Campus de São Carlos

"SÍNTESE DE SONS MUSICAIS BASEADA NA TRANSFORMADA
DE KARHUNEN - LOÈVE"

Autor:

CARLOS ALBERTO YNOGUTI

Orientador:

Prof. Dr. MARCELO BASÍLIO JOAQUIM

UNIVERSIDADE DE SÃO PAULO

**ESCOLA DE ENGENHARIA
DE SÃO CARLOS**

Universidade de São Paulo
Escola de Engenharia de São Carlos
Departamento de Engenharia Elétrica

SÍNTESE DE SONS MUSICAIS
BASEADA NA TRANSFORMADA DE
KARHUNEN-LOÈVE

ENG. CARLOS ALBERTO YNOGUTI

Dissertação apresentada à Escola de Engenharia de São Carlos, da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de mestre em Engenharia Elétrica.

ORIENTADOR: Prof. Dr. Marcelo Basílio Joaquim

DEDALUS - Acervo - EESC

31100017052

São Carlos

1995

Class.	Tese-EESC
Curr.	6339
Tempo	074/95

Eng. Elétrica

074/898

**Ficha catalográfica preparada pela Seção de Tratamento
da Informação do Serviço de Biblioteca - EESC-USP**

Y535 Ynoguti, Carlos Alberto
Síntese de sons musicais baseada na transformada
de Karhunen-Loève / Carlos Alberto Ynoguti.--São
Carlos, 1995.
92p. + apêndice

Dissertação (Mestrado) -- Escola de Engenharia de
São Carlos-Universidade de São Paulo, 1995.
Orientador: Marcelo Basílio Joaquim

1. Síntese digital-sons musicais. I. Título

AGRADECIMENTOS

Ao Marcelo por suas valiosas idéias e excelente orientação durante todo o planejamento e execução deste trabalho.

À CAPES pela bolsa de estudos concedida.

Ao grupo de som do CAASO pelo empréstimo de material, tanto eletrônico como humano.

Aos professores e funcionários do Departamento de Engenharia Elétrica.

Aos colegas da pós-graduação pela força e amizade.

À Adriana, por seu carinho e grande apoio em todas as horas.

Aos amigos, sem os quais nada teria sentido.

*A meus pais Mituyosi (in memoriam)
e Clara, por tudo que fizeram por
mim*

FOLHA DE APROVAÇÃO

Dissertação defendida e aprovada em 10/3/1995
perante a Comissão Julgadora:

Prof. Dr. MARCELO BASILIO JOAQUIM - Orientador
(Escola de Engenharia de São Carlos - USP)

Prof. Dr. JOSÉ CARLOS PEREIRA
(Escola de Engenharia de São Carlos - USP)

Profa.Dra. ROSELI A. FRANCELIN ROMERO
(Instituto de Ciências Matemáticas de S.Carlos - USP)

Presidente da Comissão de Pós-Graduação
Prof. Dr. EDUARDO CLETO PIRES

Coordenador da área - Engenharia Elétrica
Prof. Dr. AMILCAR CARELI CESAR

SUMÁRIO

LISTA DE FIGURAS.....	i
LISTA DE TABELAS.....	v
RESUMO.....	vi
ABSTRACT.....	vii
1. INTRODUÇÃO.	
1.1. Motivações e objetivos.....	1
1.2. Organização do trabalho.....	2
2. SONS MUSICAIS.	
2.1. O que são sons musicais ?.....	4
2.2. Sons harmônicos e sons inarmônicos.....	5
2.3. Análise no domínio do tempo.....	6
2.4. Parâmetros de caracterização.....	9
2.5. Conclusões.....	12
3. SÍNTESE DIGITAL DE SONS MUSICAIS.	
3.1. Tipos de síntese.....	14
3.2. Tempos de processamento.....	15
3.3. Técnicas de síntese.....	16
3.4. Síntese de fórmula de somatório.....	16
3.5. Técnicas de síntese baseada em análise.....	19

3.5.1.	Síntese subtrativa baseada em análise.....	19
3.5.2.	Síntese aditiva baseada em análise.....	24
3.6.	Novas técnicas de síntese.....	28
3.6.1.	PARSHL.....	28
3.6.2.	SMS.....	29
3.6.3.	Transformada <i>wavelet</i>	29
3.6.4.	Síntese pelo uso de autômatos lineares (LASy).....	30
3.6.5.	FOF.....	31
3.6.6.	VOSIM.....	31
3.7.	Conclusões.....	32
4.	A TRANSFORMADA DE KARHUNEN-LOÈVE.	
4.1.	Introdução.....	33
4.2.	Derivação da transformada.....	33
4.3.	Propriedades da transformada de Karhunen-Loève.....	39
4.4.	Conclusões.....	39
5.	SISTEMA EXPERIMENTAL.	
5.1.	Introdução.....	41
5.2.	Aquisição de dados.....	42
5.3.	Visualização gráfica dos sinais.....	44
5.4.	Determinação da frequência fundamental do sinal.....	44
5.5.	Normalização em frequência.....	48

5.6.	Classificação dos sinais.....	50
5.7.	Alinhamento em fase.....	51
5.8.	Cálculo da base de Karhunen-Loève.....	53
5.9.	Deteção de envoltória.....	54
5.10.	Reconstrução dos sinais.....	66
5.11.	Volta à frequência original.....	68
6.	PARTE EXPERIMENTAL.	
6.1.	Testes iniciais	69
6.1.1.	Verificação da influência da normalização em frequência e do alinhamento em fase.....	70
6.1.2.	Verificação da influência da classificação dos sinais.....	73
6.1.3.	Apreciação dos resultados.....	77
6.2.	Testes de audição.....	78
6.2.1.	Material para os testes.....	79
6.2.2.	Procedimentos.....	79
6.2.3.	Apreciação dos resultados.....	80
7.	CONCLUSÕES.....	86
	REFERÊNCIAS BIBLIOGRÁFICAS.....	89
	APÊNDICE 1: Método dos multiplicadores de Lagrange.	
	APÊNDICE 2: Método das potências para o cálculo de ϕ_1.	
	APÊNDICE 3: Redução de uma matriz simétrica para a forma tridiagonal.	
	APÊNDICE 4: Autovalores e autovetores de uma matriz tridiagonal.	

