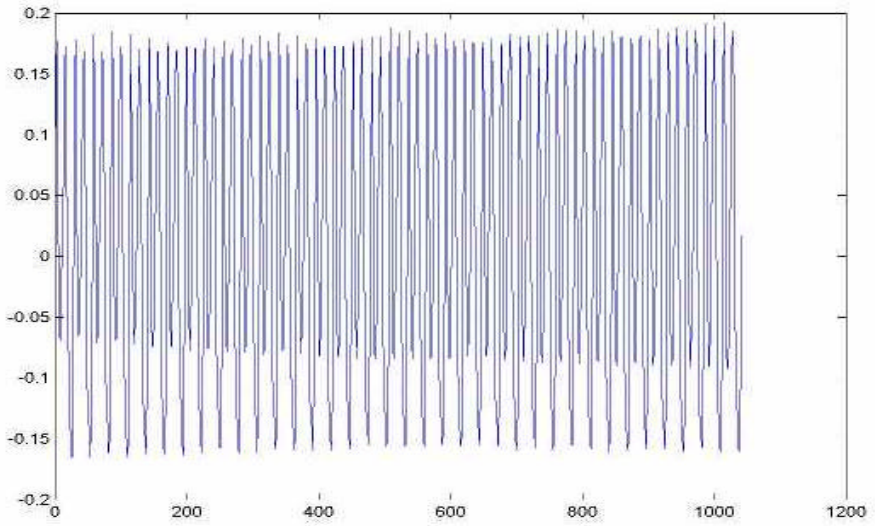


(c)

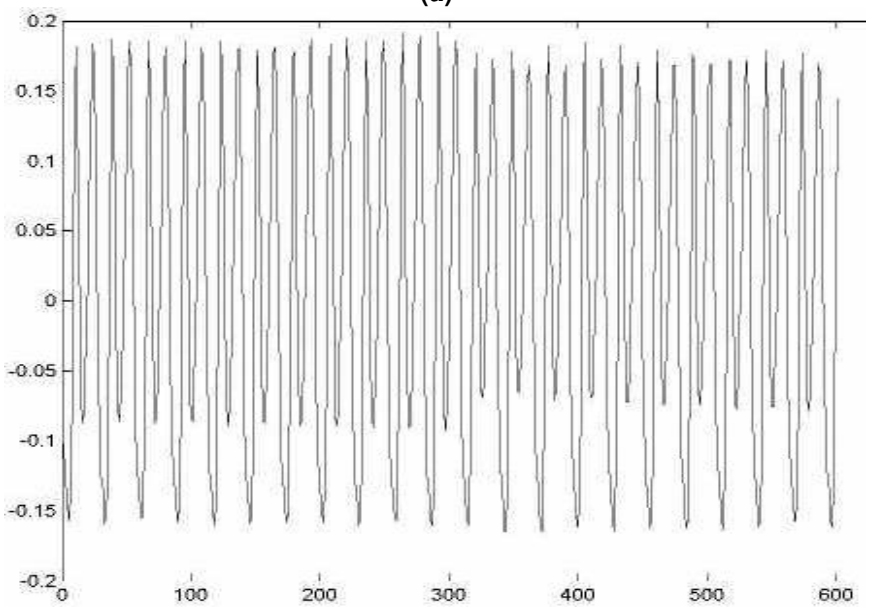
Figura 4.1 Reconstrução do segmento sonoro por *overlap-and-add* (a) segmento sonoro de ruído branco (b) segmento janelado (c) overlap-and-add de 50% do segmento janelado.

Apesar de visualmente parecer que a técnica de *overlap-and-add* reconstitui o sinal, em termos sonoros ela deixa a desejar. Pode-se ouvir a oscilação do loudness a cada união dos segmentos. Além disso, para sons mais comportados ou melódicos, pode ocorrer na união dos segmentos um cancelamento de componentes sonoras, tornando perceptível as regiões de união entre segmentos.

A outra técnica analisada vem da organização das amostras sonoras utilizada pela síntese *wavetable*. O método da síntese *wavetable* baseia no acesso e manipulação de amostras contidas numa tabela de segmentos sonoros, também conhecida por *lookup table*. As amostras de sons melódicos são armazenadas neste tabela em dois segmentos sonoros, o ataque e o ciclo. O ataque é a amostra do inicio do som que é a sua parte não-periódica e rica em padrões. O ciclo é a parte aproximadamente periódica do som, que se repete por um longo período de tempo. O segmento ciclo é arranjado de maneira a descrever um período completo de repetição do som. Este pode ser repetido indefinidamente de modo a ser percebido pela audição como se fosse um som contínuo.



(a)



(b)

Figura 4.2 Um segmento sonoro de 1024 pontos da amostra de uma nota de flauta, com taxa de amostragem fs=11025Hz, na forma de um ciclo periódico. (a) um ciclo completo. (b) detalhe da junção entre dois ciclos, (aproximadamente no meio da figura, em t_k = 300).

Utilizaremos este método de repetição de sons de curta duração para a simulação da síntese evolutiva, devido à sua simplicidade e eficiência.

4.2 Cálculo das curvas psicoacústicas do indivíduo

No capítulo 3 convencionamos que as curvas psicoacústicas de loudness, pitch e espectro constituem o genótipo do segmento sonoro. Pode-se derivar deste conceito que cada curva psicoacústica passe a representar um "gene" do indivíduo, ou então, no caso de se discriminar seções das curvas psicoacústicas, considera-las como um "cromossomo" contendo diversos genes. Até onde avançamos nos experimentos que realizamos aqui, consideramos, por simplicidade, cada curva psicoacústica equivalente a um gene. Pesquisas futuras poderão vir a expandir essa classificação de curvas psicoacústicas para cromossomos cujos operadores genéticos irão manipular as suas seções, ou genes.

O capítulo 2 não apresentou uma definição formal de genótipo pois as operações genéticas são feitas no próprio segmento sonoro. Neste primeiro modelo não houve como estabelecer uma distinção entre genótipo, cromossomo e indivíduo. No entanto, nesse primeiro modelo, a conceituação de gene fica clara pois o operador genético crossover pode operar sobre seções do segmento sonoro, embora não seja dada uma classificação atribuindo características genéticas para as partes do segmento sonoro.

Uma possibilidade, ainda que um tanto aproximada, seria considerar como genes do cromossomo *loudness*, as seções ADSR (*attack, decay, sustain, release*) da classificação de envelope de amplitude utilizada na síntese *wavetable*. Como genes do cromossomo espectro, poderia se considerar as regiões comumente atribuídas aos graves, médios e agudos. Pode-se a princípio considerar um gene sonoro como sendo uma seção de uma curva psicoacústica porém é ainda difícil estabelecer as fronteiras entre os genes (onde um gene termina e o outro começa), apesar de existirem seções do segmento sonoro que concentrem informação perceptual significativa e específica.

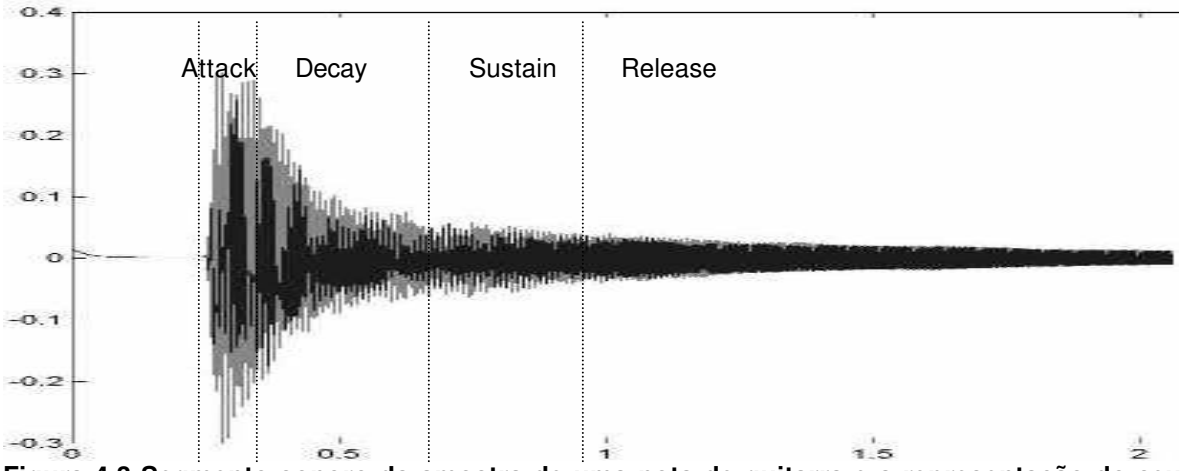


Figura 4.3 Segmento sonoro da amostra de uma nota de guitarra e a representação do seu envelope de amplitude ADSR (*attack, decay, steady-state, release*) como proposta de representação de genes da curva de loudness, onde esta representaria então um cromossomo.

Neste trabalho nos limitamos a classificar as curvas psicoacústicas de *loudness*, *pitch* e espectro como o genótipo do som, e cada curva como um gene, deixando para futuras pesquisas a classificação e distinção entre cromossomos e genes nas curvas psicoacústicas. Note que o método de síntese evolutiva proposto no capítulo 2, não faz distinção entre segmento sonoro, genótipo e cromossomo. No capitulo 3, com a inclusão das curvas psicoacústicas, o método da síntese evolutiva passa a diferenciar o indivíduo (como segmento sonoro) do seu genótipo (as três curvas psicoacústicas) e gene (cada curva psicoacústica).

Para simular a extração do genótipo do indivíduo, desenvolvemos a função do chamada lpe.m. Esta função é dada a seguir:

```
%   function lpe.m - Loudness, Pitch e Espectro
%   Calcula o genotipo que sao as curvas psicoacusticas do segmento sonoro
%   (3.3 Extraçao das curvas psicoacusticas dos individuos)

function [genotipo] = lpe(xi,fs,N)

% mapeamento da curva de fletcher-munson de 40dB
fm=(0.05+(400./(1:20000))).*(10.^((0.05+(400./(1:20000))))))/max((0.05+(400./
(1:20000))).*(10.^(-(0.05+(400./(1:20000))))));

% inicializa variaveis iniciais
Nt =length(xi);
if N>Nt
    xi(N)=0; % faz size(x)=N
end;
x=xi(1:N);
Loudness(N) =0;       % determina tamanho das curvas psicoacusticas
Pitch(N) =0;
Espectro(N) =0;

firstP =0; secondP =1;
f1 =0; f2 =0; f3 =0;
pico1 =0; pico2 =0; pico3 =0;

%Calculo da curva psicoacustica espectro
Espectro =fft(x);
Espectro =Espectro'/max(Espectro);

% Pre-processamento, normaliza e suprime espurios
x2 = subplus(x);        % descarta parte negativa
x2 = x2/max(x2);        % normaliza
x2 =x2-p4;               % elimina pontos abaixo de p4
x2 = subplus(x2);
x2 = x2/max(x2);

% Nivel de detecçao da janela
frameSize = round(fs*p2);       % determina o tamanho da janela
hframeSize = round(frameSize/2);    % metade do tamanho da janela
numberOfFrames = round(N/frameSize);
k1 = 1; k2 = frameSize;

for i=1:numberOfFrames,
   Frame = x2(k1:k2);
   frameMax = max(Frame);
   if frameMax ~= 0
        Frame = Frame/frameMax;
        Frame = Frame.^p1;               % eleva a potencia p1
        Frame = Frame/max(Frame);
   end;
   count = 1;
   for j=k1:k2,
       if Frame(count)<p3 x2(j) = 0; end; % descarta direto em x2 se pto de
Frame < p3
       count = count+1;
   end;
```

```
   k1 = k2 + 1;
   k2 = k2 + frameSize;
   if k2>N break; end;
end;

% Detecçao do maior elemento da janela - detecçao de picos
i=1;
while i<N-2,
   i=i+1;
   if x2(i-1) < x2(i) & x2(i) > x2(i+1) % detecta um pico
      firstP = secondP;                  % posiçao do primeiro pico
      secondP  = i;                           % posiçao do segundo pico
      f1 = f2; f2 = f3; f3 = fs/(secondP - firstP); % carrega freqs
      if f1 ~=0
         if f3>(f2*p5) | f3<(f2/p6)       % maxima variacao de pitch
            f3=f2;
         end;
      end;
      pico1 = pico2; pico2 = pico3; pico3 = x2(i);   % carrega pico
  end;
      frequencia = (f1+f2+f3)/3;  % media das tres ultimas frequencias
      pico       = (pico1+pico2+pico3)/3; % media dos 3 ultimos picos
      if pico > p7
           Pitch(i) =frequencia;      % calculo da curva de Pitch
      else Pitch(i) = 0;
      end;
      %Loudness(i) = pico;
      if Pitch(i)>=1 && Pitch(i)<20000;
         Loudness(i)  =pico.*fm(round(Pitch(i)));    %  mapeamento  fletcher-
munson
      else Loudness(i) =0;
      end;
end;

genotipo.loudness = Loudness;
genotipo.pitch    = Pitch;
     genotipo.espectro = Espectro;
```

A função lpe.m recebe como entrada o segmento sonoro, sua taxa de amostragem e o comprimento desejado (que pode ser maior ou menor que o comprimento original) e retorna uma estrutura chamada genótipo que contêm as três curvas psicoacústicas. As curvas de *loudness* e *pitch* são calculadas ao mesmo tempo, pela detecção de picos no segmento sonoro. A curva de espectro é calculada através da transformada discreta de Fourier do segmento, dada pela função fft.m. Para segmentos sonoro em ciclo, que podem ser considerados como um sinal periódico, a curva de espectro mostra as componentes em freqüência deste segmento sonoro. Para sons não periódicos, a curva de espectro fornece a distribuição espectral das componentes em freqüência. Três resultados típicos do cálculo do genótipo através dessa função são dados abaixo:

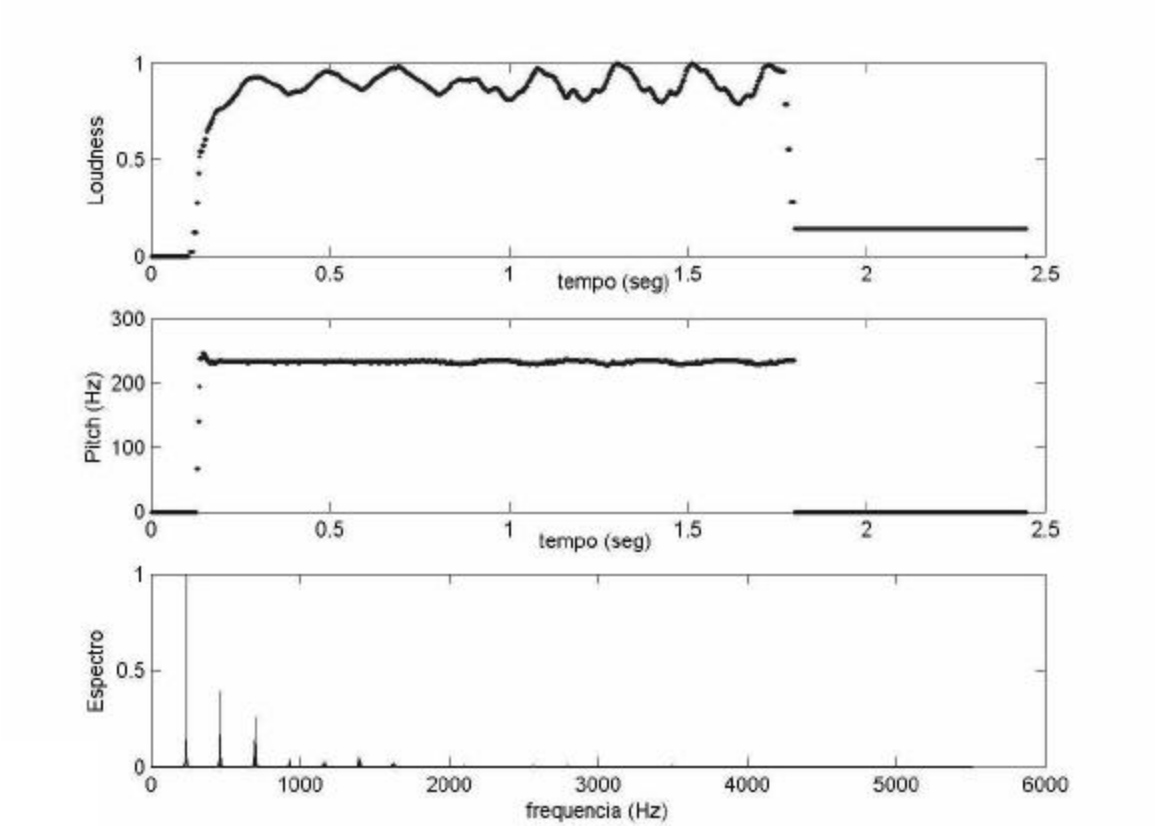


Figura 4.4 As curvas psicoacústicas que compõem o genótipo do segmento sonoro de uma nota de saxofone alto.

Este indivíduo representa o som emitido por uma nota de saxofone alto. O som é aproximadamente periódico, com uma leve variação de intensidade ao final, descrito pela curva de *loudness*. Nota-se que a curva de *pitch* permanece aproximadamente constante, uma vez que as componentes em freqüência não variam significativamente ao longo do tempo. A curva de espectro define claramente três componentes em freqüência deste indivíduo, sendo que a primeira componente é a fundamental do som, que é representada pela curva de *pitch* (observe que a freqüência é a mesma). A partir de 1000Hz tem-se na curva de espectro alguns componentes de pouca intensidade que praticamente não aparecem no gráfico acima. Estes são provavelmente resultados da "quase-periodicidade" do segmento sonoro.

No próximo exemplo veremos o resultado do cálculo do genótipo para um indivíduo que é um segmento sonoro em ciclo, ou seja, estritamente periódico

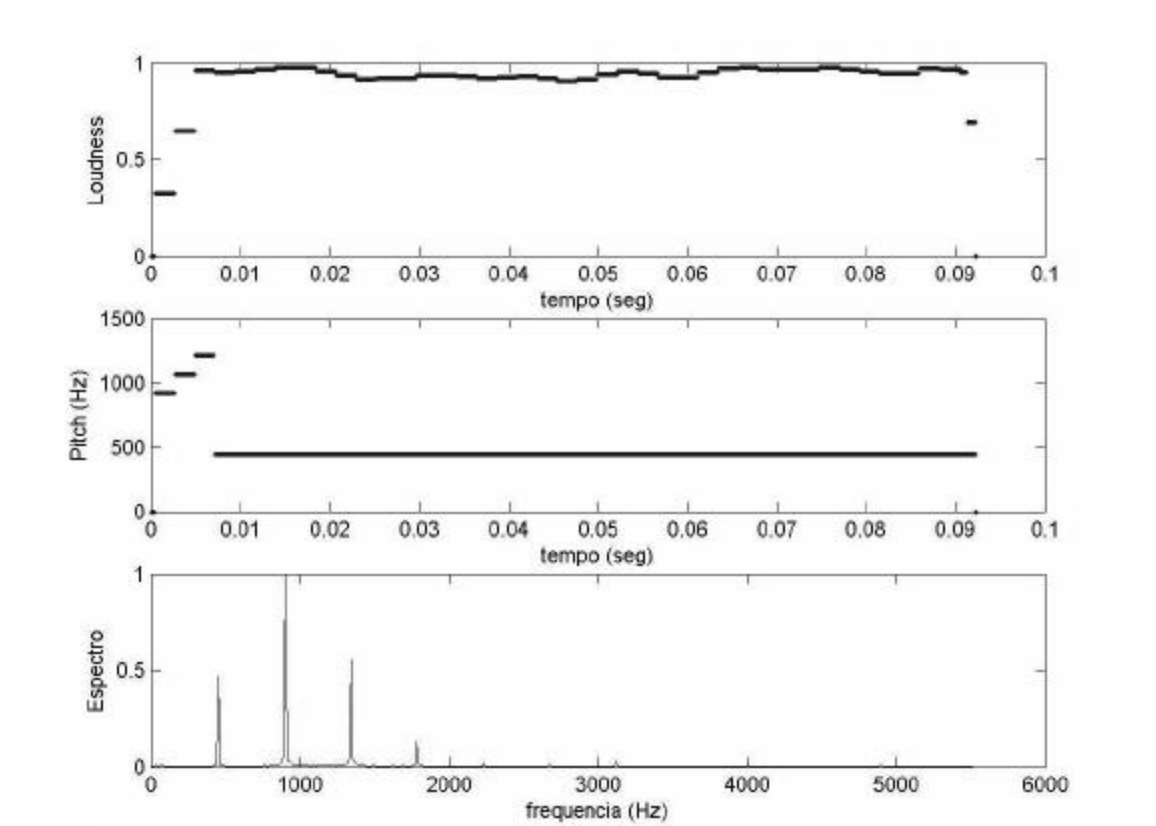


Figura 4.5 O genótipo de um segmento sonoro, em ciclo, do som de uma voz feminina.

Observe que a curva de *pitch* é mais linear que a anterior. A curva de espectro também define mais claramente três componentes em freqüência. É interessante notar que, neste caso particular, a fundamental não é a componente em freqüência de maior intensidade no espectro do segmento sonoro.

Devido a natureza do algoritmo de detecção do *pitch* e *loudness*, dado na função lpe.m descrita acima, existe um atraso de detecção de aproximadamente 5ms onde a curva de *pitch* apresenta valores incorretos. Observe que no início da curva de *pitch* tem-se a detecção de freqüências das componentes superiores à fundamental, mostradas pelas curva de espectro. Também nesse mesmo período a curva de *loudness* "caminha" para a detecção do valor correto de *loudness*.

O próximo exemplo mostra o cálculo do genótipo para um segmento sonoro estritamente não periódico; o ruído branco.

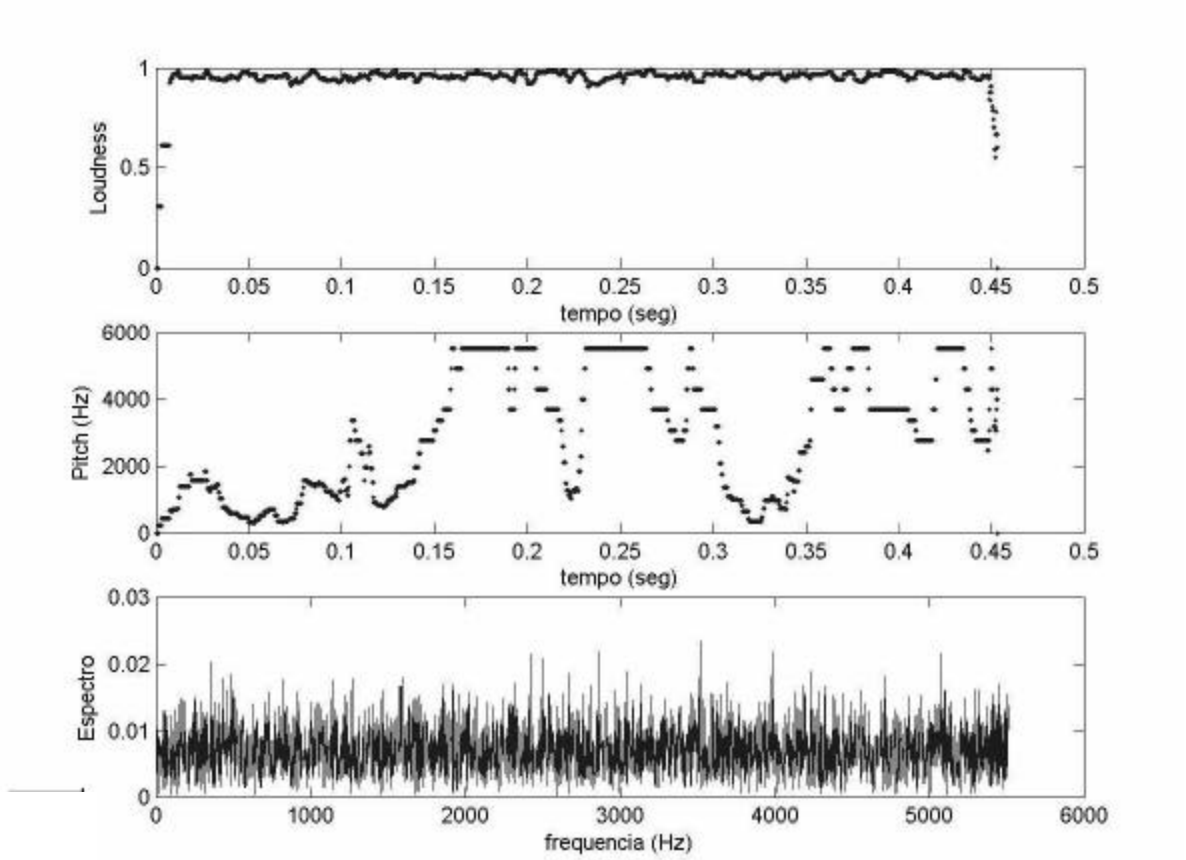


Figura 4.6 O genótipo de um segmento não-periódico (ruído branco).

Este indivíduo apresenta uma curva de espectro em forma de distribuição espectral. A curva de *loudness* se mantêm aproximadamente constante e a curva de *pitch* apresenta muita variação. Sendo rigorosos, a curva de *pitch* deveria estar em zero pois o ruído não define qualquer pitch, no entanto o algoritmo utilizado para a detecção de *pitch* tenta encontrar a fundamental do segmento sonoro o que faz com que seja apresentada uma curva de *pitch* que de fato não existe.

Finalizando, o exemplo abaixo mostra o genótipo de um indivíduo representado pelo som de 5 notas cantadas por voz feminina, separadas em aproximadamente meio tom cada.

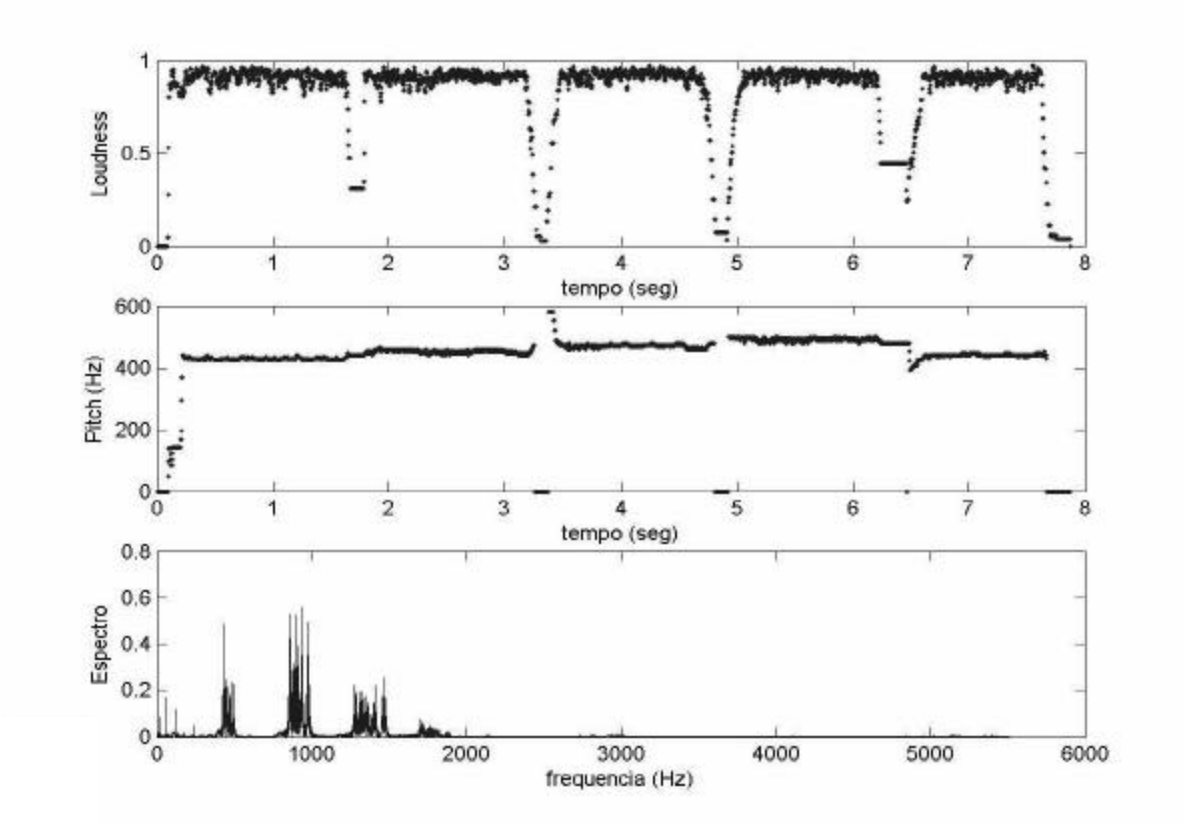


Figura 4.7 O genótipo da amostra de um trecho de voz cantado (5 notas próximas).

Observe que a curva de espectro apresenta claramente uma distribuição espectral das componentes em freqüência do segmento sonoro. A curva de *pitch* mostra a variação da fundamental entre as notas contadas, que são próximas entre si.

4.3 A operação genética sobre o indivíduo

Foi visto que os operadores genéticos crossover e mutação constituem o processo de reprodução na síntese evolutiva. No capitulo 2 estes operadores genéticos modificam o indivíduo alterando o seu segmento sonoro. Para o operador crossover, tem-se a necessidade de dois indivíduos, pois no método da síntese evolutiva o operador crossover permuta características sonoras entre o r-ésimo indivíduo da população e o melhor indivíduo, escolhido pelo processo de seleção.

Para simular as operações genéticas sobre o indivíduo, criou-se a função opg.m. Esta recebe como entrada um indivíduo da população, o melhor indivíduo, a taxa de crossover alfa e a taxa de mutação beta. A saída é o segmento sonoro modificado pelos operadores genéticos. Esta função é mostrada a seguir:

```
%   function opg.m - Operaçao Genetica sobre o segmento sonoro
%   Calcula o crossover e a mutaçao sobre o segmento sonoro
%   (4.3 A operação genética sobre o indivíduo)
%

function [novosegmento] = opg(segmento,melhorseg,alfa,beta)

Ns =length(segmento);
Nm =length(melhorseg);

k1 = round(rand(1,1)*Ns);   % gera k1 e k2 e garante que k1 < k2 < Ns
k2 = round(rand(1,1)*Ns);
if k1>k2 k=k1; k1=k2; k2=k; end;

if Nm>Ns  % separa a parte do melhor segmento de tamanho igual a do segmento
    melhor = melhorseg(1:Ns);
elseif Ns>Nm
    melhor = [melhorseg;segmento(Nm+1:Ns)];
else melhor = melhorseg;
end;

novosegmento = ((1-alfa)*segmento)+(alfa*melhor);   % crossover

mut = 1 - (rand(1,Ns).*beta);   % vetor mutaçao
mut=1-(mut.*beta);  % dimensionado pela taxa de mutaçao beta

novosegmento = novosegmento.*mut';   % novo individuo, apos mutaçao
```

Se o melhor indivíduo tem tamanho diferente do indivíduo da população que está sendo modificado, a operação genética é feita no tamanho do menor indivíduo porém a saída da função (indivíduo modificado) tem o mesmo tamanho do indivíduo de entrada, de N pontos. Isto porque o indivíduo de saída (modificado) irá substituir o indivíduo de entrada na próxima geração da população. A primeira operação genética é a de crossover, onde dois números k1 e k2 são escolhidos aleatoriamente, onde 0 < k1 < k2 < N. Estes definem a seção do segmento sonoro onde o crossover irá ocorrer, ou seja, a região do indivíduo da população que será misturada com a parte correspondente do melhor indivíduo. A mistura ocorre de acordo com a taxa de crossover alfa, que varia entre 0 a 100%. Dentro da seção definida aleatoriamente por k1 e k2, o operador crossover faz uma mistura entre os pontos das seções dos indivíduos e o resultado é copiado na seção do indivíduo da população. O melhor indivíduo permanece inalterado.

Em seguida faz-se a operação de mutação sobre o indivíduo. A mutação é definida por uma taxa de mutação, beta, que também varia entre 0 e 100%. Para isso, cria-se um vetor de números aleatórios de N pontos, chamado de vetor mutação, que é multiplicado pelo indivíduo. Os elementos desse vetor estão limitados por beta. Se beta=0 todos os elementos do vetor mutação são unitários e o resultado da multiplicação deste com o indivíduo é o próprio indivíduo. Se beta=1, todos os elementos do vetor mutação variam entre 0 e 1 e o resultado da multiplicação do vetor mutação pelo indivíduo é um vetor de elementos aleatórios. O resultado é o segmento sonoro operado, ou seja, o indivíduo modificado pelos operadores genéticos de acordo com as taxas alfa e beta, saída da função.

Para exemplificar a função opg.m, dois resultados típicos são mostrados a seguir. As figuras apresentam três curvas. A superior é o segmento sonoro representado pelo indivíduo original, antes da modificação feita pelos operadores genéticos. A segunda curva (do meio) representa o melhor indivíduo da população. A ultima curva é o novo indivíduo, o resultado da operação genética sobre o segmento sonoro original pelo melhor indivíduo. As operações genéticas são determinadas pelas taxas de crossover (alfa) e mutação (beta). Note porém que os resultados não se repetem para os mesmos parâmetros de alfa e beta uma vez que a seção de operação do crossover k1 e k2 é escolhida aleatoriamente toda vez que esta função é executada.

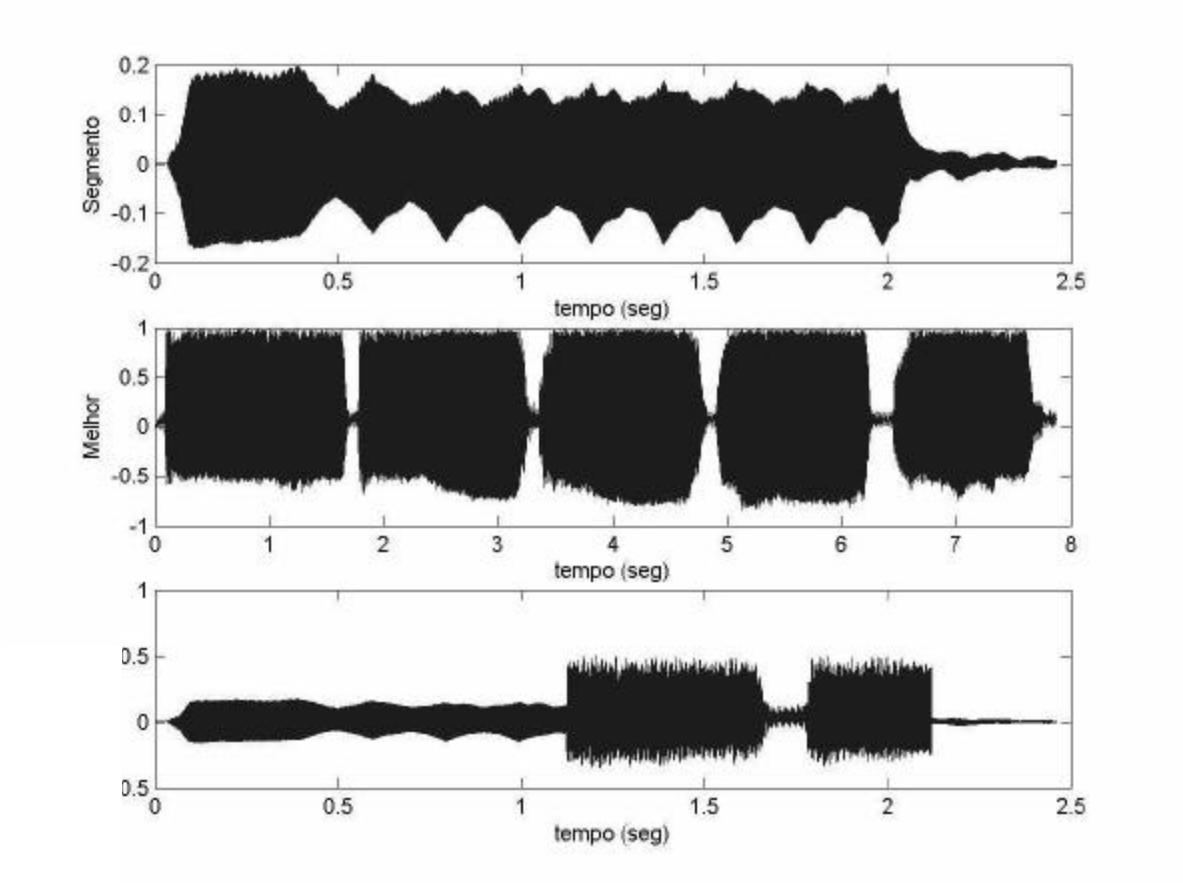


Figura 48 Operação genética sobre segmento sonoro. Indivíduo é nota de flauta, melhor indivíduo é voz feminina. As taxas de operação genética são: alfa=50% e beta=10%.

No exemplo a seguir temos duas senóides como indivíduo. O melhor indivíduo é uma senóide de 1000Hz e o indivíduo a ser manipulado é uma senóide de 440Hz. A taxa de crossover é de 90% o que faz com que a seção do segmento tenha quase que uma substituição pela seção correspondente do melhor indivíduo. A taxa de mutação é 50% o que insere um ruído ao segmento sonoro que não existia anteriormente, uma vez que os dois indivíduos são senóides.

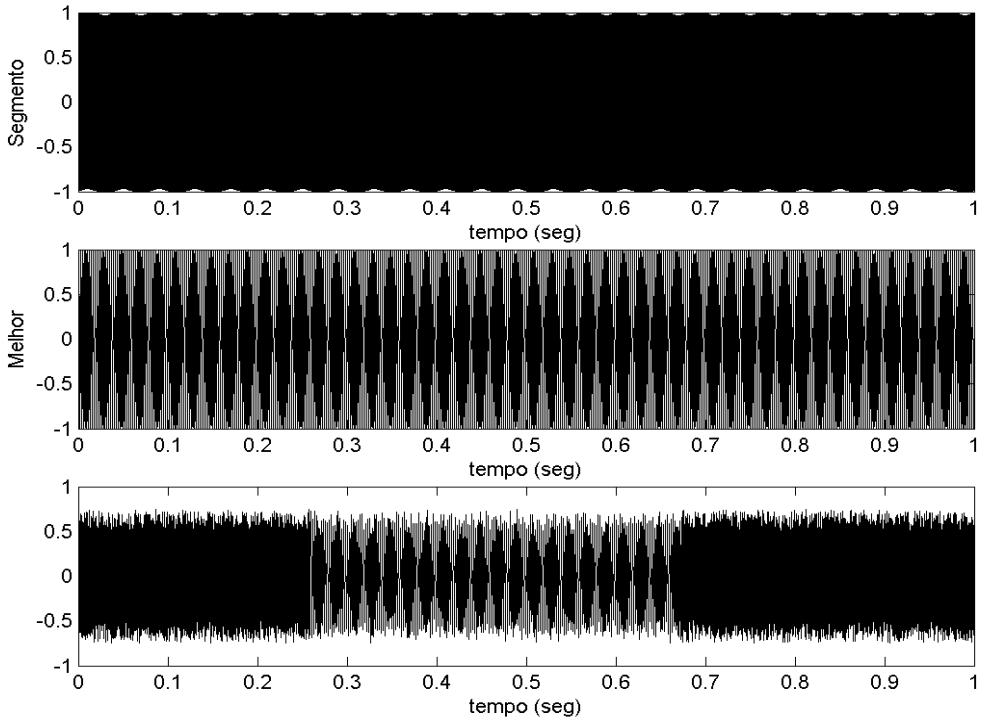


Figura 4.9 Operação genética sobre segmento sonoro. Indivíduo é uma senoide de 1KHz, melhor indivíduo é uma senoide de 440Hz. As taxas de operação genética são: alfa=90% e beta=50%.

4.4 A operação genética sobre as curvas psicoacústicas do indivíduo

A inclusão das curvas psicoacústicas como genótipo do som, no modelo de síntese evolutiva do capítulo 3, fez com que as operações genéticas passassem a ser executadas sobres as curvas psicoacústicas, ao invés de diretamente sobre os segmentos sonoros. Isto porque cada curva psicoacústica representa uma característica perceptual do segmento sonoro e portanto a sua manipulação por operadores genéticos leva a resultados sonoros que são percentualmente mais condizentes aos propósitos de uma síntese sonora.

Conforme dissemos anteriormente, consideramos neste trabalho cada curva psicoacústica como o equivalente sonoro a um gene biológico, que carrega uma informação estrutural da percepção sonora deste indivíduo, e é indivisível durante a operação genética. Neste segundo modelo da síntese evolutiva tivemos que abandonar temporariamente a noção de seccionamento do segmento na operação de crossover, que introduzimos no capítulo 2, e tratar cada curva como se fosse um gene sonoro, uma seção perceptual atômica. Futuros desenvolvimentos do modelo de síntese evolutiva podem voltar a utilizar o seccionamento, desta vez para a operação de crossover nas curvas psicoacústica, passando então a considera-las como correspondente a cromossomos sonoros, e as suas seções como genes.

Desenvolvemos uma função para a simulação das operações genéticas de crossover e mutação sobre as curvas psicoacústicas do indivíduo, a função cog.m. Esta função recebe como entrada duas variáveis, do tipo estrutura, e as taxas de crossover e mutação, alfa e beta. A primeira estrutura corresponde ao genótipo do indivíduo a ser modificado, que contêm as suas curvas psicoacústicas de *loudness*, *pitch* e espectro. A outra estrutura é o genótipo do melhor indivíduo da população. A função é dada a seguir:

```
%   function cog.m - Calculo das Operaçoes Geneticas
%   Calcula o crossover e mutaçao sobre o genotipo
%   (4.4 A operação genética sobre as curvas psicoacústicas do indivíduo)
%   g_ind_pop:  struct do genotipo de um individuo na populaçao
%   g_melhor_ind:  struct do genotipo do melhor individuo da populaçao
%   alfa: taxa de crossover  beta: taxa de mutaçao
%    g_novo_ind:   struct do genotipo do novo individuo na populaçao (apos
operaçoes geneticas)

function [g_novo_ind] = cog(g_ind_pop,g_melhor_ind,alfa,beta)
% inicializa variaveis
l1 =g_ind_pop.loudness;
p1 =g_ind_pop.pitch;
e1 =g_ind_pop.espectro;
l2 =g_melhor_ind.loudness;
p2 =g_melhor_ind.pitch;
e2 =g_melhor_ind.espectro;
N=length(l1);
% crossover
r =3;% round(rand(1,1)*2+1);    % gera numero aleatorio 1, 2 ou 3
switch r
    case 1
l3 =((1-alfa)*l1)+(alfa*l2);
p3 = p1;
e3 =e1;
    case 2
l3 = l1;
p3 =((1-alfa)*p1)+(alfa*p2);
e3 =e1;
    case 3
l3 = l1;
```

```
p3 = p1;
e3 =((1-alfa)*e1)+(alfa*e2);
end;
% mutaçao
ml =1-(rand(1,N).*beta);  % vetor mutaçao para o loudness
mp =1-(rand(1,N).*beta);  % vetor mutaçao para o pitch
me =1-(rand(1,N).*beta);  % vetor mutaçao para o espectro
g_novo_ind.loudness  =  l3.*ml; % novo loudness apos mutaçao
g_novo_ind.pitch     =  p3.*mp;  % novo pitch apos mutaçao
g_novo_ind.espectro  =  e3.*me;  % novo espectro apos mutação
```

Da mesma maneira que a função opg.m fazia para o segmento sonoro, a função cog.m faz a operação genética sobre o genótipo do indivíduo no trecho de mesmo tamanho entre o genótipo do indivíduo da população a ser modificado e o genótipo do melhor indivíduo. A operação de crossover escolhe aleatoriamente uma das três curvas psicoacústicas para manipular, porém sem seciona-la, como fazia a função opg.m. Esta curva psicoacústica é então misturada com a curva psicoacústica correspondente do genótipo do melhor indivíduo, à uma taxa de crossover alfa. Em seguida, a operação de mutação altera as três curvas psicoacústicas, pela sua multiplicação com três vetores de números aleatórios. Estes números são gerados em um intervalo determinado pela taxa de mutação beta.