LISTA DE FIGURAS

Figura 1. Espectro de um tom de clarineta. (harmônico).....	5
Figura 2. Espectro de um tom de tambor (inarmônico).....	6
Figura 3: Tom de flauta dividido nas partes de ataque, sustentação de decaimento.....	8
Figura 4: Tom de violão onde está ausente a parte de sustentação.....	8
Figura 5: Diagrama em blocos do modelo de síntese subtrativa.....	19
Figura 6: Sistema de análise heteródino.....	25
Figura 7: Tom de clarineta digitalizado e separado em harmônicas.....	27
Figura 8: Tom de clarineta linearizado por partes.....	27
Figura 9: Diagrama em blocos do sistema de análise/síntese.....	42
Figura 10: Estrutura do formato de áudio WAV.....	43
Figura 11. Exemplo de erro na detecção do período fundamental devido a variações na amplitude do sinal.....	45
Figura 12. Sinal original multiplicado por $y(n) = \exp(-Sn)$, onde $S = 0.004$	45

Figura 13. Ilustração de problemas de detecção de pitch devido a irregularidades na forma de onda.....	45
Figura 14: Alternativa para detecção do período fundamental na passagem por zero.....	47
Figura 15: Problemas na detecção do período fundamental somente através dos pontos de cruzamento.....	47
Figura 16: Ondas quadrada, senoidal e triangular com fases aleatórias.....	53
Figura 17: Ondas quadrada, senoidal e triangular alinhadas em fase.....	53
Figura 18: Função de frequência linear.....	56
Figura 19: Espectro depois da cisão harmônica.....	59
Figura 20: Envoltória de frequência para um tom de violão.....	64
Figura 21: Envoltória de fase para um tom de violão.....	65
Figura 22: Envoltória de amplitude para as 6 pimeiras funções base de um tom de violão.....	65
Figura 23. Tom de guitarra elétrica.....	67
Figura 24. Tom de guitarra elétrica reconstruído com 5 funções base.....	67

Figura 25: Comparação do erro quadrático médio de truncamento para os casos de sinais sem pré-processamentos, normalizados em frequência e alinhados em fase.....	71
Figura 26: Funções amostra da base formada com sinais sem pré-processamentos.....	71
Figura 27: Funções amostra da base formada com sinais normalizados em frequência.....	72
Figura 28: Funções amostra da base formada por sinais alinhados em fase.....	72
Figura 29: Comparação do erro quadrático médio de truncamento para verificar a influência da classificação dos sinais na concentração de energia nos primeiros autovetores.....	74
Figura 30: Funções amostra da base formada por sinais com correlação menor que 0.5.....	75
Figura 31: Funções amostra da base formada por sinais com correlação entre 0.5 e 0.9.....	75
Figura 32: Funções amostra da base formada por sinais com correlação maior que 0.9.....	76
Figura 33. Tom de violão reconstruído a partir de base formada por funções amostra sem pré-processamentos.....	77

Figura 34. Tom de violão reconstruído a partir de base formada por funções amostra normalizadas em frequência.....	77
Figura 35. Tom de violão reconstruído a partir de base formada por funções amostra alinhadas em fase.....	78
Figura 36. Tom de violão reconstruído a partir de base formada por funções amostra com correlação maior que 0.9.....	78
Figura 37: Espectro de um tom de guitarra durante o ataque.....	81
Figura 38: Espectro do tom reconstruído, durante o ataque.....	82
Figura 39: Espectro de um tom de guitarra no final da região de decaimento.....	82
Figura 40: Espectro do tom reconstruído no final da região de decaimento.....	82
Figura 41. Espectro do décimo autovetor da base onde foi reconstruído o tom de guitarra.....	83
Figura 42. Espectro do décimo primeiro autovetor da base onde foi reconstruído o tom de guitarra.....	83
Figura 43. Espectro do décimo segundo autovetor da base onde foi reconstruído o tom de guitarra.....	84
Figura 44. Espectro do décimo terceiro autovetor da base onde foi reconstruído o tom de guitarra.....	84

LISTA DE TABELAS

Tabela 1: Quadro comparativo para a verificação da influência da normalização em frequência e alinhamento em fase no desempenho da transformada.....	73
Tabela 2: Quadro comparativo para a verificação da influência da correlação entre as funções amostra no desempenho da transformada.....	76
Tabela 3: Numero suficiente de autovetores para síntese de alta qualidade para cada uma das classes.....	80

RESUMO

Tradicionalmente, a técnica de síntese aditiva com funções base de Fourier é a que tem apresentado melhores resultados no que diz respeito à qualidade dos sons gerados. Entretanto, a carga computacional imposta por esta técnica é extremamente alta, dificultando assim a sua implementação em tempo real. Substituindo as funções de Fourier por outras mais complexas, derivadas das técnicas estatísticas de Karhunen-Loève, consegue-se uma redução na quantidade de operações necessárias. Neste trabalho foi estudado e implementado em microcomputador um modelo de síntese aditiva baseada em análise utilizando a transformada de Karhunen-Loève.

Palavras-chave: síntese, sons musicais, Karhunen-Loève, processamento de sinais, áudio

ABSTRACT

Traditionally, the Fourier based additive synthesis is the method that have achieved the best results concerning to the quality of the generated sounds. However, the computational load imposed by this technique is extremely high, difficulting its real time implementation. Substituting the Fourier functions by another set of more complex functions, derived from the Karhunen-Loève statistical techniques, one achieves a reduction on the amount of the necessary operations so as to viabilize its real time implementation.