Para ilustrar as operações genéticas de crossover e mutação sobre as curvas psicoacústicas do genótipo do indivíduo, utilizaremos os mesmos indivíduos cujos genótipos são mostrados nas figuras 4.4 (saxofone alto, nota contínua) 4.5 (voz feminina, segmento sonoro em ciclo) e 4.6 (ruído branco), de modo a tornar possível a comparação entre os resultados da modificação destes genótipos.

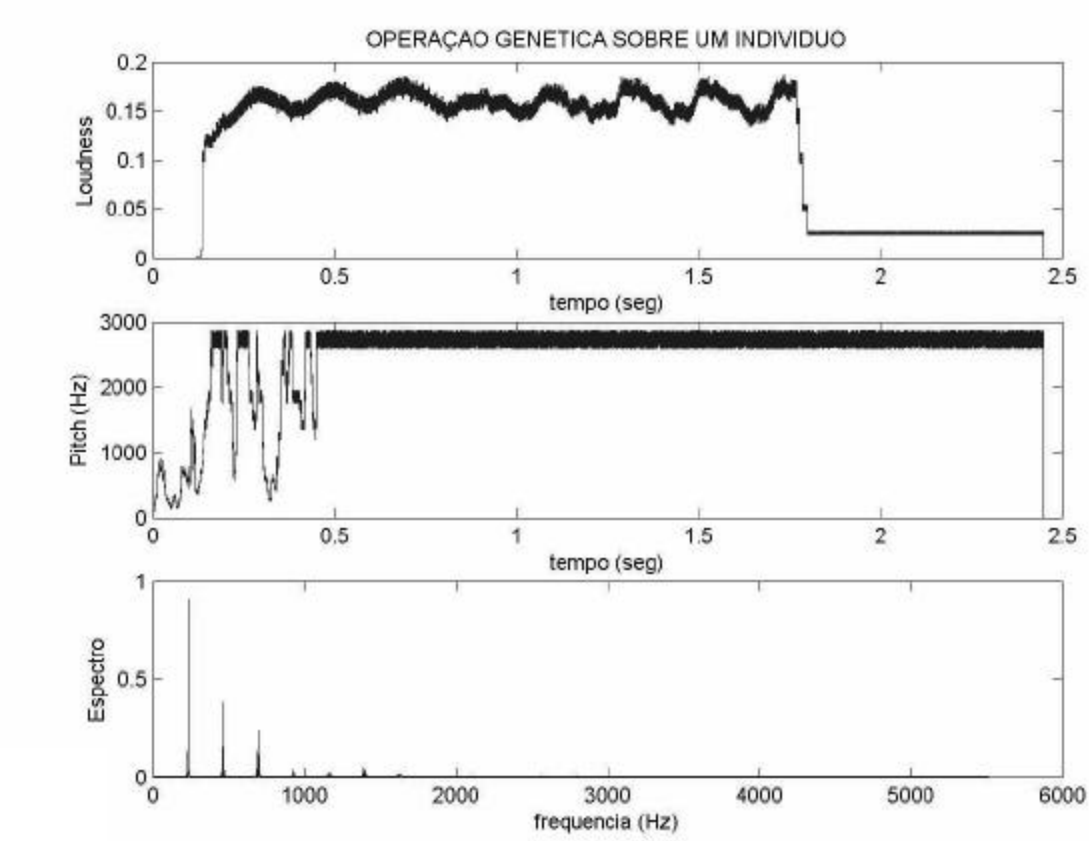


Figura 4.10 Exemplo de operação de crossover e mutação sobre o indivíduo da figura 4.4 (saxofone alto) e melhor indivíduo aquele dado na figura 4.6 (ruído branco), nas taxas alfa=50% e beta=10%.

Observa-se na figura anterior que a operação de crossover foi feita na curva de *pitch*, onde o genótipo do indivíduo (sax alto) tem a curva de *pitch* misturada com a curva de *pitch* do genótipo do melhor indivíduo (ruído branco). O efeito da operação mutação é mais visível na curva de *loudness*. Pode-se observar que o traçado da curva, que se tornou mais grosso que aquele observado na figura 4.4 (genótipo original). Isto ocorre devido à operação de mutação, dada pela multiplicação da curva de *loudness* original com o vetor mutação de *loudness*, à taxa beta=10%.

No próximo exemplo, temos o resultado das operações genéticas para o genótipo do indivíduo da figura 4.5 (voz feminina, segmento sonoro em ciclo) pelo genótipo do melhor indivíduo dado na figura 4.4 (nota contínua do saxofone alto). A operação de crossover ocorreu sobre a curva de espectro do indivíduo onde a curva de espectro modificada (dada na figura 4.11) apresenta uma nova componente em freqüência, da ordem de 250 Hz, que não fazia parte da curva de espectro do genótipo do indivíduo original (figuras 4.5) mas que fazia parte da curva de espectro do genótipo do melhor indivíduo (figura 4.4).

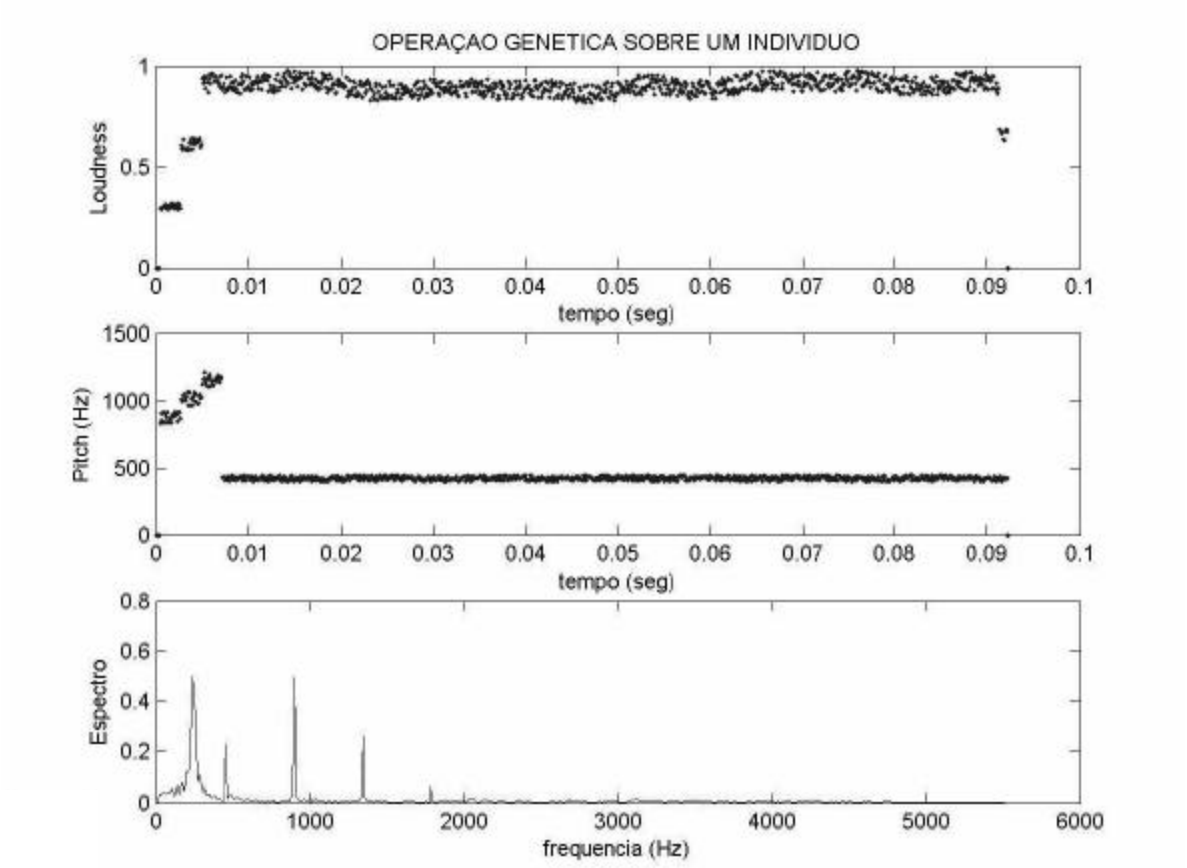


Figura 4.11 Exemplo de operação de crossover e mutação sobre o indivíduo da figura 4.5 (voz feminina em ciclo) e melhor indivíduo aquele dado na figura 4.4 (saxofone alto), nas taxas alfa=50% e beta=10%.

No exemplo a seguir utilizamos como indivíduo o segmento sonoro de números aleatórios, correspondente à ruído branco (figura 4.6) e melhor indivíduo a voz feminina em ciclo (figura 4.5). A operação de crossover foi feita sobre a curva de *pitch* o que é evidenciado pela compressão de sua variação original observada na figura 4.6.

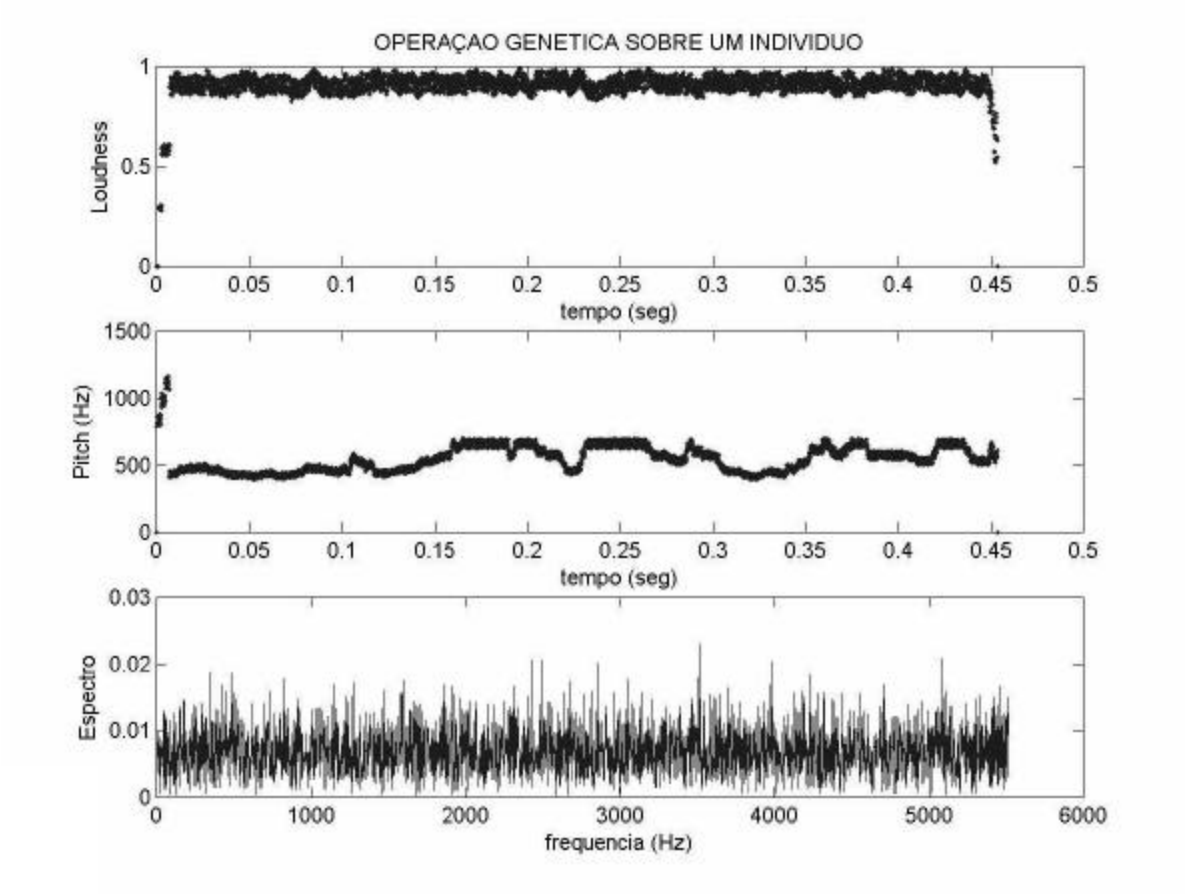


Figura 4.12 Exemplo de operação de crossover e mutação sobre o indivíduo da figura 4.6 (ruído branco) e melhor indivíduo aquele dado na figura 4.5 (voz feminina em ciclo), nas taxas alfa=95% e beta=10%.

4.5 A medida da distância entre o indivíduo e o conjunto alvo

Nos itens 4.3 e 4.4 mostramos os resultados experimentais das operações genéticas, respectivamente sobre o indivíduo (segmento sonoro) e sobre o genótipo (curvas psicoacústicas). Dissemos que as operações genéticas são as responsáveis pelo processo de reprodução no método da síntese evolutiva. Nesta seção iremos mostrar os resultados experimentais do outro processo que compõe o método da síntese evolutiva, o processo de seleção. Na síntese evolutiva a seleção dos indivíduos é dada pelo grau de adequação deste indivíduo em comparação aos indivíduos do conjunto alvo. O grau de adequação do indivíduo é dado pela medida da distância de Hausdorff. No capítulo 2 esta distância é medida entre os indivíduos (segmentos sonoros). No capítulo 3 esta distância é calculada entre os genótipos dos indivíduos, compostos pelas curvas psicoacústicas do segmento sonoro. Pode-se dizer que a medida da distância é o ponto mais importante do método de síntese evolutiva pois é através desta medida que toda a evolução é conduzida. A medida da distância determina qual será melhor indivíduo de cada geração, que não só reproduz com todos os outros indivíduos da população mas também é o som sintetizado.

Para ambos os modelos de síntese evolutiva, dos capítulos 2 e 3, a medida é dada ela distância entre vetores de N pontos (segmentos sonoros ou curvas psicoacústicas). Estudamos quatro medidas de distância vetorial, a saber: 1) d1: distância euclidiana sem peso, 2) d2: distância euclidiana com peso decrescente, 3) d3: distância euclidiana diferencial sem peso 4) d4: distância euclidiana diferencial com peso decrescente. A primeira medida da distância, d1, é a distância dada pela equação 2.3. A segunda medida d2 é a versão com peso. Utilizamos pesos com valores decrescentes (N, N-1, N-2, … 3, 2, 1) afim de valorizar a informação contida no inicio do vetor. A terceira

distância d3 é a mesma métrica da d1 porém para o diferencial dos vetores, para medir a variação do vetor, ao invés de sua área. A quarta medida d4 é a d3 acrescida de peso decrescente, como em d2.

Criamos a função mds.m para o calculo das distâncias entre os segmentos sonoros. Esta função tem como entrada dois segmentos sonoros de tamanhos N1 e N2. A distância é medida sobre o tamanho de interseção, ou seja, o menor valor de N. A função retorna na saída as quatro distâncias entre os dois segmentos. O algoritmo da função é dada abaixo:

```
%   function mds.m - Medida da distancia entre segmentos
%   Calcula as 4 distancia entre dois segmentos:
% (1) distância euclidiana (L2) sem peso, (2) L2 com peso (decrescente),
% (3) diferencial sem peso e (4) diferencial com peso decrescente.
%   (4.5 A medida da distância do indivíduo ao conjunto alvo)

function [d1,d2,d3,d4] = mds(segmento1,segmento2)

% inicializa variaveis iniciais
N1=length(segmento1);   % tamanho do segmento 1
N2=length(segmento2);   % tamanho do segmento 2
if N1>N2    % calcula o tamanho de segmento para a medida da distancia
    N =N2;
else
    N =N1;
end;

% distancias
d1 =sqrt(sum(( segmento1(1:N)-segmento2(1:N)).^2))/N;
d2 =sqrt(sum(((segmento1(1:N)-segmento2(1:N)).*(N:-1:1)').^2))/sum(1:N);
d3 =sqrt(sum((diff( segmento1(1:N))-diff(segmento2(1:N))).^2))/(N-1);
d4              =sqrt(sum(((diff(segmento1(1:N))-diff(segmento2(1:N))).*(N:-
1:2)').^2))/sum(1:N-1);
```

As distâncias estão normalizadas, afim de se estabelecer um valor relativo entre elas. Para compararmos as medidas de distâncias vetoriais, criamos uma população de 36 indivíduos, que são amostras de segmentos sonoros, em 16 bits e taxa de amostragem de 11025KHz, dados em formato WAV. A letra **c** ao lado do nome do som significa que o segmento é dado em ciclo, e a letra **a** significa o ataque do segmento. Por exemplo, “baixo a.wav” é o ataque de “baixo.wav”, e “baixo c.wav” é o seu ciclo.

1 'baixo a.wav'	**14 'sax baritono.wav'**	**27 'trompete c.wav'**
2 'baixo c.wav'	**15 'sax c.wav'**	**28 'trompete.wav'**
3 'baixo.wav'	**16 'sax soprano.wav'**	**29 'voz fem a.wav'**
4 'cello'	**17 'sax tenor.wav'**	**30 'voz fem c.wav'**
5 'flauta a.wav'	**18 'seno 1K.wav'**	**31 'voz fem 3 notas.wav'**
6 'flauta c.wav'	**19 'seno 440Hz.wav'**	**32 'voz fem cromatica.wav'**
7 'flauta.wav'	**20 'seno slide.wav'**	**33 'voz masc port.wav'**
8 'guitar a.wav'	**21 'senos 110 a 880Hz.wav'**	**34 'voz masc 3 notas.wav'**
9 'guitar c.wav'	**22 'senos 3 notas.wav'**	**35 'voz masc escala.wav'**
10 'guitarra.wav'	**23 'strings.wav'**	**36 'ruido.wav'**
11 'pizzicato.wav'	**24 'synth bells.wav'**	
12 'sax a.wav'	**25 'synth voice.wav'**	
13 'sax alto.wav'	**26 'trompete a.wav'**	

Figura 4.13 População de indivíduos, composta por 36 segmentos de som no padrão 16 bits 11025KHz.

Para exemplificar as medidas de distância entre segmentos sonoros, consideramos como referência o 18º indivíduo (seno 1KHz) da população de 36 indivíduos dada na figura 4.13. O resultado da medida das quatro distâncias entre os indivíduos da população e o indivíduo de referência é mostrado no gráfico abaixo:

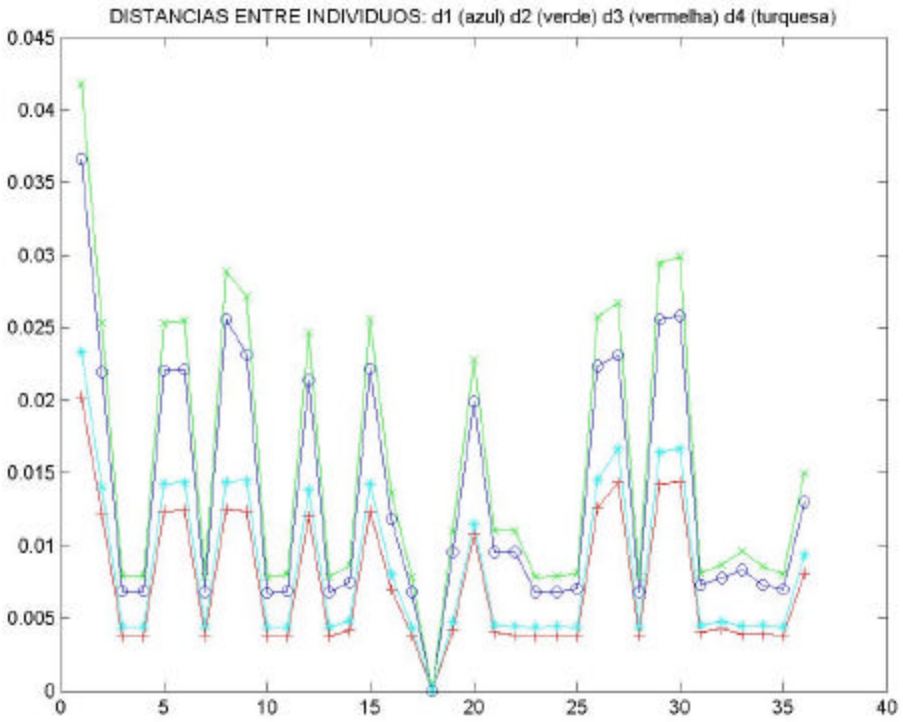


Figura 4.14 Distâncias entre 36 indivíduos na população em relação ao 18º indivíduo (seno 1KHz). A distância d1 (euclidiana sem peso) tem pontos em "o", a d2 (euclidiana com peso) pontos em "x", a d3 (diferencial sem peso) pontos em "+" e a d4 (diferencial com peso) pontos em " * ".

Observa-se que todas as medidas de distância possuem aproximadamente o mesmo formato de variação, assim não se demonstrou existir uma diferença significativa quanto ao tipo de distância utilizado no método da síntese evolutiva apresentado no capítulo 2. Por simplicidade, optou-se por utilizar a distância d1 (euclidiana sem peso) como critério de medida da distância de Hausdorff para a seleção do melhor indivíduo na síntese evolutiva.

4.6 A medida da distância entre as curvas psicoacústicas do indivíduo

No capítulo 3, o critério de adequação do indivíduo (segmento sonoro) passou a ser dado pela distância entre as suas curvas psicoacústicas de *loudness*, *pitch* e espectro, extraídas deste segmento sonoro. Diferente do modelo de síntese evolutiva do capitulo 2, onde a medida da distância é feita entre segmentos sonoros, no capítulo 3 a medida da distância é feita entre os genótipos dos indivíduos, compostos pelas suas curvas psicoacústicas. Por simplicidade, consideramos nesse trabalho que a medida da distância entre genótipos é dada pela média aritmética das distâncias das três curvas psicoacústicas, ou seja: $d = (dl + dp + ds)/3$.

Da mesma forma que fizemos no item 4.5, foram aqui testados os quatro tipos de distâncias entre vetores: d1: euclidiana sem peso, d2: euclidiana com peso decrescente, d3: diferencial sem peso e d4: diferencial com peso decrescente.