Keywords: synthesis, musical sounds, Karhunen-Loève, signal processing, audio.

1. INTRODUÇÃO.

1.1 Motivações e objetivos.

O estudo da música digital é uma área interdisciplinar que engloba várias áreas do conhecimento tais como a física, a computação, a psicologia, a eletrônica, e é claro, a música.

Com o avanço extraordinário experimentado pela tecnologia digital nas últimas décadas, está se tornando cada vez mais viável a sua utilização em aplicações musicais e de áudio em geral. Alguns exemplos de aplicações que estão sendo realizadas com esta tecnologia são gravações em estúdios totalmente digitais, síntese de sons musicais, escrita e execução de peças musicais utilizando o computador, uso do computador em composições musicais, e muitas outras mais.

Dentre todas estas áreas, este trabalho concentra-se no problema de análise e síntese de sons musicais. Assim, os objetivos a serem alcançados são: fornecer uma idéia do estado atual de desenvolvimento dos sistemas de análise e síntese digital de sons musicais, estudar e implementar uma das técnicas utilizadas para tal fim.

A técnica escolhida para ser estudada e implementada foi a de síntese aditiva baseada em análise, utilizando funções obtidas a partir da transformada de Karhunen-Loève.

A razão para a escolha do modelo aditivo baseado em análise é que este é o que tem apresentado a síntese de melhor qualidade dentre os métodos conhecidos (Moorer, 1977). A escolha das funções base de Karhunen-Loève deve-se ao fato de esta ser a

transformada ótima em relação ao erro quadrático médio de truncamento. Devido a esta propriedade, esta transformada é a que consegue a maior taxa de compressão entre todas as transformadas ortogonais, minimizando o grande esforço computacional tradicionalmente requerido pelas técnicas de síntese aditiva.

1.2. Organização do trabalho.

Esta dissertação é organizada da seguinte maneira:

No capítulo 2 trata-se dos sons musicais: o que são, e quais os parâmetros importantes para a sua análise.

No capítulo 3 é feito um levantamento sobre as técnicas de síntese mais utilizadas, suas características principais, bem como as vantagens e desvantagens de cada método.

No capítulo 4 discorre-se sobre a transformada de Karhunen-Loève: como é obtida, suas propriedades e as razões para a sua escolha neste trabalho.

No capítulo 5 é descrito o sistema experimental de análise\síntese desenvolvido, que consiste de vários programas para microcomputadores do tipo PC. O sistema de análise vai desde a aquisição dos sinais via conversor A/D até a geração dos parâmetros de síntese para os sinais sob análise. O sistema de síntese consiste na reconstrução dos sinais a partir dos dados obtidos pelo sistema de análise. A fase de análise dos sons musicais inclui ainda alguns pré-processamentos dos sons musicais, tais como normalização em frequência, alinhamento em fase e classificação dos sinais. Para cada uma destas tarefas foi também desenvolvido um programa específico.

No capítulo 6 são apresentados os resultados obtidos com a análise e síntese de 90 sons colhidos de seis instrumentos diferentes. Os testes foram divididos em duas etapas. Na primeira foram realizados ensaios para verificar a influência dos pré-processamentos de normalização em frequência, alinhamento em fase e classificação dos sinais no desempenho da transformada. Os critérios de desempenho adotados foram o erro quadrático médio de truncamento e a porcentagem da energia total do sinal

armazenada no primeiro autovetor. Em uma segunda etapa foi verificado o desempenho do sistema na reconstrução de tons inteiros. A avaliação foi realizada através de testes de audição.

No capítulo 7 são feitas as conclusões gerais sobre o trabalho. Os resultados dos ensaios são avaliados, analisando-se a influência dos pré-processamentos no desempenho do sistema, bem como a sua adequação para a produção de sons musicais. São levantadas também as restrições e vantagens da técnica de KL em comparação com a de Fourier.

2. SONS MUSICAIS.

2.1. O que são sons musicais ?

De modo a fornecer subsídios para analisar a escolha do método empregado e apreciar os resultados obtidos, será feita neste capítulo uma breve introdução sobre a natureza dos sons musicais, suas características principais e particularidades.

Tradicionalmente são considerados como sendo sons musicais aqueles produzidos pelos instrumentos de orquestra, e pela voz humana quando cantando. Os instrumentos de orquestra são divididos em três grupos: cordas, sopros e percussão. Em cada um destes grupos também existem subdivisões: os instrumentos de corda são divididos em instrumentos de arco (violino, viola, violoncelo, contrabaixo), percutidos (piano) e beliscados (harpa, cravo). Dentre os instrumentos de sopro temos os de palheta (saxofone, oboé) e os de embocadura (flauta, flautim). Para os instrumentos de percussão, tem-se os de pele (tímpanos, caixa), de metal (prato, gongo, triângulo) e os de madeira (marimba). Além dos instrumentos de orquestra existem ainda muitos outros tais como o violão, o cavaquinho, o bandolim, o pandeiro, só para citar alguns. Também existem aqueles próprios de regiões específicas do mundo, tais como o banjo norte-americano, o okotô e o shamisen japoneses, o ukelele havaiano e a guitarra portuguesa entre tantos outros. Com o desenvolvimento da tecnologia e o advento de novos estilos musicais, surgiram vários instrumentos novos, tais como a guitarra e o contrabaixo elétricos.

Apesar da grande variedade dos sons produzidos pelos instrumentos musicais, pode-se definir um som musical, pelo menos os mais comuns, como um sinal complexo que apresenta uma altura determinável e um conjunto de parciais (não necessariamente harmônicos) relacionadas e dependentes do tempo, variáveis tanto em intensidade como em frequência (Pollard, 1988).