Para a análise comparativa entre os quatro tipos de distâncias, criamos a função mdg.m. Similar à função mds.m, esta função recebe de entrada o genótipo de dois indivíduos e retorna na saída as 4 distâncias entre os dois genótipos. Esta função é dada abaixo:

```
%   function mdg.m - Medida da distancia entre genotipos
%   Calcula as 4 distancia entre dois genotipos:
% (1) distância euclidiana (L2) sem peso, (2) L2 com peso (decrescente),
% (3) diferencial sem peso e (4) diferencial com peso decrescente.
%   (4.6 A medida da distância das curvas psicoacústicas do indivíduo)

function [d1,d2,d3,d4] = mdg(g_ind_pop,g_melhor_ind)

% inicializa variaveis
l1 =g_ind_pop.loudness;
p1 =g_ind_pop.pitch;
e1 =g_ind_pop.espectro;

l2 =g_melhor_ind.loudness;
p2 =g_melhor_ind.pitch;
e2 =g_melhor_ind.espectro;

N1=length(l1);   % tamanho do segmento 1 g_ind_pop
N2=length(l2);   % tamanho do segmento 2 g_melhor_ind

if N1 > N2    % calcula o tamanho de segmento para a medida da distancia
    N = N2;
else
    N = N1;
end;

% distancias loudness
dl1 =sqrt(sum(( l1(1:N)-l2(1:N)).^2))/N;
dl2 =sqrt(sum(((l1(1:N)-l2(1:N)).*(N:-1:1)).^2))/sum(1:N);
dl3 =sqrt(sum((diff(l1(1:N))-diff(l2(1:N))).^2))/(N-1);
dl4 =sqrt(sum(((diff(l1(1:N))-diff(l2(1:N))).*(N:-1:2)).^2))/sum(1:N-1);
% distancias pitch
dp1 =sqrt(sum(( p1(1:N)-p2(1:N)).^2))/N;
dp2 =sqrt(sum(((p1(1:N)-p2(1:N)).*(N:-1:1)).^2))/sum(1:N);
dp3 =sqrt(sum((diff(p1(1:N))-diff(p2(1:N))).^2))/(N-1);
dp4 =sqrt(sum(((diff(p1(1:N))-diff(p2(1:N))).*(N:-1:2)).^2))/sum(1:N-1);
% distancias espectro
de1 =abs(sqrt(sum(( e1(1:N)-e2(1:N)).^2))/N);
de2 =abs(sqrt(sum(((e1(1:N)-e2(1:N)).*(N:-1:1)).^2))/sum(1:N));
de3 =abs(sqrt(sum((diff(e1(1:N))-diff(e2(1:N))).^2))/(N-1));
```

```
de4  =abs(sqrt(sum(((diff(e1(1:N))-diff(e2(1:N))).*(N:-1:2)).^2))/sum(1:N-
1));
% distancias finais
d1=dl1+dp1+de1;
d2=dl2+dp2+de2;
d3=dl3+dp3+de3;
d4=dl4+dp4+de4;
```

Para exemplificar as medidas de distância entre genótipos também consideramos como referência para a medida das distâncias o 18º indivíduo (seno 1KHz) da população de 36 indivíduos dadas anteriormente. O resultado das distâncias entre os genótipos dos indivíduos desta população e o genótipo do indivíduo de referência é dado a seguir:

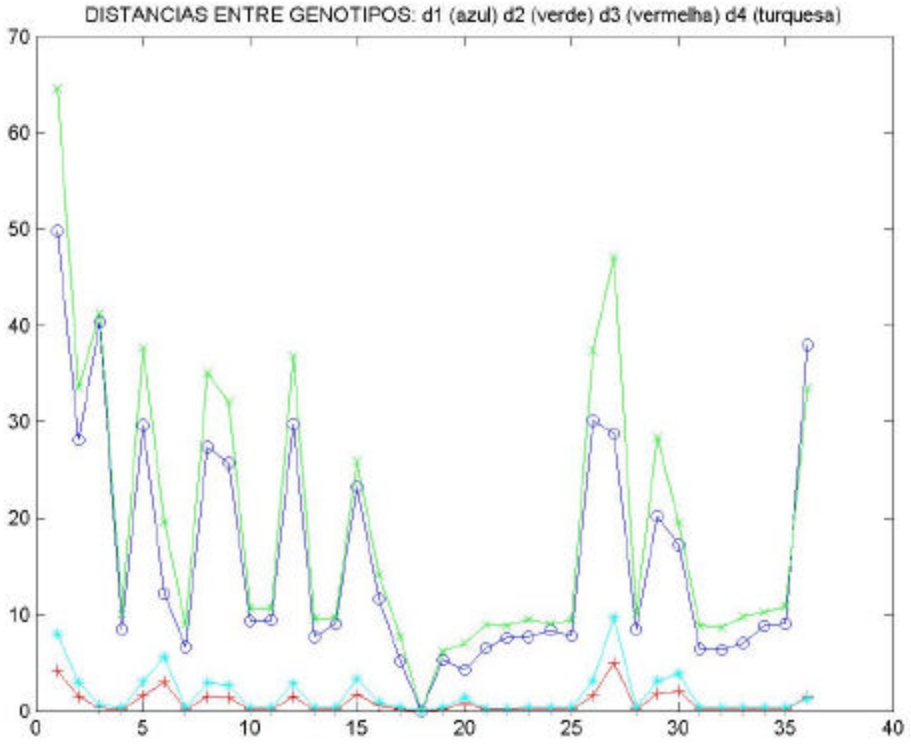


Figura 4.15 Distâncias entre os genótipos dos 36 indivíduos na população em relação ao genótipo do 18º indivíduo (seno 1KHz). A distância d1 (euclidiana sem peso) tem pontos em "o", a d2 (euclidiana com peso) pontos em "x", a d3 (diferencial sem peso) pontos em "+" e a d4 (diferencial com peso) pontos em " * ".

Observa-se que as distâncias diferenciais d3 e d4 foram menos sensíveis à variação de genótipos que as distâncias euclidianas d1 e d2. Também se apresentam algumas diferenças entre as distâncias euclidianas sem peso (d1) e com peso (d2). Observe a distância d1 e d2 entre os indivíduos 26 (trompete a.wav), 27 (trompete c.wav) e o indivíduo de referência, 18 (seno 1KHz.wav). Para a medida da distância d1 os indivíduos 18 e 26 são mais distantes que 27 e 18, ao passo que para a medida da distância d2, os indivíduos 26 e 18 são mais próximos que 27 e 18.

Levando em conta que 27 é um ciclo periódico do som de uma nota de trompete e 18 é uma senóide, chegamos a conclusão que a medida da distância entre genótipos d1 é mais correta em termos de percepção sonora.

Nos dois próximos exemplos, mostramos os resultados gráficos para a análise comparativa entre as quatro distâncias de genótipos dos indivíduos da população de 36 indivíduos. Na figura 4.16 o indivíduo de referência é o 36º (ruído branco). Na figura 4.17 o indivíduo de referência é o 30º (voz feminina). Apesar das diferenças entre as medidas de distância de genótipos,

especialmente evidenciadas na figura 4.16, optamos por simplicidade em utilizar a distância d1 (euclidiana sem peso) no que segue da experimentação deste trabalho.

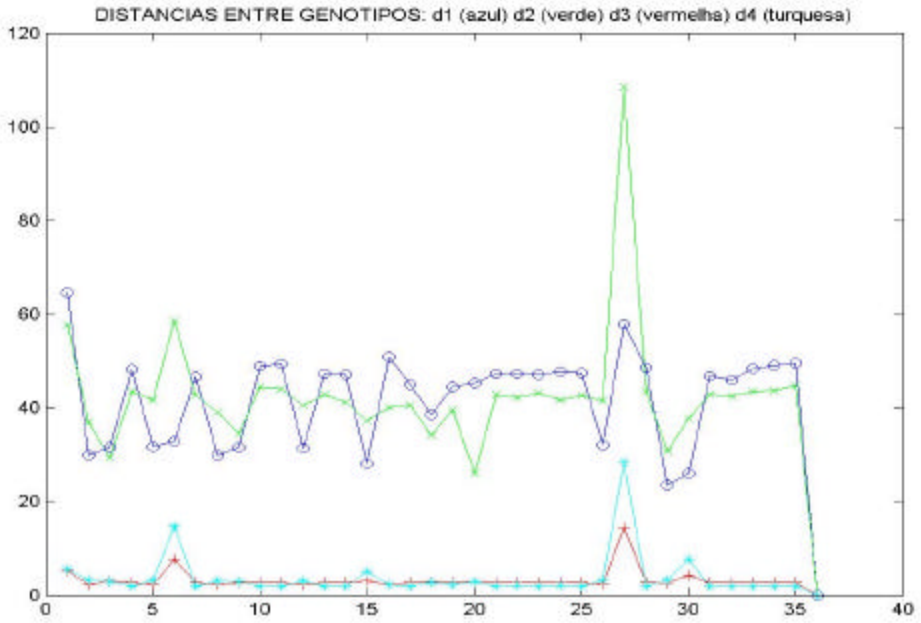


Figura 4.16 Distâncias entre os genótipos dos 36 indivíduos na população em relação ao genótipo do 36º indivíduo (ruído branco). A distância d1 (euclidiana sem peso) tem pontos em "o", a d2 (euclidiana com peso) pontos em "x", a d3 (diferencial sem peso) pontos em "+" e a d4 (diferencial com peso) pontos em " * ".

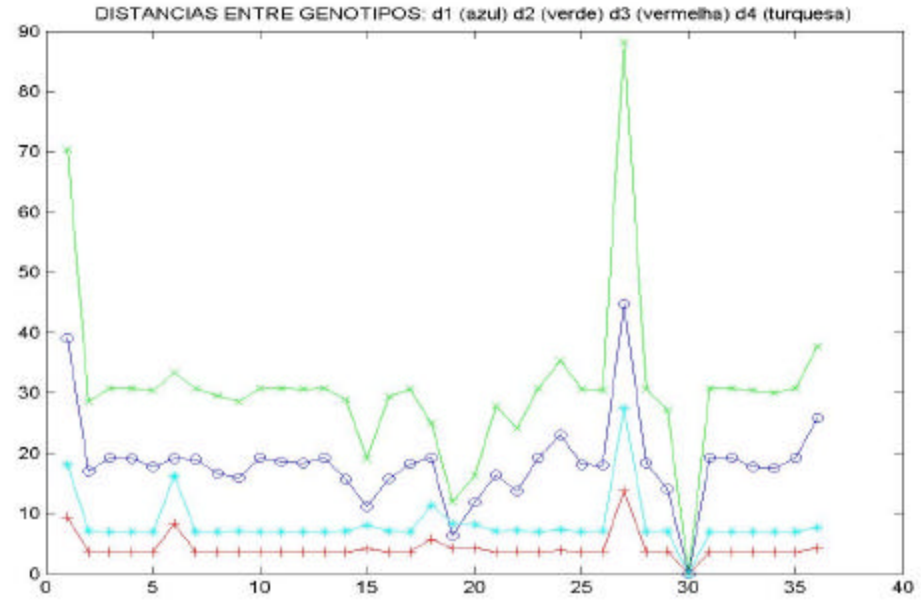


Figura 4.17 Distâncias entre os genótipos dos 36 indivíduos na população em relação ao genótipo do 30º indivíduo (voz feminina em ciclo). A distância d1 (euclidiana sem peso) tem pontos em "o", a d2 (euclidiana com peso) pontos em "x", a d3 (diferencial sem peso) pontos em "+" e a d4 (diferencial com peso) pontos em " * ".

4.7 A construção do novo indivíduo a partir da variação das curvas psicoacústicas

Vimos que o primeiro modelo da síntese evolutiva, apresentado no capítulo 2, faz as operações genéticas diretamente sobre o segmento sonoro, enquanto o modelo da síntese evolutiva utilizando curvas psicoacústicas como genótipo, apresentado no capítulo 3, faz as operações genéticas sobre as curvas psicoacústicas, e o segmento sonoro permanece inalterado.

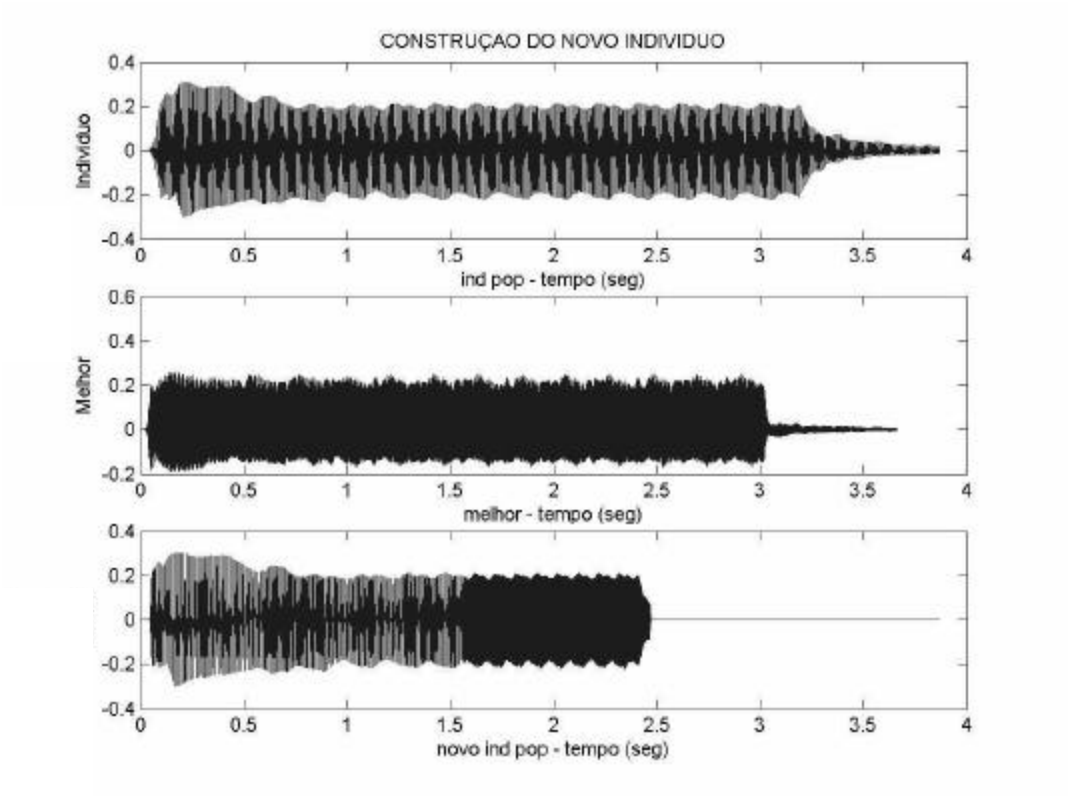
Desse modo, para a síntese evolutiva utilizando curvas psicoacústicas, é necessário construir o novo segmento sonoro a partir da variação de suas curvas psicoacústicas pelo processo de reprodução (dado pelos operadores genéticos).

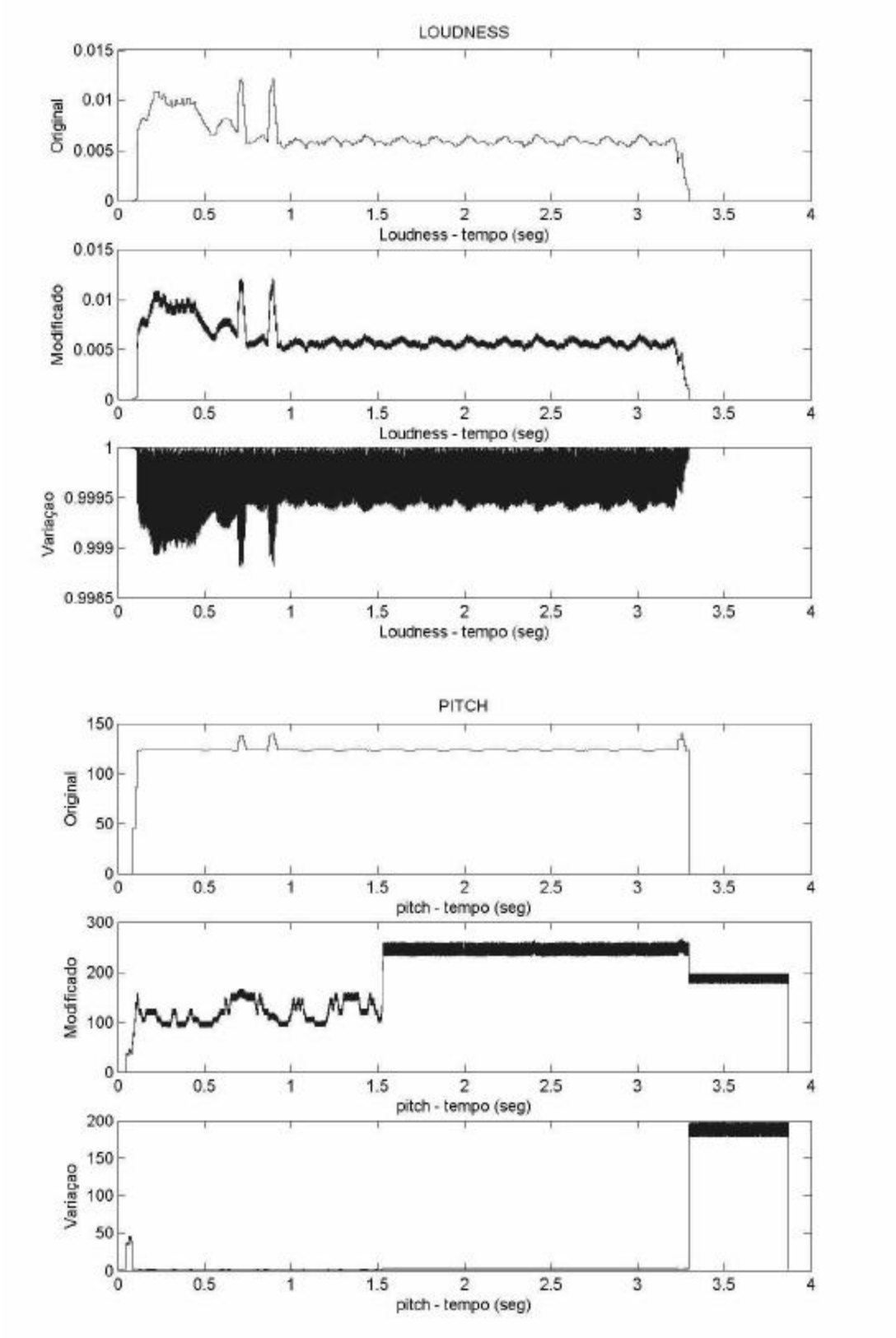
Criamos a função cns.m que calcula o novo segmento sonoro a partir da variação de suas curvas psicoacústicas. A função cns.m recebe como entrada o segmento sonoro original, do indivíduo a ser reconstruído e a variação de seu genótipo. A saída desta função é o novo segmento sonoro reconstruído. Esta função é dada a seguir:

```
%   function cns.m - Construçao do Novo Segmento
%   Constroi novo segmento sonoro a partir das curvas psicoacusticas
%    (4.7 A construção do novo indivíduo a partir da variação das curvas
psicoacústicas)
%   Obs: v_g_modificado sao as variaçoes das curvas psicoacusticas
%        verificar o espectro, a ifft tem q receber sequencia complexa e
%        simetrica

function [novo_segmento] = cns(seg,v_genotipo)

% inicializa variaveis
vloudness =v_genotipo.loudness;
vpitch    =v_genotipo.pitch;
vespectro =v_genotipo.espectro;
N =length(vloudness);
Ns =length(seg);
if Ns<N
    seg(N)=0; segmento=seg;
else
    segmento=seg(1:N);
end;

x = segmento'.*vloudness;   % Modificaçao do Loudness

% Modificaçao do Pitch
y(N) =0;
i =1; j =1; ratio =0;
while j < N-1,
   ratio =ratio + vpitch(i);
   if ratio>1
      inteiro =floor(ratio);    % parte inteira do ratio
      fracio =ratio - inteiro;      % parte fracionária do ratio
      i =i +inteiro;
      if i>N break; end;
      ratio =fracio;                % ratio recebe parte fracionaria
   end;
   xa =x(i); xb =x(i+1);
   y(j) =((xb-xa)*ratio) + xa;      % interpolação linear
   j =j + 1;
end;
yr =y;

% Modificaçao do Espectro
z=fft(yr);
w=z.*vespectro;
```

```
novo_segmento =real(ifft(w));
```

O exemplo dado a seguir demonstra a construção do novo indivíduo a partir da operação genética sobre um indivíduo (cello.wav) em relação ao melhor indivíduo (trompete.wav). Ambos indivíduos representam uma nota contínua de cada instrumento representado. A operação de crossover ocorre sobre a curva de *pitch* e a taxa de crossover é alfa=50% e mutação beta=10%. Seguindo o gráfico dos segmentos original e modificado, tem-se também os gráficos das curvas psicoacústicas de *loudness*, *pitch* e espectro para os seus três estados: genótipo original, genótipo modificado e a variação do genótipo.




LOUDNESS
Original
Loudness - tempo (seg)
Modificado
Loudness - tempo (seg)
Variaçao
Loudness - tempo (seg)
PITCH
Original
pitch - tempo (seg)
Modificado
pitch - tempo (seg)
Variaçao
pitch - tempo (seg)

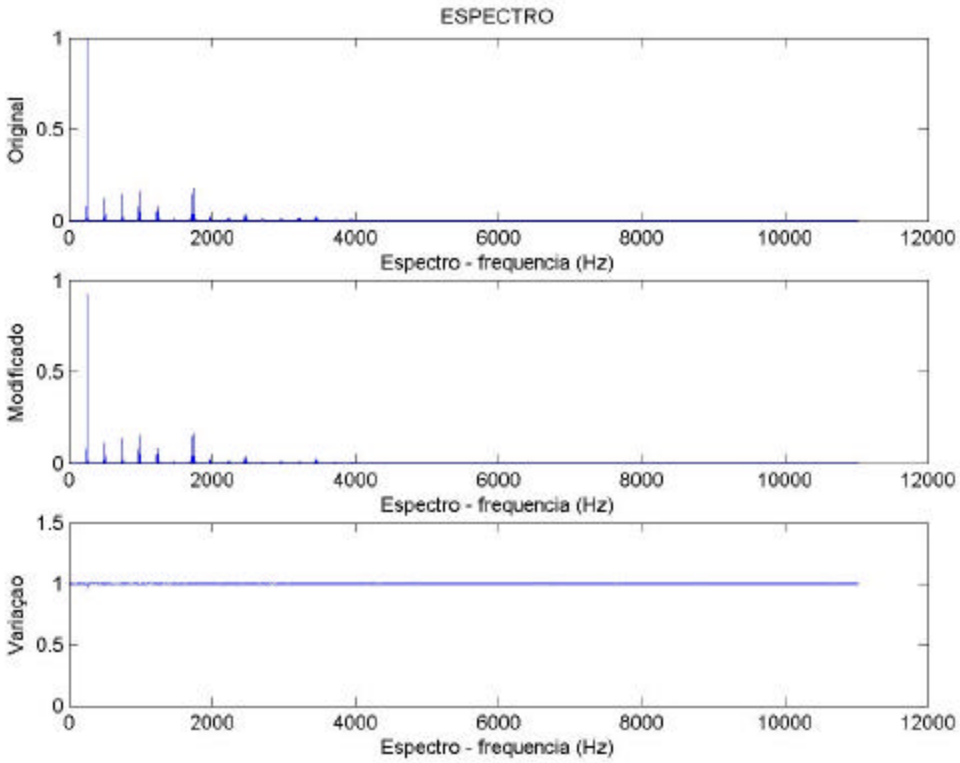


Figura 4.18 Construção de um novo indivíduo na população a partir da modificação do genótipo do indivíduo 4 (cello.wav) por crossover com indivíduo 28 (trompete.wav). As taxas de operação de crossover se dá sobre a curva de pitch, com taxas de operação genetica: alfa e beta.

4.8 Simulação do método da síntese evolutiva sobre o indivíduo

Vimos que o método de síntese evolutiva se baseia em dois processos: seleção e reprodução. A seleção é feita pela medida da distância entre indivíduos e a reprodução pelos operadores genéticos. No método de síntese evolutiva do capítulo 2 o melhor indivíduo é o único que passa de uma geração a outra inalterado. O restante dos indivíduos da população tem o seu segmento sonoro modificado pelos operadores genéticos durante o processo de reprodução.

Simulamos o método da síntese evolutiva sobre o indivíduo através de um script do MATLAB. Utilizamos na simulação os 36 segmentos sonoros citados na figura 4.13 como indivíduos dos conjuntos população e alvo. Apesar dos indivíduos terem tamanhos diferentes, a restrição de tamanho do segmento é dada pela operação genética que é feita sobre o tamanho do indivíduo da população que está sendo manipulado pelos operadores genéticos (reprodução) ou medido em relação à distância com o conjunto alvo (seleção). Pode-se também utilizar os mesmos indivíduos no conjunto população e alvo, lembrando apenas que no conjunto população, com exceção do melhor indivíduo, todos os outros indivíduos são modificados a cada geração pelos operadores genéticos e no conjunto alvo os indivíduos permanecem inalterados (até que o usuário os modifique). A seqüência de etapas da simulação é descrita pelo diagrama abaixo. Os trechos correspondentes da script de simulação do MATLAB é apresentado ao lado:

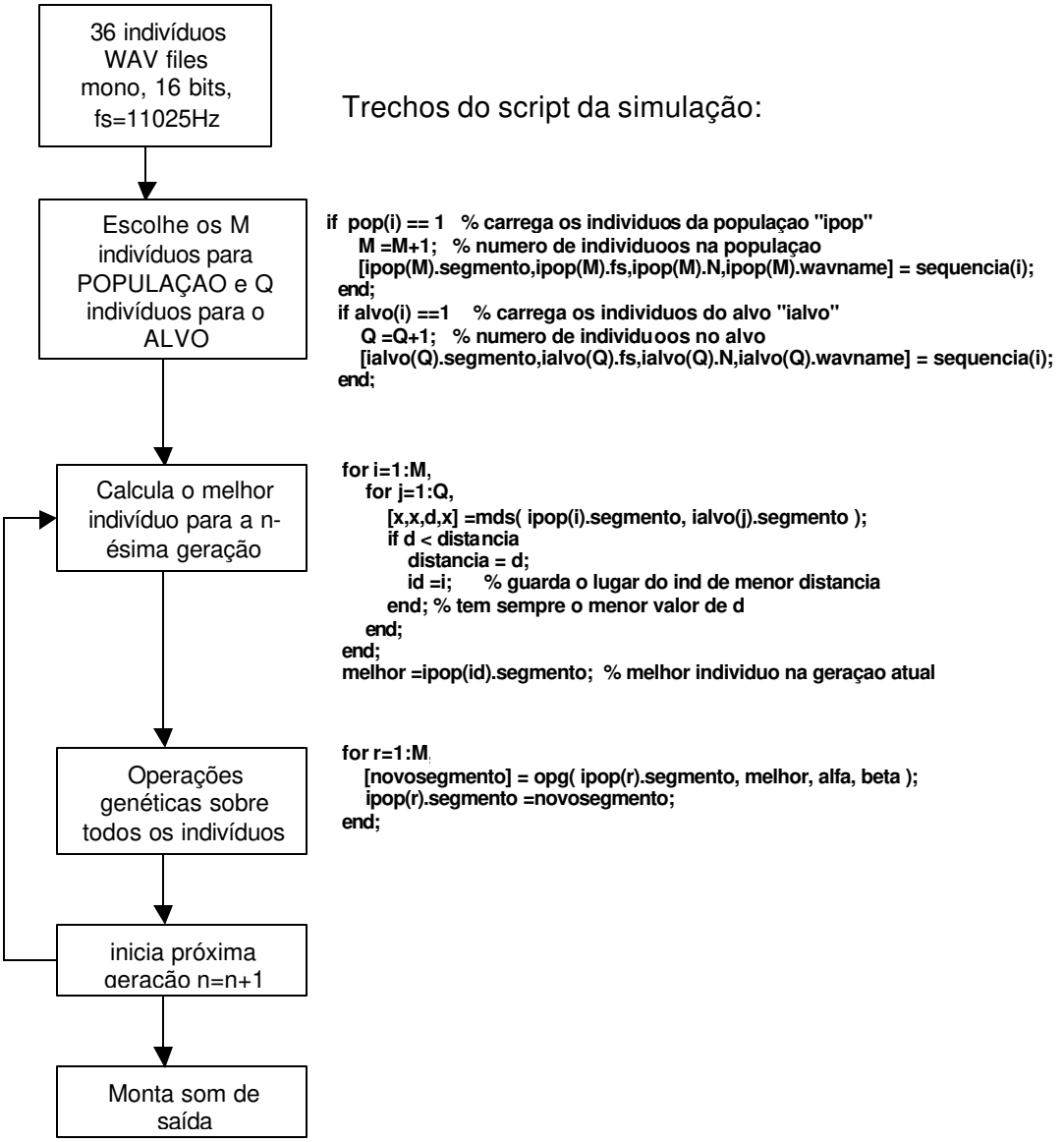


Figura 4.19 Diagrama da simulação do método da síntese evolutiva descrita no capítulo 2.

Esta simulação utiliza as funções opg.m e mds.m, descritas anteriormente. A função opg.m realizar a operação genética sobre o segmento e a função mds.m faz a medida da distância entre os segmentos. O funcionamento de ambas é dado respectivamente nos itens 4.3 e 4.5 deste capítulo.

Para ilustrar o resultado da simulação do primeiro modelo da síntese evolutiva, realizamos uma simulação com 100 gerações, medida da distância euclidiana sem peso, d1, para uma população de 5 indivíduos que são os segmentos sonoros em ciclo dados abaixo:

'baixo c.wav', 'flauta c.wav', 'guitar c.wav', 'trompete c.wav', 'voz fem c.wav'

e o conjunto alvo com outros 5 indivíduos em ciclo, que representam um grupo de diversos tipos de saxofones:
'sax alto c.wav', 'sax baritono c.wav', 'sax c.wav', 'sax soprano c.wav', 'sax tenor c.wav'

As taxas de operação genética são mantidas em alfa=50% e beta=10%. Após 100 gerações, o melhor indivíduo é dado a seguir:

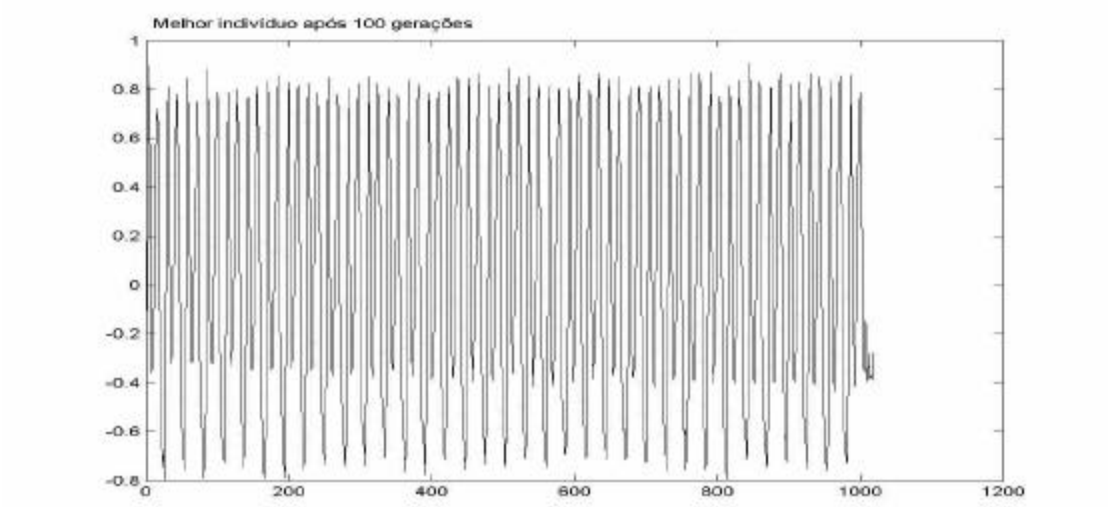


Figura 4.20 Melhor segmento após 100 gerações de síntese evolutiva para segmentos sonoros.

Escutando este segmento sonoro em repetição (*looping*) podemos dizer que este som se assemelha ao som de um saxofone com algum ruído. Como o conjunto alvo é dado por uma serie de segmentos sonoros de saxofones, a evolução do primeiro modelo da síntese evolutiva guiou o resultado das operações genéticas para o som especificado no conjunto alvo.

É importante ressaltar que o objetivo da síntese evolutiva vai além da criação dinâmica de sons com riqueza de padrões. É também considerado como informação importante o caminho percorrido pelo processo evolutivo, sob a forma da sucessão dos melhores indivíduos ao longo das gerações. Para as 100 gerações desta simulação, tivemos 100 melhores indivíduos, onde apenas o último deles é dado acima. A maior contribuição da síntese evolutiva é o processo de evolução sonora descrito pela sucessão dos 100 melhores indivíduos que culminaram no último melhor indivíduo da simulação dado na figura 4.20. Se, durante estas gerações, o conjunto alvo mudasse, a evolução tomaria outro rumo. O efeito do condicionamento dinâmico da evolução da síntese pelos indivíduos do conjunto alvo é um dos fatores mais importantes que fazem da síntese evolutiva um método inovador da síntese sonora.

4.9 Simulação da síntese evolutiva utilizando as curvas psicoacústicas como genótipo do indivíduo

No método de síntese evolutiva do capítulo 2, vimos que o melhor indivíduo é o único a passar de uma geração para a outra inalterado. No método de síntese evolutiva do capítulo 3, a utilização das curvas psicoacústicas como genótipo do indivíduo faz com que o melhor indivíduo seja o único a ter o seu segmento sonoro alterado pela construção de um novo segmento com base na variação de seu genótipo. Afim de tornar o processo mais eficiente, o restante dos indivíduos na população tem apenas os seus genótipos alterados a cada geração. O segmento sonoro destes indivíduos permanece o mesmo, até que este seja eventualmente escolhido pelo processo de seleção (distância de Hausdorff) como o melhor indivíduo em uma dada geração da população.

Da mesma forma que fizemos no modelo de síntese evolutiva do capítulo 2, simulamos o modelo do capítulo 3 através de um *script* do MATLAB. O diagrama da simulação do modelo da síntese evolutiva utilizando as curvas psicoacústicas como genótipo do indivíduo é dado abaixo:

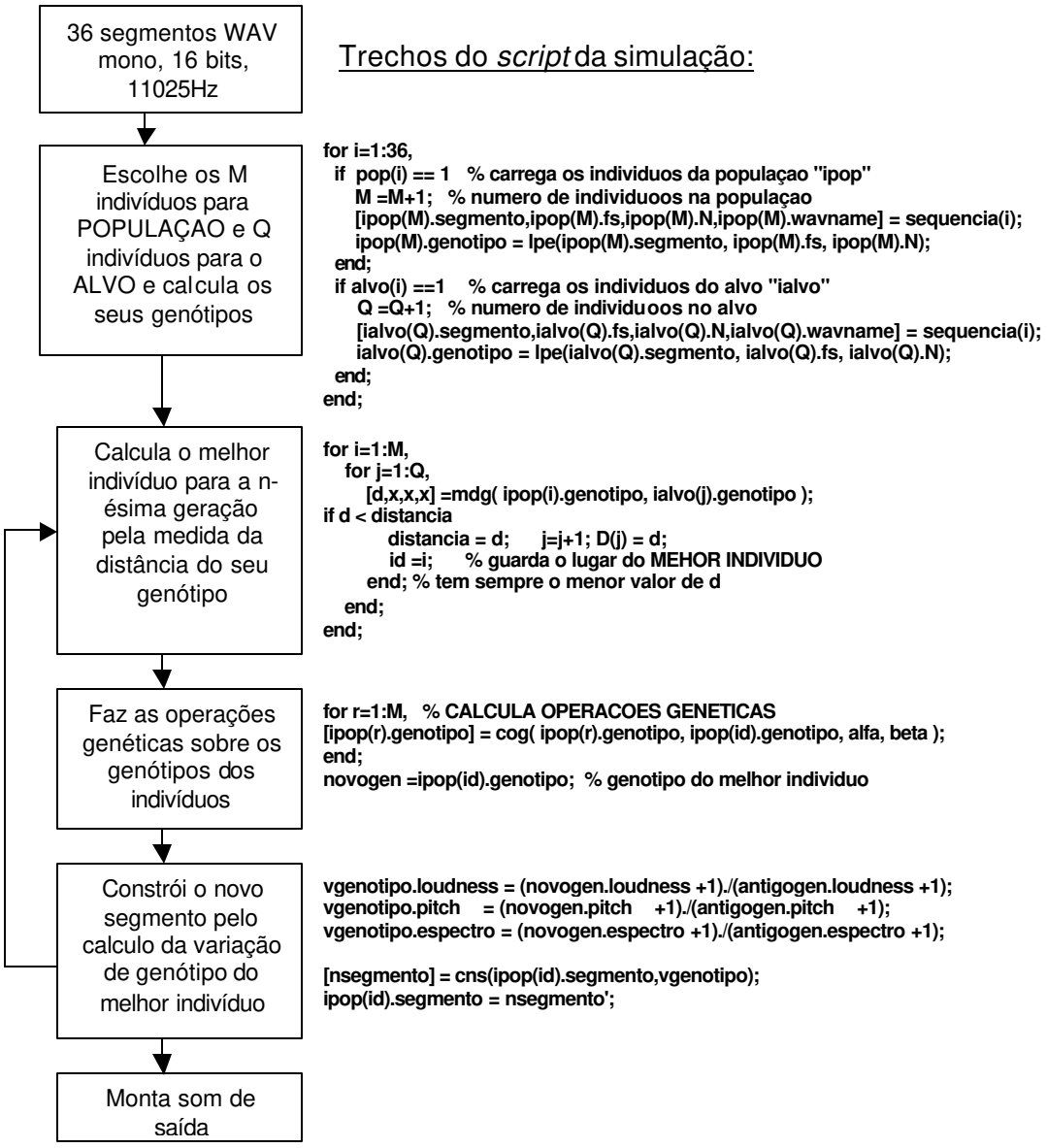


Figura 4.21 Diagrama da simulação do método da síntese evolutiva descrita no capítulo 3.

As funções desenvolvidas para simulação da síntese evolutiva utilizando curvas psicoacústicas são descritas a seguir:

Função	Comentário
[genotipo] = lpe(individuo,fs,N)	**Cálculo das curvas de Loudness, Pitch e Espectro (Extração do Genótipo)** **Entrada: indivíduo, taxa de amostragem e tamanho do segmento.** **Saída: Calcula as curvas de loudness, pitch e espectro**
[d1,d2,d3,d4] = mdg(gIndPop, gMelhorInd)	**Medida da distância entre dois genótipos.** **Entrada: genótipo do indivíduo da população e genótipo do melhor indivíduo.** **Saída: Os quatro tipos de distância entre genótipos.**
[gNovoInd] = cog(gIndPop,gMelhorInd,alfa,beta)	**Cálculo da Operação Genética.** **Entrada: genótipo do indivíduo da população; genótipo do melhor indivíduo, e as taxas de crossover (alfa) e mutação (beta)** **Saída: Genótipo do novo indivíduo.**
[novoInd] = cns(ind,vgenotipo)	**Construção do Novo Segmento.** **Entrada: indivíduo e variação do genótipo do indivíduo.** **Saída: novo indivíduo (segmento sonoro)**

Para exemplificar a simulação deste modelo de síntese evolutiva, mostramos abaixo o resultado da sua simulação para as mesmas condições da simulação do modelo anterior. O melhor indivíduo após 100 gerações é dado abaixo:

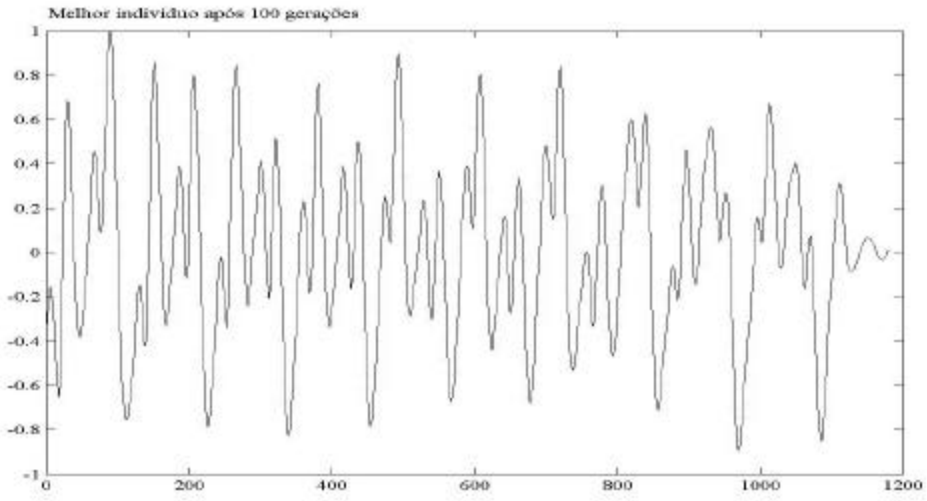


Figura 4.22 Melhor segmento após 100 gerações de síntese evolutiva utilizando curvas psicoacústicas.

Ao escutar a repetição desse segmento sonoro, embora diferente, também se parece ao som de um saxofone. Na simulação de ambos os modelos, observou-se que estes convergem rapidamente para um segmento sonoro que permanece inalterado até o fim das gerações da síntese. Isto talvez ocorra devido ao pequeno tamanho da população adotado nas simulações (5 elementos). O fato é que os dois modelos foram capazes de sintetizar um indivíduo cujo segmento sonoro se assemelha perceptualmente ao som dos indivíduos do conjunto alvo (saxofones), cujo papel é condicionar a evolução dos indivíduos na síntese evolutiva.

5 Conclusões e Comentários Finais

Apresentamos as possíveis aplicações para a síntese evolutiva de segmentos sonoros e citamos algumas possibilidades para futuras pesquisas que encontramos ao longo do desenvolvimento deste trabalho.

5.1 Resultados da síntese evolutiva

Observamos que o primeiro modelo da síntese evolutiva [cap 2], onde as operações genéticas e medidas de distância são feitas sobre os segmentos sonoros, é computacionalmente menos complexo que o modelo utilizando curvas psicoacústicas.

No entanto, este primeiro modelo de síntese evolutiva é mais limitado nas possibilidades de manipulação sonora dos segmentos. O operador genético *crossover* pode apenas misturar seções dos segmentos sonoros e o operador mutação apenas impõe uma perturbação na amplitude de cada segmento sonoro operado. Outras características psicoacústicas destes segmentos sonoros, como o pitch ou a composição espectral, permanecem praticamente inalteradas pelos operadores genéticos durante o processo de síntese.

5.1.1 O segmento sonoro como indivíduo

Neste primeiro modelo de síntese evolutiva o segmento sonoro é o indivíduo da população de M indivíduos. Os processos de reprodução e seleção ocorrem diretamente sobre os indivíduos da população. Cada indivíduo é um segmento sonoro, que representa um trecho amostrado de um dado som. Convencionamos amostra-los em 16 bits, com taxa de amostragem 11025 Hz. Desse modo o segmento sonoro pode representar sons com relação sinal-ruído de 96dB e conter componentes em freqüência até aproximadamente 5KHz. Todos os segmentos sonoros são normalizados em amplitude dentro do intervalo [-1,1].

Para a restrição do tamanho do segmento sonoro, convencionou-se limita-lo em 1024 pontos, o que corresponde a um intervalo de tempo de 0,092s para a taxa de amostragem utilizada. Este intervalo é muito curto para representar segmentos sonoros que não estejam em ciclo, pois o mesmo deve ser repetido diversas vezes (em *looping*) a fim de que se possa perceber o resultado das operações genéticas. O formato utilizado é o padrão WAV, com 1 canal de áudio.

5.1.2 As operações genéticas sobre o indivíduo

Os operadores genéticos utilizados são os operadores clássicos da computação evolutiva: *crossover* e a mutação. O *crossover* mistura partes do segmento sonoro do indivíduo com o melhor indivíduo da geração anterior da população. A taxa de crossover alfa determina o grau de mistura entre as seções dos indivíduos. Se alfa=0, não há mistura. Se alfa=1 ocorre a substituição da seção no indivíduo pela seção correspondente do melhor indivíduo.