2.2. Sons harmônicos e sons inarmônicos.

Segundo Moorer (1977), os sons musicais podem ser classificados *em quasi-periódicos* (ou harmônicos) e *aperiódicos* (ou inarmônicos). A maioria dos instrumentos de orquestra são quasi-periódicos. Isto significa que eles podem ser descritos por uma soma de um número finito de senóides. A natureza quasi-periódica é modelada fazendo-se com que tanto a amplitude quanto a frequência de cada harmônica variem no tempo de maneira independente. Nas figuras 1 e 2 tem-se um exemplo do espectro para cada um destes tipos de sons. Na figura 1 é mostrado um som de clarineta (harmônico) e na figura 2, um som de tambor (inarmônico).

Figura 1. Espectro de um tom de clarineta. (harmônico)

Figura 2. Espectro de um tom de tambor (inarmônico).

A maioria dos instrumentos de percussão (tambores, sinos) são inarmônicos. Isto significa que para descrevê-los com senóides geralmente requer-se um número grande delas em comparação com sons harmônicos. Um modelo que funciona bem com sons inarmônicos é o chamado subtrativo, onde os sons são descritos através de uma função de excitação aplicada a filtros variantes no tempo. Como será visto na seção seguinte, esta técnica trata tanto com sons harmônicos como inarmônicos indiferentemente, o que facilita o seu uso com sons inarmônicos.

2.3. Análise no domínio do tempo.

Dentre os sons audíveis ao ouvido humano, os tons musicais formam uma classe especial que difere marcadamente em estrutura tanto dos tons puros como do ruído, dois tipos de sinais que têm sido alvo de estudos intensos no passado. Os métodos de análise concebidos tanto para tons puros como para o ruído são geralmente inadequados para lidar com a análise de sons musicais pelo fato destes apresentarem uma estrutura complexa, com parâmetros que variam tanto no domínio do tempo como no da frequência.

No domínio do tempo, um som musical consiste essencialmente de três partes (Pollard & Jansson, 1982):

1. o **ataque**, com duração entre 5 ms e 350 ms, no qual ocorrem complicados processos de interferência entre as vibrações livres dos modos do instrumento e o movimento forçado imposto pela fonte de energia, juntamente com algum ruído instrumental de fundo que é geralmente um elemento importante na determinação das características gerais do som;
2. uma **região de sustentação**, correspondente ao movimento forçado do instrumento com duração entre 100 ms e vários segundos. (Esta região consiste geralmente de um conjunto de parciais que podem estar ou não relacionadas harmonicamente);
3. o **decaimento** de vibrações livres dos modos do instrumento quando a função forçante é removida. O padrão de decaimento é geralmente definido por interferências entre os modos do instrumento e os da sala. Sob condições normais, os modos da sala predominam durante esta fase, estendendo o decaimento aparente por um ou mais segundos.

Na figura 3 é mostrado o desenvolvimento temporal de um tom de flauta, onde podem ser vistas as regiões de ataque, sustentação e decaimento.

Alguns sons musicais não apresentam todas estas três regiões. Para alguns tipos de sons (percussivos por exemplo) a região de sustentação não existe. Nestes existe um ataque rápido, seguido de um decaimento que depende das características da parte ressonante do instrumento. Exemplos de sons percussivos são os produzidos por tambores, pianos, violões, entre outros. Na figura 4 é mostrado, um som de violão, onde pode ser notada a ausência da região de sustentação.

Figura 3: Tom de flauta dividido nas partes de ataque, sustentação e decaimento.

Figura 4: Tom de violão onde está ausente a parte de sustentação.

2.4. Parâmetros de caracterização.

Costuma-se também descrever um som musical em termos de quatro características principais:

1. **duração**, que envolve a importância relativa entre o transiente inicial, a região de sustentação e o decaimento;
2. **volume**, que é relacionado à intensidade do som (e é uma medida da potência elétrica transmitida pelo nervo auditivo);
3. ***pitch***, que envolve uma avaliação em uma escala unidimensional relacionada à frequência;
4. **timbre**, que é um parâmetro multidimensional relacionado ao nível das componentes espectrais.

A duração e o volume são grandezas que podem ser medidas diretamente através de aparelhos adequados. A duração pode ser dada em unidades de tempo (s) e o volume em unidades de pressão sonora (W/m^2 ou dB SPL). Em relação ao volume, vale ressaltar que o ouvido humano não tem resposta plana dentro da faixa de sons audíveis: ele é mais sensível na faixa de frequências médias (em torno de 3 kHz) do que nas outras regiões do espectro. Desta maneira, a sensação de volume depende também da frequência do som. Já o *pitch* e o timbre são grandezas mais complexas que não dependem de apenas um parâmetro, mas de uma série destes, embora, no caso do *pitch*, a frequência seja o fator mais importante. Estes dois parâmetros merecem uma atenção especial.

Segundo Stoll & Parncutt (1987), o *pitch* é definido como a altura percebida de um som. Ao contrário da frequência, o *pitch* não é mensurável fisicamente. Não existe uma relação simples entre as frequências espectrais de um som e o *pitch* que ele evoca. Apenas alguns *pitch*'s evocados por um som são realmente notados, geralmente aquele correspondente à sua frequência fundamental. A percepção do *pitch* de um som

complexo envolve um processo de reconhecimento de padrões para o qual a entrada consiste nas frequências das suas componentes perceptíveis.

A frequência fundamental de um tom complexo inteiro ou limitado em banda pode ser determinada a partir da diferença entre as frequências de duas harmônicas sucessivas quaisquer. Se, entretanto, é sabido que algumas harmônicas podem estar ausentes ou completamente mascaradas por outros sons, qualquer submúltiplo do máximo divisor comum das frequências componentes torna-se um candidato para a fundamental. A situação torna-se mais incerta quando se sabe que mais de um som pode estar soando simultaneamente. Neste caso, qualquer par de frequências poderia ser parte de uma série harmônica baseada em qualquer dos divisores comuns daquelas frequências.