A operação mutação ocorre no indivíduo precede a operação *crossover* e independe de outro indivíduo para ser realizada. A quantidade de mutação aplicada em um indivíduo é determinada por uma taxa beta, que corresponde ao grau de perturbação na amplitude do segmento sonoro. Se beta=0 não há perturbação. Se beta=1 a perturbação é total. A perturbação é feita pela multiplicação do segmento sonoro com um segmento de mesmo tamanho, contendo números aleatórios.

Observou-se que a operação de *crossover*, da maneira que ela está definida no capítulo 2, substitui uma seção do segmento sonoro operado por uma seção que é o resultado da mistura da seção original (sem operação) com a seção correspondente do segmento sonoro do melhor indivíduo da geração anterior. O resultado sonoro dessa operação pode ser descrito como similar à sensação de edição ou inserção de um som dentro de outro som, porém como uma justaposição de trechos sonoros, o que, via de regra, não é percebido como um novo timbre sonoro. Se a seção escolhida aleatoriamente por k1 e k2 é do mesmo tamanho do segmento sonoro operado, o resultado da

operação de *crossover* é o mesmo que o da mixagem dos dois segmentos sonoros. Se a seção for menor que o segmento sonoro operado, como acontece na maioria dos casos, então pode-se perceber auditivamente o que foi descrito acima; algo que também pode ser definido como uma colagem sonora.

O resultado perceptual sonoro do operador mutação é a sensação de inserção de um certo grau de ruído no segmento sonoro, que varia em intensidade de acordo com a taxa de mutação beta.

Mesmo assim, apenas se valendo de mistura e inserção de ruído, este modelo de síntese evolutiva chega a resultados sonoros inusitados, uma sucessão imprevisível de trechos sonoros, com elementos misturados entre si, que vão evoluindo de acordo com sua proximidade sonora, atribuída pelas consecutivas medidas de distância entre os indivíduos do conjunto população e alvo.

5.1.3 Medida de distância entre os indivíduos

A adequação dos indivíduos é medida pela distância de Hausdorff entre cada indivíduo da população e os indivíduos do conjunto alvo. No caso, estamos calculando a distância de Hausdorff entre um conjunto unitário (o r-ésimo indivíduo da população) e o conjunto alvo com Q indivíduos. Testamos quatro tipos de distância: a distância euclidiana sem peso, a distância euclidiana com peso, a distância diferencial sem peso e a distância diferencial com peso. O peso utilizado é decrescente do primeiro ao último ponto do segmento sonoro, o que privilegia o começo do segmento como sendo o de maior relevância para a percepção auditiva.

Os experimentos que fizemos mostraram que a distância sem peso é o melhor critério para a medida da distância entre segmentos. Outras métricas podem e devem ser testadas em futuras pesquisas da síntese evolutiva.

5.2 Resultados da síntese evolutiva utilizando curvas psicoacústicas como genótipo

O segundo modelo de síntese evolutiva [cap 3], utiliza curvas psicoacústicas como genótipo do indivíduo. Os processos de seleção e reprodução passam a ser feitos sobre o genótipo do indivíduo. Isto significa que a manipulação dos operadores genéticos e a medida de distância passam a serem feitas sobre o trio de funções que corresponde ao genótipo: as curvas de *loudness*, *pitch* e espectro. Conseqüentemente, o processamento do método da síntese evolutiva exige mais recursos computacionais para ser executado.

No entanto, a utilização de curvas psicoacústicas provou ampliar em muito a capacidade da síntese evolutiva. Cada curva é vista como um gene que codifica informação relevante à percepção do som representado pelo indivíduo. A manipulação de uma curva psicoacústica equivale à modificação de uma característica perceptual do som sintetizado. Além disso, a medida da distância, agora feita entre genótipos, está mais condizente com a percepção auditiva, pois com a utilização das curvas psicoacústicas a medida da distância é feita sobre os fatores que são relevantes à percepção e reconhecimento do som pela audição do usuário.

5.2.1 Extração do genótipo do indivíduo

A extração adequada das três curvas psicoacústicas do segmento sonoro é fundamental para o desempenho da síntese evolutiva. Estas curvas irão representar as características mais importantes para a percepção daquele som. Se as curvas não forem extraídas corretamente, os processos de reprodução e seleção do método terão o seu desempenho comprometido.

Dos experimentos realizados concluímos que é muito difícil garantir uma fidelidade da representação das grandezas psicoacústicas pelas curvas psicoacústicas extraídas dos segmentos sonoros, mesmo porque estas são características pessoais, e como representam características perceptuais, estas variam de acordo com cada ouvinte.

Para obter as curvas psicoacústicas, desenvolvemos algoritmos específicos. Estes foram inspirados no processo de percepção auditiva humana. Os resultados desses algoritmos são aproximações das curvas psicoacústicas que representam as grandezas psicoacústicas de um segmento sonoro. Em alguns momentos estes algoritmos não obtem as características psicoacústicas adequadas mas em geral apresentam um bom desempenho. O refinamento dos algoritmos de obtenção das curvas psicoacústicas pode e deve ser feito em pesquisas futuras.

5.2.2 As operações genéticas sobre o genótipo do indivíduo

Uma vez obtidas as 3 curvas psicoacústicas, o passo seguinte é a sua manipulação pelos operadores genéticos. Não foi testada a manipulação aleatória de mais de uma curva psicoacústica por operação de *crossover*, ou seja, a possibilidade do *crossover* escolher aleatoriamente uma ou mais curvas psicoacústicas para operar.

A operação de mutação insere uma perturbação na amplitude de cada curva psicoacústica. Em termos da curva de loudness, essa perturbação é bastante parecida com o ruído branco inserido pela operação de mutação no modelo da síntese evolutiva do capítulo 2. Na curva de *pitch*, essa perturbação é percebida como uma variação aleatória do *pitch* do som representado pelo segmento sonoro. Na curva de espectro a perturbação modifica a amplitude das componentes em freqüência do som. Cada perturbação é dada por um segmento de números aleatório distinto mas a taxa de mutação é a mesma para as 3 curvas psicoacústicas. Poderiamos testar diferentes taxas de mutação para cada curva psicoacústica. Apesar de mais sofisticada que no modelo do capítulo 2, observou-se que a operação de mutação ainda precisa ser melhorada para corresponder mais adequadamente à operação de mutação como ela é na biologia. Ao invés de inserir uma perturbação através da multiplicação por números aleatórios, a perturbação pudesse ser feita de uma maneira mais adequada à representação psicoacústica. É possível que o desenvolvimento do modelo de genes e cromossomos para as curvas psicoacústicas venha a facilitar o entendimento do que seria uma perturbação psicoacústica adequada para cada curva psicoacústica.

Uma possibilidade não testada na parte experimental, mas que poderá facilmente ser executada, é a utilização de taxas de operação genética escolhidas aleatoriamente. Nossos experimentos sempre utilizaram taxas alfa e beta fixas ao longo das gerações.

5.2.3 A medida de distância entre genótipos

Utilizamos como medida de distância entre os genótipos a média aritmética das três distâncias das curvas psicoacústicas entre os dois genótipos, entre as curvas de *loudness*, *pitch* e espectro. As medidas de distância testadas foram as mesmas quatro anteriormente citadas: euclidiana sem peso, euclidiana com peso, diferencial sem peso e diferencial com peso. Não obtivemos grandes diferenças entre as distâncias assim optamos por usar a distância euclidiana sem peso. Podem-se testar outras métricas e também outros pesos entre as distâncias de cada curva que não foram ainda investigadas.

Uma outra possibilidade de medida da distância é a multiplicação por uma matriz de correlação que misture as entradas da matriz genotípica. O genótipo pode ser visto como uma matriz 3xN, uma vez que as três curvas psicoacústicas são vetores de N elementos. Poderíamos assim utilizar uma nova notação para definir o genótipo dos indivíduos do conjunto população w e alvo t como uma matriz 3xN.

É importante salientar que cada curva psicoacústica procura estabelecer uma analogia entre o seu formato e a percepção da grandeza psicoacústica que representa. Assim a medida de distância ideal seria aquela que pudesse medir a semelhança gráfica entre o formato das curvas psicoacústicas correspondentes em cada genótipo. Pesquisa nessa área ainda deve ser feita.

5.2.4 A construção do novo indivíduo pela variação de suas curvas psicoacústicas

Uma vez que as operações genéticas tenham sido feitas sobre os genótipos dos indivíduos e a medida de distância tenha encontrado o indivíduo cujo genótipo tenha a menor distância com o os genótipos dos indivíduos do conjunto alvo o passo seguinte é a reconstrução deste indivíduo com base na variação de seu genótipo.

A construção do novo indivíduo pela variação das suas curvas psicoacústica dado pelo algoritmo que desenvolvemos, dado na função cns.m, ocorre em três etapas de construção que são seqüenciais. Tem-se primeiro a construção do *loudness*, depois a construção do *pitch* e por fim a construção do espectro, ou seja, constrói-se o novo indivíduo na seqüência de operações: [l,p,e]. No entanto, a construção do novo indivíduo poderia ocorrer em outras seqüências dessas três operações, a saber: [l,e,p], [p,l,e], [p,e,l], [e,l,p], [e,p,l] e [l,e,p]. Ainda está por ser pesquisado se estas operações são de fato comutativas.

5.3 Possíveis utilizações para a síntese evolutiva

Como acontece no princípio de qualquer nova tecnologia, a síntese evolutiva de sons atualmente se encontra nos primeiros estágios de desenvolvimento e a sua aplicação prática ainda é pouco definida, embora as possibilidades sejam muito grandes. A grande diferença da síntese evolutiva em relação às outras sínteses é que este é um método evolutivo de síntese sonora, enquanto os outros métodos são causais. Enquanto um sistema causal apresenta uma única saída para uma entrada, um sistema evolutivo tem memória, e aprende com cada entrada, desse modo a qualidade de sua saída evolui ao longo do tempo.

5.3.1 Sintetizador dinâmico de sons

A utilização mais imediata que antevemos para a utilização da síntese evolutiva é em conjunto com a síntese *wavetable*. Nesta, cada som é armazenado na tabela *lookup*, em dois segmentos sonoros: o ataque e o ciclo. Desse modo a tabela *lookup* passa a ser a população de indivíduos onde cada indivíduo é representado por dois segmentos sonoros. O controle do sistema *wavetable* é feito pela linguagem MIDI.

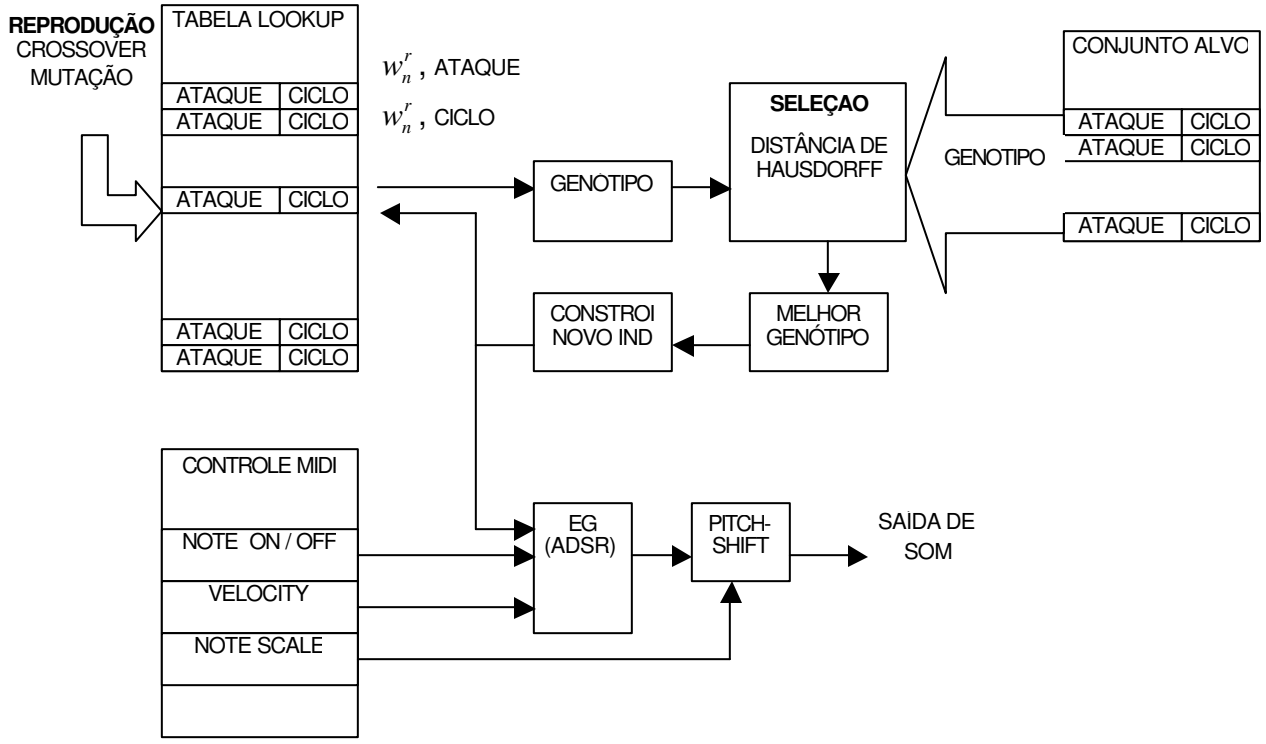


Figura 5.1 Diagrama simplificado de um sintetizador evolutivo.

5.3.2 Automação de controle timbrístico

O método da síntese evolutiva pode ser utilizado não apenas como sintetizador de sons mas também para adequação automática do timbre de acordo com a dinâmica da música (seqüência de notas, velocidade). Os processadores de áudio se baseiam em sistemas que, até onde sabemos, não se valem de características evolutivas. Por exemplo, o compressor de *loudness*, não faz muito além de multiplicar o segmento sonoro por uma função que força a sua intensidade dentro de um limite pré-estabelecido. O conhecimento desenvolvido com a síntese evolutiva pode ser adaptado para buscar e modificar características psicoacústicas determinadas pelo usuário, de modo a adaptar os segmentos sonoros. Por exemplo, pode-se criar um sistema de controle automático de equalização de um som de acordo com segmentos sonoros de referência no conjunto alvo. Neste sistema, os operadores genéticos modificariam os parâmetros do equalizador e a medida de distância compararia o genótipo do segmento sonoro com os genótipos dos segmentos de referência no conjunto alvo.

5.3.3 Reconhecimento automático de seqüências sonoras

Um dos grandes desafios do processamento digital de áudio é a criação de sistemas de reconhecimento de padrões sonoros (voz, ruídos, sinais, etc.). O método da síntese evolutiva pode ser adaptado para encontrar o segmento sonoro mais parecido de um conjunto de outros segmentos sonoros. O conjunto alvo receberia os segmentos a serem reconhecidos e a medida de distância de seus genótipos os compararia com um conjunto população de segmentos conhecidos. O ajuste fino do reconhecimento dos segmentos sonoros seria feito pelos operadores genéticos que modificariam sutilmente os segmentos referência da população de modo a facilitar seu reconhecimento.

5.3.4 Composição dinâmica de timbres sonoros

Viu-se no capítulo introdutório que, por restrições técnicas, até aproximadamente a década dos 40, a composição musical ficou restrita à manipulação das grandes estruturas de som, conhecidas em música como: melodia, harmonia e ritmo. Até então, a exploração timbrística se dava de maneira bastante limitada, na forma da orquestração e arranjo, onde o compositor pode escrever para um grupo de instrumentistas que tentam executar, na medida do possível, um número restrito e pessoal de variações no timbre de seus instrumentos.

Com o advento da música eletroacústica, que tem seu marco inicial em outubro de 1948 com a difusão do *Concert de Bruits* pela Radiodiffusion-Télévision Française, RTF, compositores passaram a manipular o timbre sonoro com recursos eletrônicos em estúdios de gravação. Hoje a composição de timbres sonoros ainda se dá de uma forma estática, ou seja, tem-se um resultado timbrístico único para cada parametrização dos processamentos sonoros aplicados.

O método da síntese evolutiva pode ser utilizado para expandir os horizontes da composição timbrística. Sob essa perspectiva, o método da síntese evolutiva passa a ser um instrumento musical que permite elaboração dinâmica de timbres sonoros. De forma similar ao que acontece no controle de um instrumento musical por um instrumentista, o sistema da síntese evolutiva também pode ser controlado de forma intuitiva pois não é necessário ao usuário conhecer os parâmetros psicoacústicos que quer obter no som de saída para controlar a evolução da síntese sonora. Basta que o usuário manipule os segmentos sonoros do conjunto alvo e as taxas de operações genéticas para controlar dinamicamente os rumos da evolução do sistema e conseqüentemente o som de saída.

5.4 Algumas possibilidades de pesquisas futuras da síntese evolutiva de segmentos sonoros

Durante o desenvolvimento deste trabalho, deparamos com uma grande quantidade de idéias e possibilidades para futuros desenvolvimentos do método da síntese evolutiva de sons. Comentamos a seguir aqueles que nos pareceram mais pertinentes para os próximos modelos desta síntese que, por ser um campo das sínteses de som ainda por ser desbravado, apresenta uma imensa quantidade de oportunidades para o desenvolvimento de pesquisas.

5.4.1 Inclusão de genes e cromossomos para as curvas psicoacústicas

No capítulo 3, introduzimos o conceito de genótipo para o segmento sonoro, composto por três curvas psicoacústicas. Este conceito pode ser expandido. Cada curva psicoacústica pode ser vistas como um cromossomo sonoro e seções destas podem ser vistas como genes. Assim as operações genéticas passam a acontecer sobre os genes (partes das curvas psicoacústicas) e não sobre o cromossomo (curva psicoacústica).

Um outro fato decorrente da associação do conceito de cromossomos às curvas psicoacústicas é a possibilidade de se criar dominância e recessividade cromossômica. Pode-se, por exemplo, criar um modelo de genótipo para o indivíduo que apresente 3 pares de curvas psicoacústicas. Cada par seria como um cromossomo diplóide onde uma curva seria dominante, enquanto a outra seria a recessiva.

5.4.2 Um processo de reprodução N-genérico

Atualmente a operação de crossover ocorre exclusivamente entre cada indivíduo e o melhor indivíduo da população. Um possível desenvolvimento seria criar uma operação de crossover que fosse generalizada, ou seja, um operador genético do tipo crossover que faça o crossover entre dois elementos quaisquer de uma população e não somente com o melhor indivíduo.

Outra possibilidade é criar o conceito de gênero entre os indivíduos. Atualmente não existe definição de gênero, ou **sexo**, entre os indivíduos da população e o processo de reprodução é similar ao da divisão celular por mitose (reprodução assexuada). Existem diversos trabalhos sobre reprodução sexuada de algoritmos genéticos. Recentemente estes tem sido catalogados com o termo de "biocomputação" [Miller,95].

Pode-se criar um modelo **N-gen**, onde diversos gêneros de indivíduos são definidos para o processo de reprodução. Decorrendo de um modelo de indivíduos com gênero, pode-se ainda estabelecer critérios de **atração** na reprodução desses indivíduos, ou seja, as características que cada indivíduo deve conter em seu genótipo para atrair um outro (ou outros) indivíduo e assim efetuar o processo de reprodução. Adiantando nesse caminho, um fator de atração poderia ser, por exemplo, a discrepância entre genótipo, ou seja, quanto mais distantes entre si, mais se atrairiam para a reprodução.

5.4.3 Uma população com tamanho variável de indivíduos

Até onde desenvolvemos o método de síntese evolutiva, o conjunto população tem um número fixo de indivíduos. Pode-se desenvolver um novo método de síntese evolutiva concebendo uma população de tamanho variável. Para isso devem-se incluir critérios de eliminação de indivíduo, ou **morte**, por tempo de permanência na população e/ou medida de distância (elimina o indivíduo de maior distância do conjunto alvo).

Também pode-se incluir o conceito de reprodução assíncrona. Atualmente a reprodução ocorre de maneira seqüencial, do primeiro ao último indivíduo da população. A reprodução aleatória pode também ser um critério para a variabilidade do tamanho populacional.

Podemos ainda incluir um parâmetro de crescimento populacional associado aos parâmetros de controle do período de maturação. Um interessante modelo matemático a ser pesquisado é o da **presa** / **predador**.

5.4.4 Período de maturação do indivíduo

No modelo atual da síntese evolutiva, quando um novo indivíduo é gerado pelos operadores genéticos, é apto a reproduzir na próxima operação genética. Pode-se refinar este modelo incluindo um período de maturação, ou **infância**, para os indivíduos recém criados. Durante a maturação o indivíduo não poderia reproduz, ou seja, não iria disseminar seu genótipo pela população. Ao invés disso, aprenderia, ou seja, receberia informação para o aprimoramento de seu genótipo através de uma interação direta com os genótipos progenitores bem como uma **interação social** (não genéticas) com outros genótipos daquela geração levando-o a um processo de amadurecimento e melhor adaptabilidade para a vida adulta, ou seja, quando este indivíduo se tornasse apto para o processo de reprodução. Caso a população tivesse um número fixo de indivíduos, poderia-se impor que, findo o período de maturação, os progenitores seriam excluídos da população (morreriam) restando apenas os indivíduos descendentes, preservando assim o número de indivíduos do conjunto população a cada geração.

5.5 Comentários finais

Os dois modelos do método da síntese evolutiva foram testados em populações com poucos indivíduos, definidos por segmentos sonoros de baixa qualidade sonora (arquivos WAV com 1 canal de som amostrados a 11025Hz). Foi possível observar que este método de síntese apresenta grandes possibilidades. A síntese evolutiva de segmentos sonoros é o primeiro método, que temos conhecimento, de computação evolutiva aplicada à síntese sonora. Tem a possibilidade de automaticamente aprender durante o processo de manipulação da população de segmentos sonoros pelos operadores genéticos como construir um segmento sonoro qualquer. Estas características, de manipulação e busca orientadas por um objetivo definido, dá ao método da síntese evolutiva um grau de inteligência que permite chegar a resultados sonoros inusitados e interessantes.

A inclusão das curvas psicoacústicas como genótipo do indivíduo contribuiu para que o método da síntese evolutiva passasse a operar mais especificamente sobre as características perceptuais do segmento sonoro. Isto focalizou os processos de seleção e reprodução sobre as características sonoras que são relevantes à percepção auditiva humana.

Sendo um método de síntese sonora original, sem similares conhecidos, observamos uma ampla gama de possibilidades para desenvolvimentos e pesquisas. Acreditamos que as possibilidades de desenvolvimento dessa área sejam muito maiores que as mencionadas neste capítulo. Introduzimos a síntese evolutiva como um novo processo de síntese sonora, que gera segmentos sonoros dinamicamente. Como estivemos desbravando uma nova fronteira da síntese sonora, tomamos muitas decisões arbitrárias do caminho por onde seguiu a pesquisa. Outras decisões foram tomadas por simplicidade computacional ou conveniência do algoritmo desenvolvido. De qualquer modo, tentamos tomar o cuidado de mencionar ao longo do texto, quando e porque essas decisões foram tomadas, pois estas são como que encruzilhadas onde uma pesquisa futura nessa área pode vir a optar por seguir outro rumo que não o nosso.

Nos parece bastante factível a aplicação comercial da síntese evolutiva. Não só como sintetizador de sons, mas como um sistema de adequação automática de características psicoacústicas especificadas pelo usuário. Para evidenciar e fomentar novas pesquisas, no capítulo 4, fizemos questão de evidenciar os passos tomados em cada etapa da simulação dos métodos de síntese evolutiva, na forma da criação em separado de funções do MATLAB para a simulação de cada etapa deste método. Cada função desenvolvida possibilitou a geração de conhecimento que pode ser reciclado e utilizado de outras maneiras e proveitos que nos passaram desapercebidos.

Fica claro para nós que a síntese evolutiva possui todas as características para vir a estabelecer um novo patamar no campo das sínteses sonoras, como um novo método, com grande potencial de contribuição para o futuro desenvolvimento das sínteses sonoras.

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.

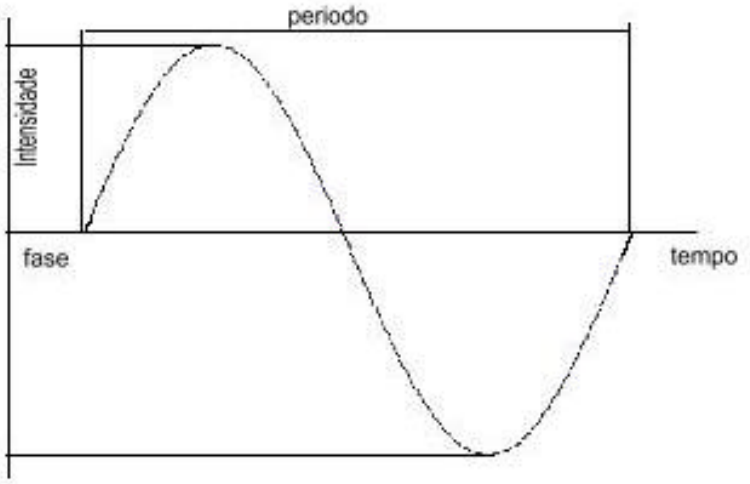


Figura A.1. A componente sonora no domínio do tempo.

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:

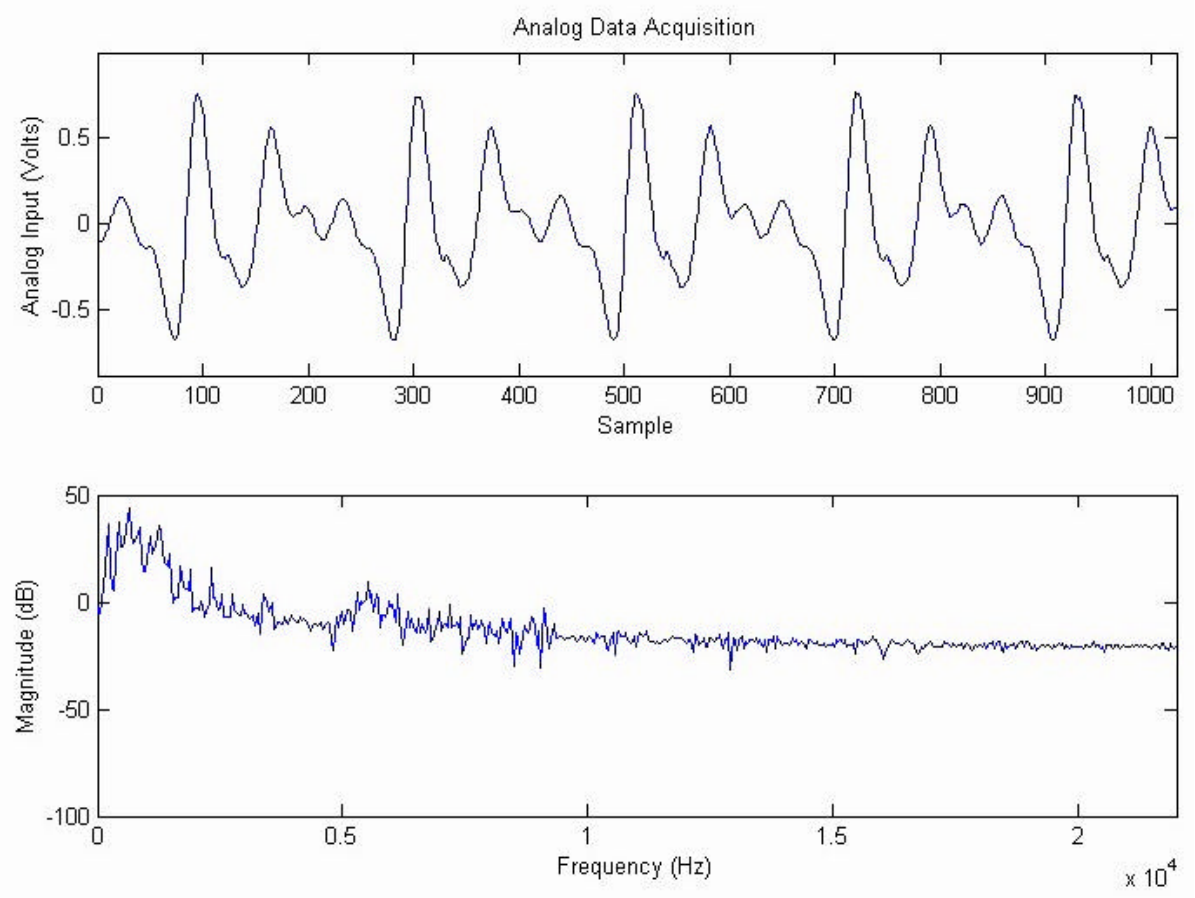


Figura A.2. Exemplo de som natural, no domínio do tempo (acima) e no domínio da freqüência (abaixo).

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:

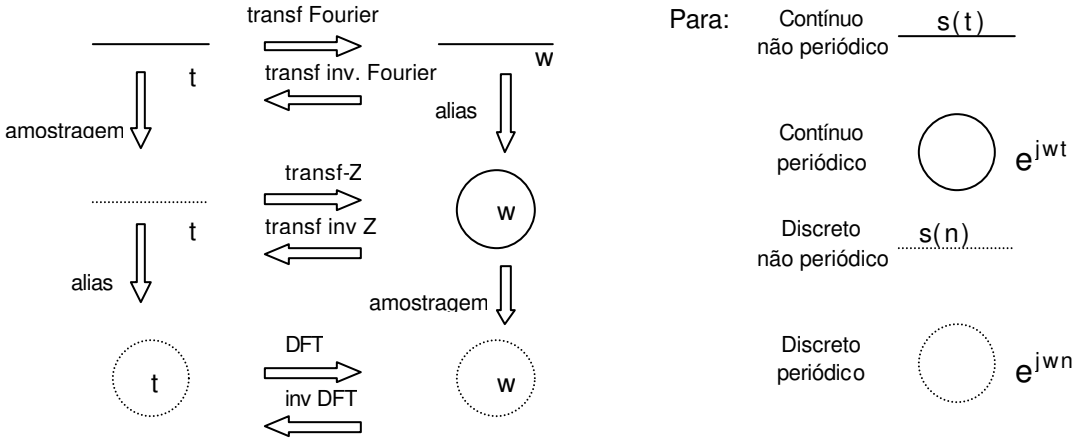


Figura A.3. Relação entre as transformações no domínio do tempo e da freqüência para sinais contínuos e discretos.

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:

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

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

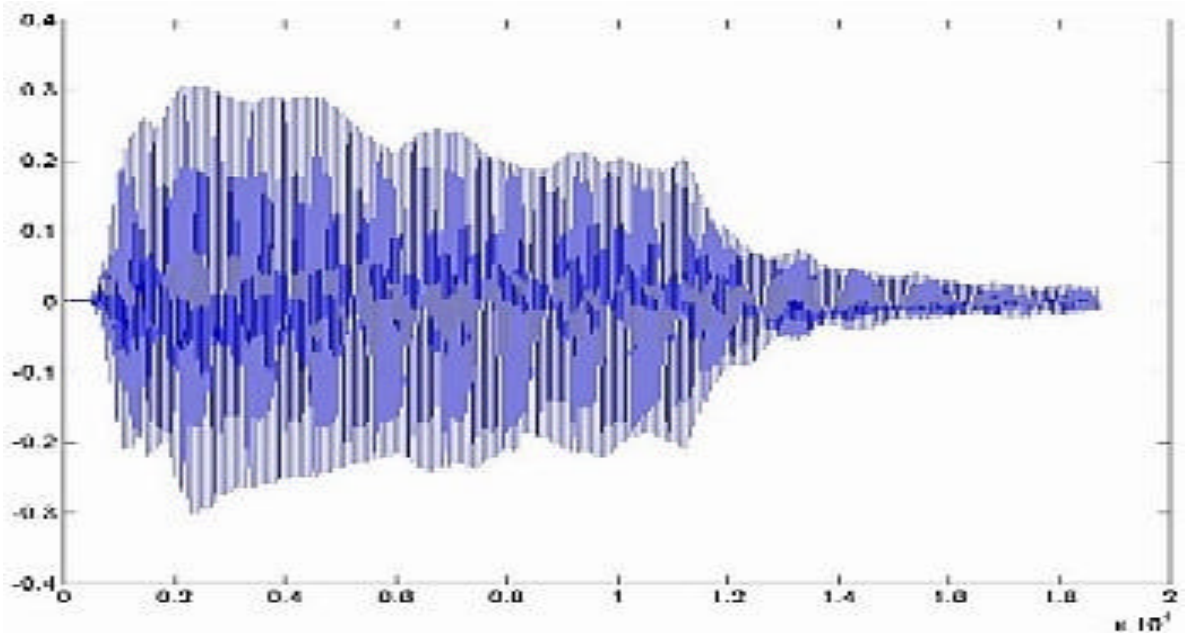
No processo natural de percepção sonora, o sistema auditivo capta a informação que está contida nas oscilações de pressão do ar e a converte de vibrações mecânicas a impulsos elétricos que são transportadas pelo nervo auditivo ao cérebro e interpretadas como a sensação fisiológica chamada de audição. 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

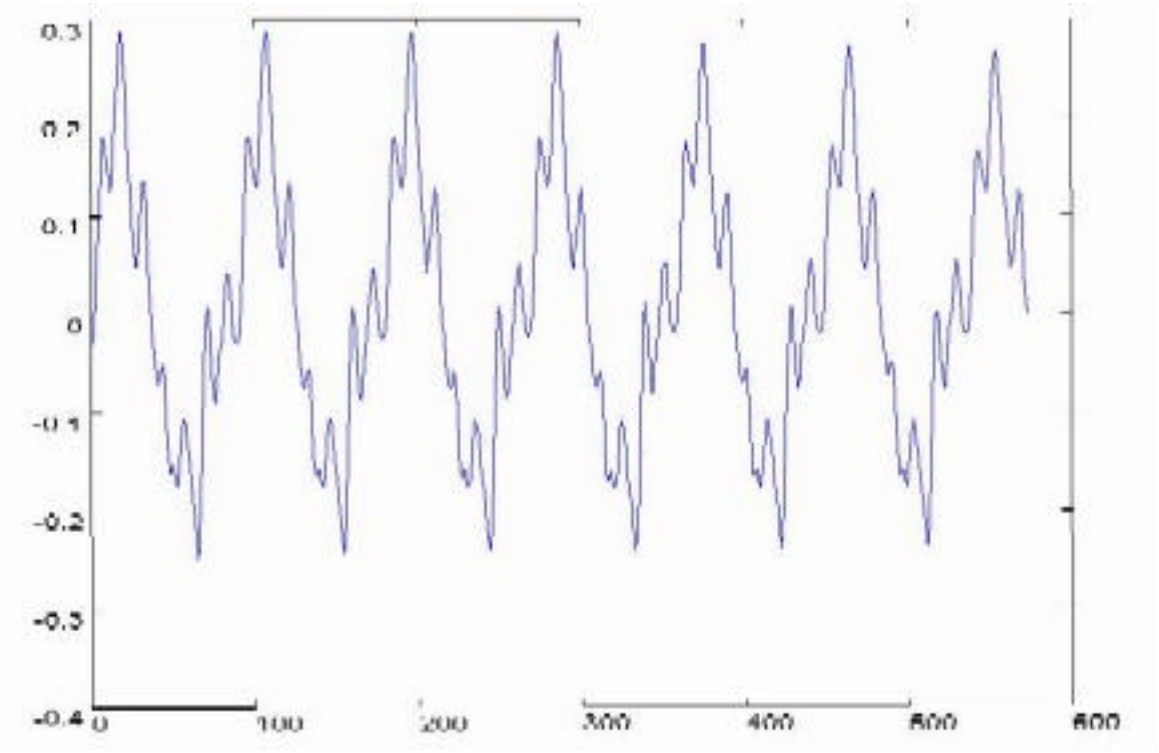
pela definição de um intervalo de tempo conhecido como persistência auditiva. Eventos sonoros que ocorram separados no tempo por intervalos menores que o da persistência auditiva são percebidos pela audição como se ocorressem simultaneamente. O intervalo da persistência auditiva médio é 30ms.

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.

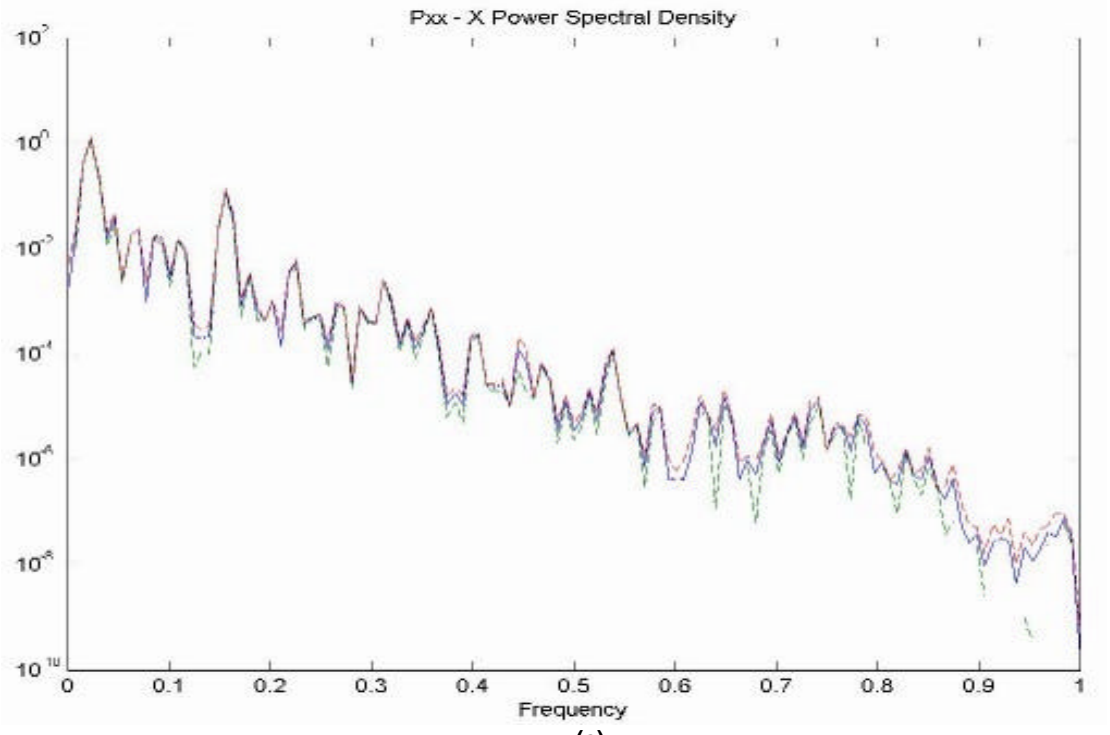
Já no nível de percepção microscópico, a audição reconhece o timbre, ou seja, as características estruturais do som, como seu ataque e a sua composição espectral. Através da percepção microscópica a audição reconhece, por exemplo, a diferença das vozes de indivíduos ou a diferença entre o som de instrumentos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

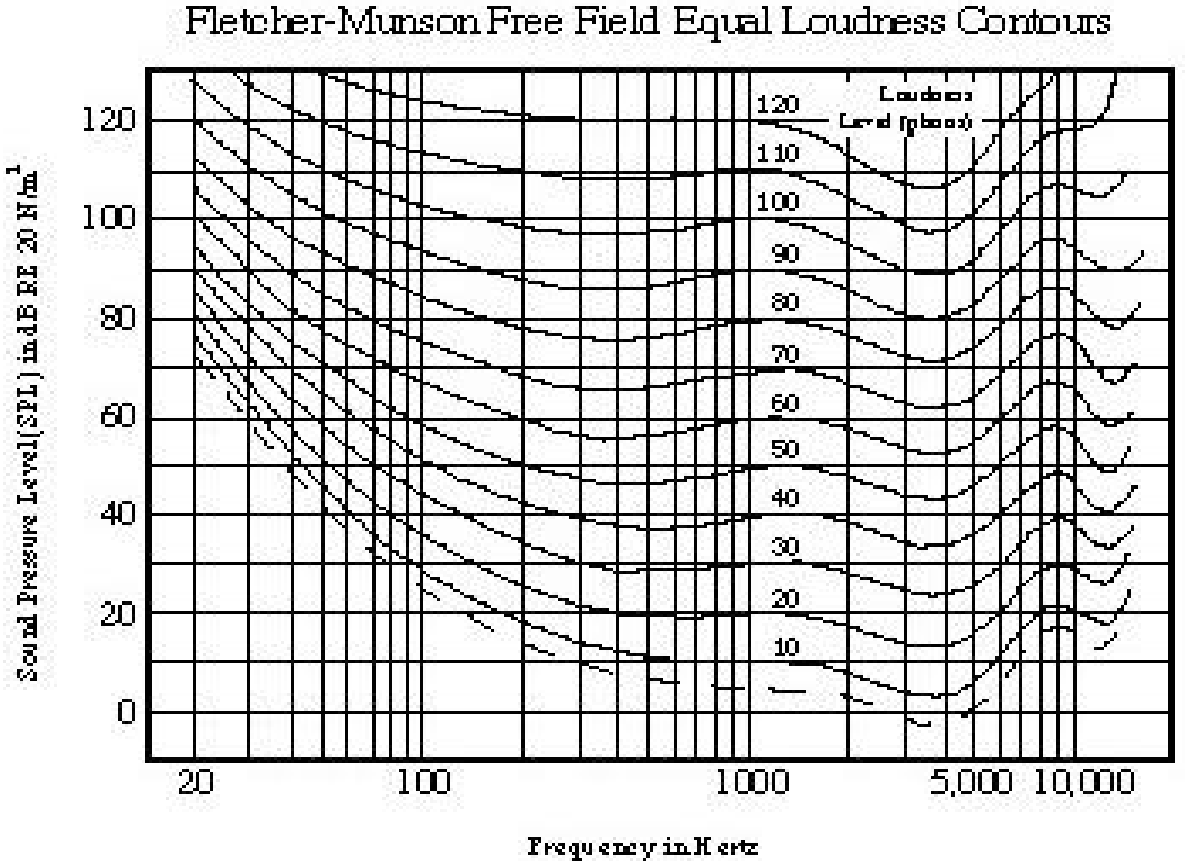


Figura A.5. Curvas Isofônicas de *Fletcher* e *Munson.*

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].

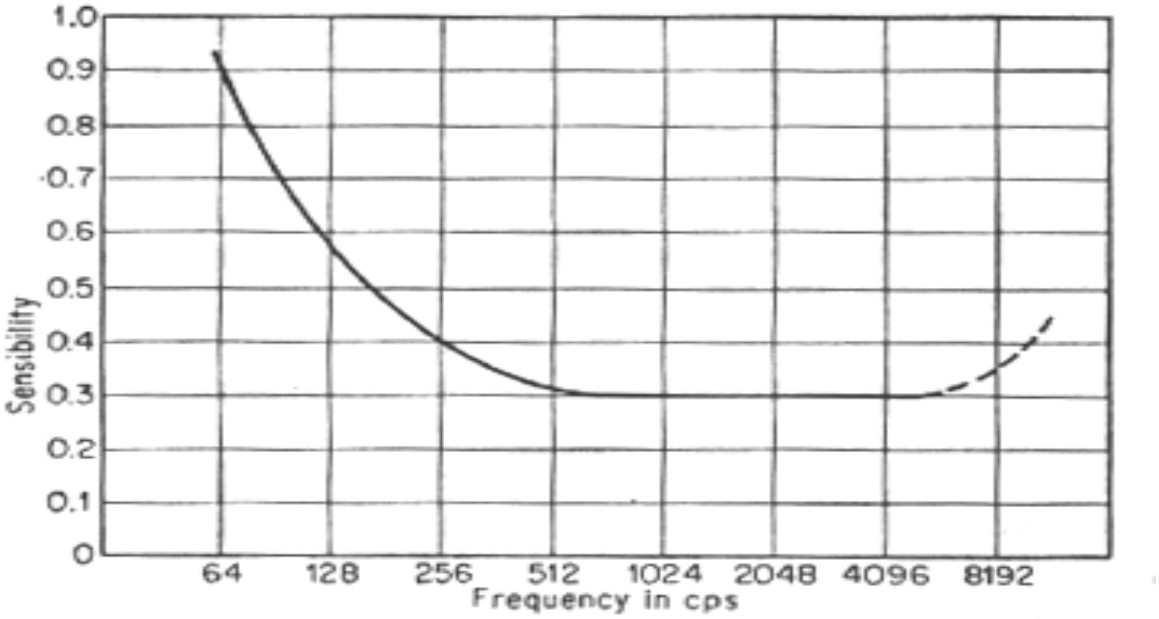


Figura A.6. Percepção da variação de freqüência. [Culver, 68].