O timbre é definido geralmente como o atributo que permite a um ouvinte discriminar dois sons com o mesmo *pitch* e o mesmo volume. Esta definição, no entanto, deixa as bases acústicas do timbre quase completamente indefinidas. Qualquer atributo acústico que não contribua exclusivamente para a percepção do *pitch*, volume ou duração poderia contribuir para a percepção do timbre.

Algumas pesquisas de percepção têm tentado determinar quais destes possíveis atributos acústicos são os mais importantes usando técnicas de julgamento subjetivo de similaridade (Iverson & Kruhmansl, 1993). Estes estudos permitiram que os pesquisadores isolassem um pequeno número de fatores acústicos que seriam mais importantes para a percepção do timbre.

Em todos os experimentos de julgamento de similaridade com sons musicais naturais, as amplitudes relativas das harmônicas de baixa e alta frequência no espectro estático tem contribuído para os julgamentos de similaridade. Grey e Gordon (1978) foram bem sucedidos na quantização de alguns destes atributos acústicos. Eles descobriram que a frequência média (centróide) do espectro tem uma alta correlação com os julgamentos de similaridade. Então a frequência centróide é uma das principais responsáveis pela percepção do timbre.

Outros estudos mostram que as variações temporais dos sons também contribuem para a percepção do timbre. Grey (1977) encontrou dois atributos dinâmicos que

mostraram contribuir para julgamentos de similaridade: a sincronia espectral, a flutuação dos tons (mais comumente nos ataques), e a presença ou ausência de energia de baixa amplitude e alta frequência nos ataques. Wessel¹ (1979) apud Iverson & Krumhansl (1993) descobriu que os julgamentos de similaridade estavam relacionados com a mordida dos ataques. Embora estes atributos dinâmicos não sejam consistentes ou fáceis de quantizar, parece que os ataques desempenham um papel particularmente importante na percepção do timbre.

Um outro tratamento para o problema foi dado por Preis (1984). Ele acredita que é possível determinar um único parâmetro como sendo uma representação quantitativa do timbre. O seu raciocínio baseia-se no fato de que tanto a intensidade como o *pitch* dependem de mais de um parâmetro, entretanto apenas um destes é o suficiente para determiná-los. Deste modo, uma aproximação análoga para o timbre não seria injustificada. Poderia-se assumir que dentre os parâmetros reconhecidos como responsáveis pela sensação do timbre, um desempenha o papel mais importante, isto é, a sua mudança seria a principal responsável pela mudança do atributo da sensação auditiva. Se isto for aceito, poderia-se apontar um certo parâmetro, o qual seria uma estimativa confiável de tons complexos. Poderia-se investigar como a variação deste parâmetro é percebida e deste modo estimar a relação entre a variação numérica e a sensação auditiva. A influência dos outros parâmetros do timbre poderia ser levada em conta mais tarde. O procedimento poderia ser similar ao da pesquisa sobre o volume, onde primeiramente leva-se em conta a influência da intensidade e somente depois é considerado que a mudança de volume depende da mudança da frequência. Segundo Bruijn (1978), esta grandeza física estaria relacionada à inclinação da envoltória espectral.

Como pode-se ver, a descrição de sons musicais a partir destes quatro parâmetros (duração, volume, *pitch* e timbre) é insuficiente para distinguir um som musical de outros tipos de sons. Existem flutuações dinâmicas dentro de cada nota (conhecidas como

¹ WESSEL, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 3, 45-52. apud IVERSON, Paul and KRUMHANS, Carol L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of Acoustical Society of America*, Vol. 94, No. 5, November. p. 235.

microestrutura do som), que podem representar um papel vital na avaliação de um som musical.. Alguns exemplos disso são:

1. Variações no nível da envoltória de tempo dos tons parciais;
2. Variações tanto no nível como na frequência das componentes espectrais;
3. O aparecimento temporário de componentes espectrais inarmônicos;
4. A presença de componentes de ruído instrumental;
5. A natureza dos transientes iniciais.

Em relação à composição musical como um todo, Seashore² (1938) apud Pollard (1988) mostrou que não se tem consciência dos elementos tonais individuais ou ondas sonoras como tal, mas do padrão ou impressão musical como um todo. Nesta mesma liha de pesquisa, Grey (1978) mostrou que a percepção do timbre é diferente em situações de tons isolados e em situações de contextos musicais, ou seja, quando são tocadas notas em sequência, isoladas ou em acordes: o fator que influencia a percepção do timbre para tons isolados parece ser a sua variação temporal, enquanto que em contextos musicais, o fator mais importante é o seu conteúdo harmônico.

2.5. Conclusões.

A complexidade dos fatores individuais envolvidos nos julgamentos relacionados a todos os valores musicais significa que, pelo menos no estado de conhecimento atual, os esforços devem ainda ser direcionados para a avaliação das impressões subjetivas de ouvintes hábeis. A disponibilidade de equipamentos de análise digital rápidos não significa que números são substitutos confiáveis para os julgamentos informais.

Das considerações acima pode-se concluir que, para uma síntese de alta qualidade devem ser levados em conta inúmeros fatores além da forma de onda do sinal. Se a

² SEASHORE, C. E. (1938). Psychology of music. McGraw-Hill. New York. apud POLLARD, H. F. (1988) Feature analysis of musical sounds. *Acustica*, No 65, p. 232.

síntese for realizada colando-se vários períodos da forma de onda do sinal um atrás do outro, o som gerado certamente será bem diferente do original. Este procedimento ignora uma das principais características dos sons musicais, que é o fato de estes serem sinais variantes no tempo.

Levando-se em conta que a consideração de todos os fatores levantados é uma tarefa extremamente dispendiosa e complicada, é razoável pensar em uma aproximação que despreze alguns destes fatores que tenham uma menor influência na qualidade do timbre ou que sejam mais complicados de determinar. Uma das aproximações possíveis é considerar apenas as variações temporais de amplitude e frequência. Com isto, consegue-se uma síntese de ótima qualidade, e com um custo computacional não muito elevado. Este procedimento preserva a porção de ataque dos sons que, como visto acima, desempenha um papel fundamental na determinação do timbre.