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:

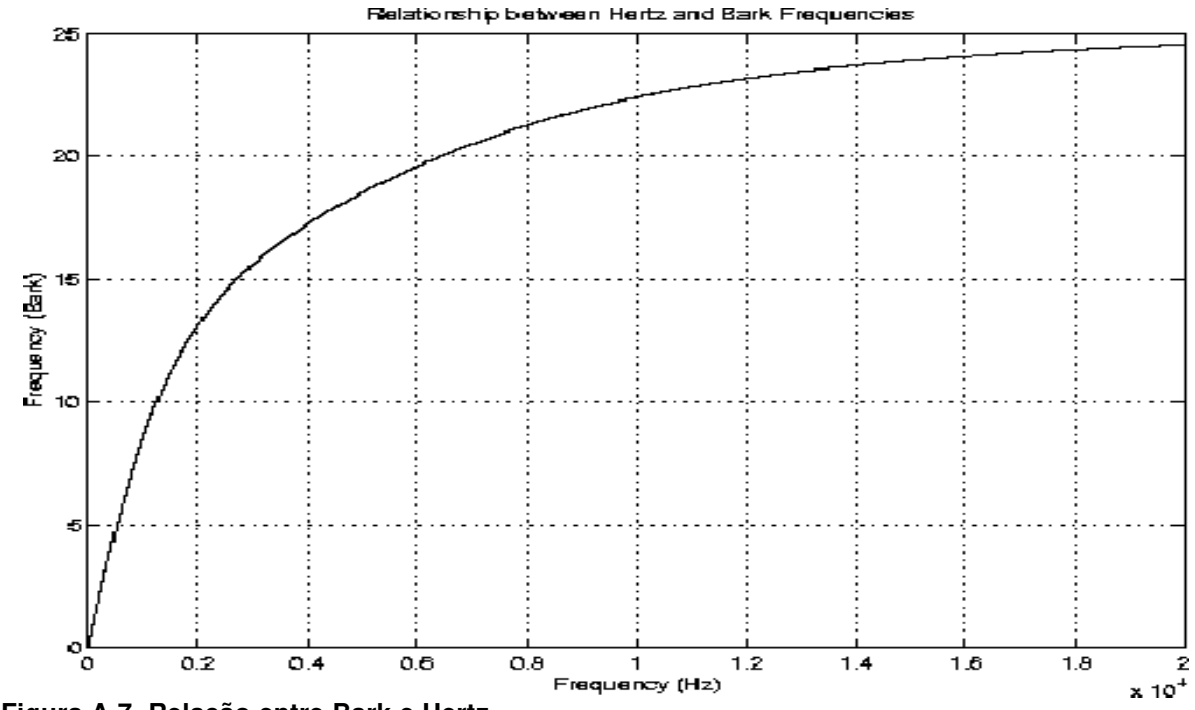


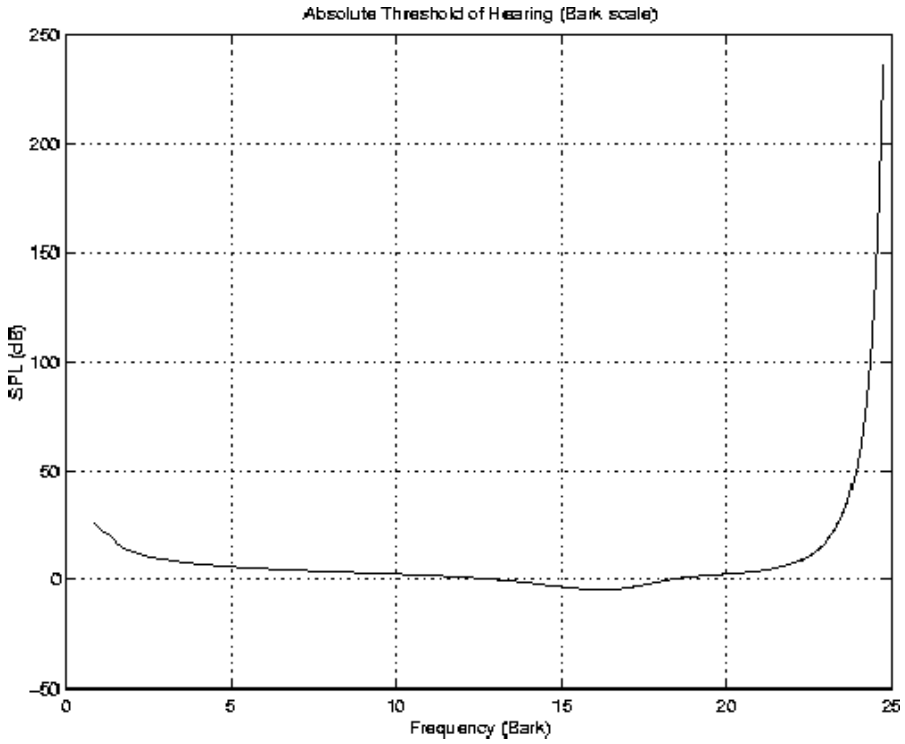
Figura A.7. Relação entre Bark e Hertz.

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

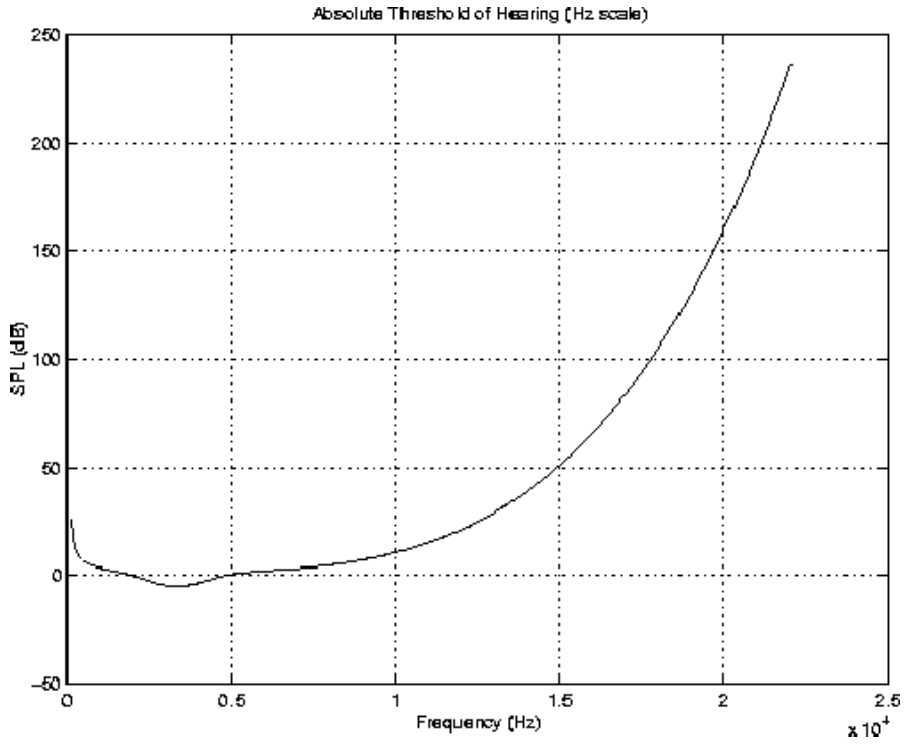
Zwicker & Terhardt (1980)	$B = 13.\tan^{-1}(0{,}76.f/1000) + 3{,}5.\tan^{-1}(f/7500)^2$ $B = 8{,}7 + 14{,}2.\log10(f/1000)$
Terhardt (1979)	$B = 13{,}3\tan^{-1}(0{,}75f/1000)$ $B = 12{,}82\tan^{-1}(0{,}78.f/1000) + 0{,}17(f/1000)^{1.4}$
Wang, Sekey & Gersho (1992)	$B = 6.\sinh^{-1}(f/600)$
Schroeder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

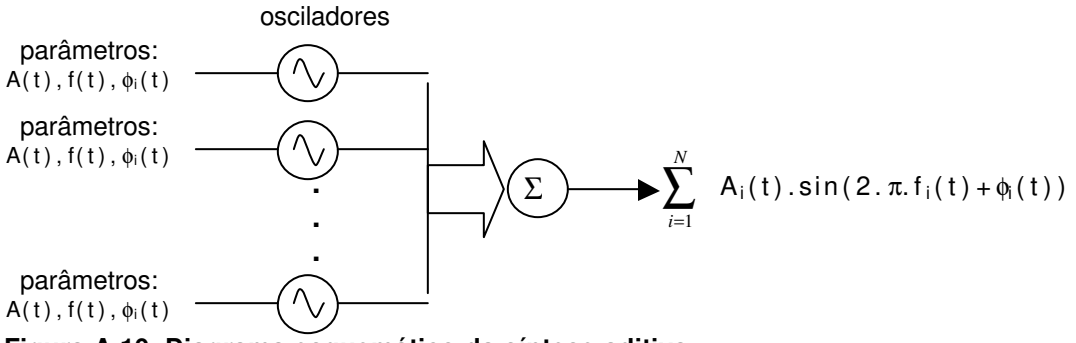


Figura A.10. Diagrama esquemático da síntese aditiva.

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.

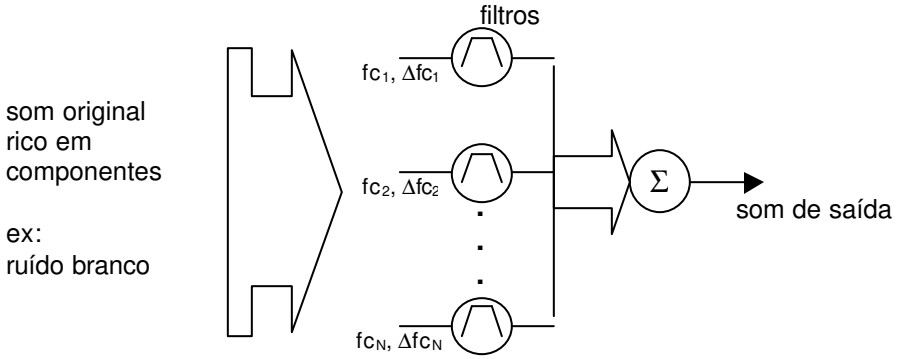


Figura A.11. Diagrama esquemático da síntese subtrativa.

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:

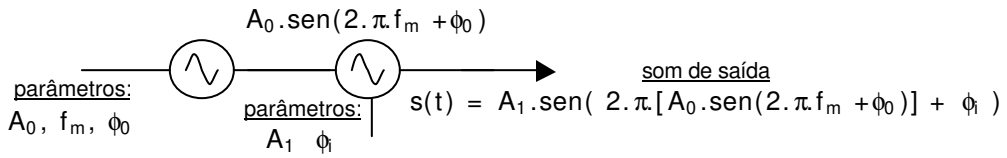


Figura A.12. Exemplo de um algoritmo de síntese FM com dois osciladores.

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.

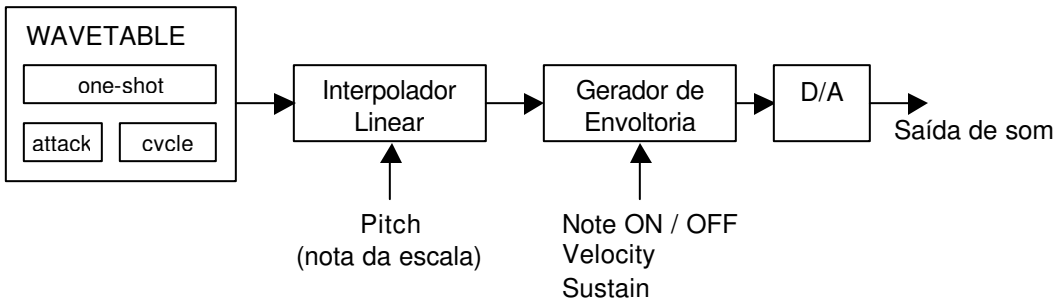


Figura A.13. Diagrama simplificado da síntese *Wavetable*.

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.

Referências Bibliográficas

[Angeline,94] Peter Angeline. "Coevolving High-Level Representations". 1994

[Begault,94] Begault, Durand R. 3-D Sound for virtual reality and multimedia. Durand R. Begault. 1957

[Bennet,02] Gerald Bennett. Notes on EM, Notes on Electroacoustic Music. 2002

[Berlyne, 66] Berlyne, D. Curiosity and exploration Science, vol 153, 23-23. 1996

[Biles,94] Biles, J. A., "Gen Jam: A Genetic Algorithm for Generating Jazz Solos", Proceedings of the 1994 International Computer Music Conference, (ICMC'94), 131-137, 1994.

[Cheung,96] Cheung, N. M., Horner, A., “Group Synthesis with Genetic Algorithms,” Journal of the Audio Engineering Society, 44(3): 130 –147, 1996.

[Chowning, 73] J. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," Journal of the Audio Engineering Society, vol. 21, pp. 526-534, 1973.

[Culver,68] Culver, C. A. "Musical Acoustics". 4°ed. cap. 5. McGraw-Hill. 1968.

[DeFatta,88] DeFatta, David J. Digital signal processing: cap. 6 e 9. A system design approach. David J. DeFatta, Joseph G. Lucas, William S. Hodgkiss. Singapore. John Willy & Sons Inc. 1988.

[Fletcher,33] Fletcher H. and Munson. "Loudness, its definition, measurement and calculation" J. Acoust. Soc. Am. 5, 82-108. 1933.

[Fogel,95] Fogel D. B.. “Evolutionary Computation - Toward a New Philosophy of Machine Intelligence”. IEEE Press, USA, 46 – 47, 1995.

[Foley,96] Foley J. D., Andries van Dam, Steven K. Feirner and John F. Hughes, "Computer Graphics Principles and Practice", Addison-Wesley Publishing Company, p. 1018, 1996.

[Fornari,01] Fornari, José, Jonatas Manzolli, Adolfo Maia, Furio Damiani. “The Evolutionary Sound Synthesis Method”, ACM Multimedia, Ottawa, Ontario, Canada, Setembro 2001.

[Fornari,01] Fornari, José, Jonatas Manzolli, Adolfo Maia, Furio Damiani. “Waveform Synthesis Using Evolutionary Computation”, Proceedings of the V Brazilian Symposium on Computer Music, Fortaleza, Ceará, BR, Julho 2001.

[Fornari,02] Fornari, José, Jonatas Manzolli, Marcio Costa, Fernando Ramos. “Solutions for Distributed Musical Instrument on the Web”, 2002.

[Fornari,95] Fornari, José, Furio Damiani. “Transformações Sonoras Através de Operações Timbrais”, Proceedings of the II Brazilian Symposium on Computer Music, Canela, RS, BR, Julho 1995.

[Fornari,95] Fornari, José, Marcelo Jara, Furio Damiani. “Reconhecimento de Timbres Musicais Através da Rede Neural Auto-Organizável de Kohonen”, Proceedings of the II Brazilian Symposium on Computer Music, Canela, RS, BR, Julho 1995.

[Fraser,59] Fraser. "Simulation of Genetic Systems by Automatic Digital Computers". Australian Journal of Biological Science, 10:484-499. 1959.

[Garcia,00] Garcia, Ricardo A. “Automatic Generation of Sound Synthesis Techniques”. Proposal for degree of Master of Science. MIT – Fall 2000.

[Homer,93] A. Horner, J. Beauchamp, and L. Haken, "Machine Tongues.16. Genetic Algorithms and Their Application to FM Matching Synthesis," Computer Music Journal, vol. 17, pp. 17-29, 1993.

[Horowitz,94] Horowitz D., “Generating rhythms with genetic algorithms”, Proceedings of the 1994 International Computer Music Conference, 142 – 143, 1994.

[Johnson,99] Johnson C. G.. “Exploring the sound-space of synthesis algorithms using interactive genetic algorithms in G. A.” Wiggins, editor, Proceedings of the AISB Workshop on Articial Intelligence and Musical Creativity, Edinburgh, 1999.

[Keltner,99] Keltner D and Gross. Functional accounts of emotions. Cognition and Emotion, 13(5). pp. 467-480. 1999.

[Koza,97] Koza, J. R., Bennett III, F. H., Andre, D., Keane, M. A., Dunlap, F., "Automated Synthesis of Analog Electrical Circuits by Means of Genetic Programming," IEEE Transactions on Evolutionary Computation, Vol. 1, NO. 2, July 1997.

[Koza,97] Koza, John R.. “Genetic Programming”. Encyclopedia of Computer Science and Technology. 1997.

[Lane,02] Richard D. Lane and Lynn Nadel. Cognitive Neuroscience of Emotion. Series in Affective Science. September 2002.

[Manzolli,99] Manzolli, J., A. Moroni, F. Von Zuben & R. Gudwin. 1999. “An Evolutionary Approach Applied to Algorithmic Composition”. Proceedings of the VI Brazilian Symposium on Computer Music, Rio de Janeiro, p. 201-210. 1999.

[Masri,98] Paul Masri, Nishan Canagarajah. "Synthesis From Musical Instrument Character Maps". Colloquium on "Audio and Music Technology". Institute of Electrical Engineers (IEE). November 1998, London. Digest No. 98/470. 1998.

[Menezes, Flo] Flo Menezes. “MÚSICA ELETROACÚSTICA: História e Estéticas”. Edusp. 1997.

[Menezes, Flo] Flo Menezes. “Atualidade Estética da Música Eletroacústica", Unesp. 1999.

[Miller,95] Geoffrey F. Miller, Peter M. Todd. The role of mate choice in biocomputation: Sexual selection as a process of search, optimization, and diversification. 1995.

[Miranda,01] Miranda, Eduardo Reck, "Composing music with Computers", Music Technology Series, 2001.

[Miranda,02] Miranda, Eduardo Reck. "Computer Sound Design: Synthesis Techniques and Programming". Focal Press. October 2002.

[Mitchell,93] Melanie Mitchell, Stephanie Forrest. "Genetic Algorithms and Artificial Life". 1993.

[Moore,89] Moore, Brian C. J., "An Introduction to the Psychology of Hearing". 3rd edition. Academic Press Limited. 1989.

[Moroni,00] Moroni, A., Manzolli, J., Von Zuben, F. & Gudwin, R., 2000, “Vox Populi: An Interactive Evolutionary System for Algorithmic Music Composition”. San Francisco, USA: Leonardo Music Journal - MIT Press, Vol. 10, p. 49-55. 2000.

[Mrazek,96] Pavel Mrazek. Genetic Algorithms and Image Search. Winter School of Computer Graphics 1996.

[Olsen,67] Olsen, Harry F.:Music, "Physics and Engineering", Dover Publications, Inc. N.Y., 2nd ed., 1967.

[Oppenheim,75] Oppenheim, Alan V. "Digital signal processing". A.V. Oppenheim, R.W. Schafer. Englewood Cliffs, N.J: Prentice-Hall. 1975.

[Plomp,70] Plomp R., "Timbre as a multidimensional attribute of complex tones. In frequency analysis and periodicity detection in hearing". Eds R. Plomp and G. F. Smoorenburg), Sijthoff, Leiden. 1970.

[Polansky,96] Polansky, L. 1996. “Morphological Metrics”. Journal of New Music Research, 25, pp. 289-368. 1996.

[Pollack,52] Pollack I., "The information of elementary auditory displays" J. Accoust. Soc. Am. 24, 745-749. 1952

[Pollock,97] John L. Pollock. Taking Perception Seriously. Proceedings of the First International Conference on Autonomous Agents. 1997.

[Risset,91] Risset, J.-C. 1991. “Timbre Analysis by Synthesis: Representations, Imitations, and Variants for Musical Compostion”. In Representations of Musical Signals, ed. De Poli, Piccialli & Road, Cambridge, Massachussetts: The MIT Press, ISBN 0-262-04113-8, pg. 7-43. 1991.

[Roads,94] Roads C. “Genetic Algorithms as a Method for Granular Synthesis Regulation”. Proceedings of the 1993 International Computer Music Conference, 1994.

[Rosemberg,82] Rosemberg, Martin E. "Sound and Hearing: Studies in Biology" n,145. cap. 2. Afd Sonologie. 1982.

[Sarkar,02] Biswajit Sarkar, Lokendra Kumar and Debranjan Sarkar. "A Genetic Algorithm-Based Approach for Detection of Significant Points for Polygonal Approximation of Digital Curves". 2002.

[Schaeffer,66] Schaeffer, P. “Traité des objets musicaux”.: Editions du Seuil. Paris 1966.

[Serra,89] Serra, Xavier, "A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic Plus Stochastic Decomposition", Dissertação de doutorado, CCRMA, Universidade de Stanford, CA, EUA, 1989.

[Smalley,90] Smalley, D. "Spectro-morphology and Structuring Processes In The Language of Electroacoustic Music", ed. Emmerson, pg. 61-93, 1990.

[Smalley,93] Smalley, D. 1990. “Spectro-morphology and Structuring Processes”. In The Language of Electroacoustic Music, ed. Emmerson, pg. 61-93. 1993.

[Smith,87] Smith, Julius, "Music Applications of Digital Waveguides", Reserch Sponsored by System Development Foundation, CCRMA, Stanford, CA, EUA, 1987.

[Smith,91] Smith, Julius O. "Physical Modeling Using Digital Waveguides." Computer Music Journal 16.4 (1992) :74-91. 1991.

[Smith,00] Julius O. Smith. "Viewpoints on the History of Digital Synthesis". CCRMA, Stanford University. 2000.

[Steiglitz96] Steiglitz, Ken. "A digital process primer, with applications to digital audio and computer music". Addison-Wesley Publishing Company. 1996.

[Stevens,80] Stevens, S.S. & Warshofsky, Fred., "Sound and Hearing", Time-Life Science Library, 1980.

[Zwicker,98] Zwicker, E., H. Fastl, "Psychoacoustics: Facts and Models", 2nd edition, ed.Springer-Verlag, 1998.