Esta aproximação, foi utilizada neste trabalho, e foram conseguidos resultados muito bons a nível de similaridade entre o sinal original e o sintetizado.

3. SÍNTESE DIGITAL DE SONS MUSICAIS.

O objetivo dos métodos de síntese digital é produzir um sinal amostrado que quando convertido em uma tensão analógica e tocado através de um sistema de som de alta fidelidade produza o som musical desejado, com a maior fidelidade possível. O problema então é sintetizar este sinal de dados amostrados. Isto é feito utilizando várias técnicas fundamentais de síntese diferentes.

Uma técnica de síntese pode ser entendida como uma fórmula para produzir as amostras de um sinal amostrado. Esta expressão tem geralmente muitas variáveis, tais como frequência, duração e volume. Para que haja um maior realismo na reprodução, é necessário que estas variáveis mudem com o tempo durante a duração do som. Isto significa que estas variáveis precisam ser funções do tempo.

O grau em que o som sintetizado se iguala ao som original é uma medida da validade do modelo e da técnica de estimação. Se o modelo não é válido para o som em questão, ou se a técnica de estimação não é adequada, não se pode esperar reproduzir o som original.

3.1. Tipos de síntese.

Podem ser identificados três tipos de síntese: direta, baseada em análise e *musique concrète*.

Na síntese direta, o compositor senta-se em um computador e escreve um programa que produza as amostras da forma de onda do som desejado. Neste tipo de síntese, as funções

de controle são introduzidas à mão pelo músico ou programador, usualmente como funções convenientemente lineares por partes.

Na síntese baseada em análise, tons naturais existentes são digitalizados e analisados. Os resultados da análise são usados como dados para dirigir a síntese. Nesta, as funções de controle são formas de onda amostradas processadas a partir dos sons dos instrumentos musicais existentes. Estas podem então ser modificadas pelo compositor. A expressão de síntese é um modelo matemático da forma de onda. Na síntese baseada em análise o procedimento de análise é uma tentativa de estimar os parâmetros do modelo para um dado som.

A *musique concrète* foi criada por Pierre Schaefer, um radialista francês, no final da década de 40. Consiste em digitalizar sons naturais e usar o computador para processá-los de várias maneiras sem realmente analisar e ressintetizar como na forma anterior. Exceto pela tremenda quantidade de dados envolvidos, esta é provavelmente a forma mais simples de síntese.

3.2. Tempos de processamento.

Os sistemas de síntese podem também ser classificados pelo seu tempo de processamento. Este pode ser *off-line*, iterativo ou tempo real. O sistema ideal de síntese musical deveria ser totalmente em tempo real, onde o compositor poderia ajustar os parâmetros de síntese dinamicamente, e ouvir o som gerado, tocando-o através de um teclado de órgão ou qualquer um de outros circuitos de entrada em tempo real. Desde que a síntese (e a linguagem) de música de alta qualidade requer um grande esforço computacional, este parece ser um objetivo impossível sem a ajuda de hardware especial.

No sistema *off-line*, o processamento computacional é feito em um dado instante e então a conversão digital para analógica é feita nas amostras acumuladas. Este é um meio bastante difícil para trabalhar! Desde que o critério de qualidade é quase inteiramente

subjetivo, é necessário um grande esforço de experimentação para se chegar aos sons desejados.

No modo iterativo, computa-se um som curto e este é tocado rapidamente, os parâmetros são reajustados e nova tentativa é realizada. É um método intermediário entre o *off-line* e o tempo real, e que é geralmente utilizado quando a capacidade do equipamento é limitada.

3.3. Técnicas de síntese.

Frequentemente distingue-se os processamentos em técnicas aditivas, técnicas subtrativas e técnicas de modulação (não linear).

Na síntese aditiva, uma forma de onda composta é produzida pela soma das partes componentes, por exemplo, somando-se as harmônicas de um tom para produzir um único som. Pelo fato de as componentes mais usadas serem senóides, esta técnica é às vezes chamada, um tanto equivocadamente, síntese de Fourier. Na síntese subtrativa toma-se uma forma de onda espectralmente rica e eliminam-se as partes não desejadas através de filtros. Este modelo é bastante popular em sintetizadores musicais analógicos convencionais. Dentre as técnicas não lineares, a mais importante é a classe conhecida como fórmulas de somatório. Estas são expressões em forma fechada de várias somas trigonométricas. Isto permite gerar um conjunto inteiro de parciais senoidais resolvendo uma única expressão às vezes de uma simplicidade incrível.

3.4. Síntese de fórmula de somatório.

Esta forma de síntese não requer análise, um fato que é tanto uma vantagem quanto uma desvantagem. A vantagem é que não é necessário nenhum equipamento de conversão analógico-digital, de modo que este método pode ser aplicado a sistemas bem pequenos. A extrema simplicidade de algumas destas fórmulas faz com que sejam

bastante atrativas quando a potência do computador é limitada. Este método é facilmente aplicável na técnica de síntese direta. A desvantagem é que desde que não é usada nenhuma análise, não se pode esperar que estes tons reproduzam fielmente os sons naturais.

O nome fórmula de somatório vem do modo como as fórmulas são derivadas. Quase todas vêm da soma de séries, um tópico que foi estudado extensivamente por Stirling no séc. XVIII (Stirling¹, 1749 *apud* Moorer, 1977).

Esta classe de técnicas de síntese pode ser vista como um novo instrumento musical o qual é necessário aprender a tocar de uma maneira análoga a aprender a tocar um instrumento qualquer. É necessário aprender quais valores dos parâmetros de controle produzem que tipos de sons. Desde que não existe análise possível, o instrumento precisa ser tratado de uma maneira empírica. As suas capacidades devem ser exploradas fazendo experiências com vários valores para os parâmetros de controle, construindo lentamente um corpo de conhecimento sobre como obter tipos de sons diferentes.

O pai destes métodos é a equação de modulação em frequência a qual foi primeiramente explorada em suas qualidades tímbricas por Chowning (1973). Quando apresentada como uma técnica de síntese, a forma de onda de saída é calculada pela seguinte fórmula:

$$X(n) = A(n) \sin\{2\pi f_c nT + I(n) \sin(2\pi f_m nT)\} \quad (3.4.1)$$

onde $A(n)$ é a amplitude da forma de onda. Em muitos casos, isto será uma função simples do tempo variando lentamente. Ela poderá ter uma porção de ataque, uma porção de sustentação e uma porção de decaimento muito parecidas com as formas de onda de instrumentos musicais comuns. $I(n)$ é o índice de modulação; ele é mostrado aqui como sendo uma função do tempo. Como a amplitude, ele é geralmente simples e varia

¹ STIRLING, J. (1749) The differential Method. London. *apud* MOORER, James A. (1977) Signal processing aspects of computer music - a survey. *Proceedings of the IEEE*, vol 8 n 8. August. p. 1977.

lentamente. f_c é chamada de frequência da portadora. f_m é chamada de frequência de modulação. T é o tempo entre amostras. Isto nos diz que o tom é completamente determinado pela escolha das funções $A(n)$ e $I(n)$, e pela escolha dos outros parâmetros f_c , f_m e T .

É interessante notar que estas fórmulas são conhecidas há bastante tempo. Apenas recentemente têm sido aplicadas para produzir sons musicais. Estudantes de engenharia aprendem as equações de modulação em frequência todos os anos, mas poucas pessoas pensaram em alguma vez ouvir o som da forma de onda modulada. Embora a modulação em frequência tenha sido usada por músicos com sintetizadores analógicos, a precisão dos osciladores controlados por tensão não é suficiente para produzir tons harmônicos. Existem também casos em que a frequência torna-se negativa o que ocorre quando o índice de modulação é suficientemente alto. Com processamento digital, estas características são bastante fáceis de se obter.

Pela sua fácil implementação em hardware e pelos sons diferentes gerados por esta técnica, esta técnica revolucionou o mercado de sintetizadores no início da década de 80 com a família DX da empresa japonesa YAMAHA, notadamente com o DX7, que foi o sintetizador mais usado pelos grupos de rock e música pop desta época. Atualmente, a técnica de FM é a mais utilizada na geração de sons em computadores, notadamente em aplicações multimídia.

Embora seja a mais famosa, a técnica de FM não é a única. Existem várias outras fórmulas variantes ou não da equação de modulação em frequência, que são utilizadas para síntese de sons musicais e que caem dentro desta denominação de fórmulas de somatório. Dentre estas destacam-se as equações que utilizam as funções de Bessel. Apesar de serem mais complicadas que a equação de FM, mantém ainda a simplicidade conceitual.

3.5. Técnicas de síntese baseada em análise.

Nas técnicas de síntese baseada em análise, os parâmetros são obtidos através da análise de sons musicais já existentes, eliminando o processo, por vezes cansativo, de tentativa e erro utilizado pelas técnicas de fórmula de somatório. Ainda, com a análise, consegue-se reproduzir os sons originais com muito mais fidelidade e qualidade.

3.5.1. Síntese subtrativa baseada em análise.

Na síntese subtrativa, a forma de onda do som é modelada com uma função de excitação enviada através de um filtro variante no tempo, como pode ser visto na figura 5.

Figura 5: Diagrama em blocos do modelo de síntese subtrativa.

Esta técnica tem sido usada extensivamente para produzir formas de onda de voz sintéticas. Para formas de onda de voz, a excitação escolhida pode ser uma forma de onda periódica como um trem de pulsos, ou um ruído branco. A forma de onda periódica é usada durante a linguagem expressiva, tais como vogais. A forma de onda do ruído é usada para a fala não expressiva, tais como sons mudos. O filtro simula então as ressonâncias do trato vocal. Este precisa mudar com o tempo, pois as ressonâncias do trato vocal mudam com as diferentes posições da língua, lábios, etc. Este modelo pode ser representado pela seguinte equação:

$$X(n) = -\sum_{k=1}^p a_k X(n-k) + G \sum_{r=0}^q b_r U(n-r) \quad (3.5.1.1)$$

onde a_k e b_r são os coeficientes do filtro, por convenção, b_0 é tomado como sendo 1, $X(n)$ é a forma de onda da saída, G é o fator de ganho global, e $U(n)$ é a forma de onda da entrada, usualmente um trem de pulsos ou ruído branco.

Para sintetizar um som é necessário então determinar em cada instante que tipo de excitação usar, e as características dos filtros. A escolha de $U(n)$ e G determina a excitação, e a escolha de a_k e b_r , as características dos filtros. Esta determinação é geralmente feita pela análise da forma de onda do som a ser sintetizado. Se a excitação é modelada por um trem de pulsos é necessário determinar a sua frequência fundamental em cada instante.

Um método popular para escolher o filtro é descrito em Makhoul (1975) que é baseado em calcular um filtro cuja resposta em frequência combine melhor com o espectro da fala em cada instante no sentido de mínimo erro quadrático médio. O meio mais simples de calcular os coeficientes do filtro assume que o modelo é somente de polos, ou seja, $b_r = 0$ para $r = 1, 2, \dots, q$ (análise preditiva linear).

Uma das vantagens deste tipo de técnica na produção de fala é que é possível sintetizar uma declaração em taxas diferentes simplesmente atualizando os parâmetros (frequência, coeficientes do filtro), mais rápido ou mais devagar do que eles ocorrem na declaração original. Pode-se também sintetizar a declaração em uma frequência fundamental diferente, mas com a temporização original, simplesmente multiplicando todos os tons por um fator de escala fixo.

Se alguém toma uma fala gravada e meramente acelera ou desacelera a forma de onda, a frequência fundamental é modificada assim como a qualidade do tom. Analisando primeiramente a fala, pode-se mudar o tempo ou a frequência fundamental independentemente. Embora não seja fácil manter a qualidade do tom exatamente constante enquanto se muda a frequência fundamental da declaração sintética, ela é muito mais constante usando este sistema de análise-síntese do que, por exemplo, acelerando ou desacelerando um gravador de fita.

Pode-se também ressintetizar a declaração usando uma forma de onda complexa como excitação ao invés de um simples trem de pulsos. Isto pode produzir uma variedade dos

resultados mais peculiares! Petersen na Universidade de Utah (Petersen², 1975, 1976a, 1976b *apud* Moorer, 1977) tem feito experiências usando pedaços inteiros de música como fontes de excitação. Isto produz o efeito de orquestra falante, onde a característica da fala é também aparente. A inteligibilidade pode ser aumentada pela escolha de fontes de banda muito larga, tais como uma orquestra inteira e corais (em oposição a, por exemplo, um quarteto de cordas).

O objetivo deste tipo de modificação é dar ao compositor um grande sortimento de novos instrumentos, todos derivados de sons naturais mas diferentes destes. O método subtrativo fornece uma excelente oportunidade para tais combinações nas quais o processo de geração de sons já é dividido em duas partes: a modelagem da excitação e a modelagem do espectro. Este método fornece também um controle fácil de tempo e tom, no qual uma vez que a análise tenha sido feita, a síntese pode ser feita em qualquer taxa, e a frequência fundamental estimada pode ser modificada de qualquer maneira desejada.

A vantagem desta técnica sobre a de síntese aditiva é que ela essencialmente separa os efeitos do tom, tempo e espectro. Na síntese aditiva o tom e o espectro são intimamente ligados de modo que a função amplitude para uma harmônica particular é aplicada a uma frequência diferente quando a harmônica muda de frequência. Então mudando o tom de uma escala de instrumento de síntese aditiva, a forma do espectro também muda, ao passo que no modelo subtrativo, o tom e o espectro são independentes.

Uma outra vantagem do método de predição linear é que este não é muito sensível à função de excitação. Isto significa que é possível calcular um filtro que combine, no sentido de um mínimo erro quadrático médio, o espectro de uma forma de onda inarmônica tão bem quanto para uma forma de onda perfeitamente harmônica. Isto

²PETERSEN, T. L. (1975) Vocal tract modulation of instrumental sounds by digital filtering, apresentado na Music Computation Conf. II, School of Music, Univ. Illinois, Urbana-Champaign, Nov. 7-9.

PETERSEN, T. L. (1976a) Dynamic sound processing. in Proc. 1976 ACM Computer Science Conf. (Anaheim, CA.), Feb. 10-12.

PETERSEN, T. L. (1976b) Analysis-synthesis as a tool for creating new families of of sound. apresentado na 54th Conference of Audio Engineering Society (Los Angeles, CA.) May 4-7.

apud MOORER, James A. (1977) Signal processing aspects of computer music - a survey. *Proceedings of the IEEE*, vol 8 n 8, August, p. 1120.

não é verdade para a análise da síntese aditiva, a qual assume uma forma de onda perfeitamente periódica. O preditor linear não é sensível ao fato de que o espectro que está sendo modelado é de natureza harmônica ou inarmônica, exceto quando a ordem do filtro é muito alta; então cada harmônica é rastreada separadamente no caso periódico. Deste modo, seria interessante tentar simular sons de tambores desta maneira. Muitos tambores têm espectros totalmente inarmônicos os quais poderiam ser bem simulados com esta técnica.

Nesta linha de pensamento, Laroche & Meillier (1994) conseguiram sintetizar sons percussivos através de síntese subtrativa. Sons percussivos podem ser entendidos como aqueles resultantes das vibrações livres de um corpo ou meio postos em movimento por uma excitação curta. Exemplos de sons percussivos são os de tambores, piano, sinos, marimba, violão, etc. Estes sons são classicamente decompostos em duas partes temporais adjacentes: o ataque, de curta duração (menor que algumas centenas de ms) e a ressonância. O ataque inclui as interações entre a excitação e o corpo ressonante e dura até que a estrutura alcance uma vibração estável. A ressonância vem após a porção de ataque e inclui as vibrações livres do corpo do instrumento.

Em teoria, sob a hipótese de pequenas perturbações, pode-se mostrar que a ressonância da estrutura gera um som composto por senóides exponencialmente amortecidas. A idéia é então modelar este tipo de som como a saída de um filtro ressonante excitado por um sinal de pequena duração, isto é, fazer uma analogia da forma

EXCITAÇÃO FÍSICA $\leftrightarrow$ SINAL DE EXCITAÇÃO
CORPO RESSONANTE $\leftrightarrow$ FILTRO RESSONANTE.

São apontadas duas razões principais para este tipo de analogia:

1. *Conteúdo senoidal*: um modo simples de gerar componentes senoidais exponencialmente amortecidas consiste em alimentar um filtro polos/zeros estável com um sinal de excitação curto. Depois que a excitação para, a saída

do filtro de resposta a impulso infinita é uma soma de senóides exponencialmente amortecidas.

2. *Independência entre a excitação e os valores de frequências e fatores de amortecimento:* as frequências e fatores de amortecimento das senóides observadas na parte ressonante de sons percussivos não dependem da natureza da excitação. São determinados pelas propriedades do corpo ressonante, suas dimensões, rigidez, etc. A excitação modifica apenas as amplitudes e fases relativas das componentes senoidais. Esta propriedade é válida também para os modelos sinal de excitação/filtro. Uma vez que a excitação para, as frequências e fatores de amortecimento das senóides que compõem a resposta do filtro *não dependem do sinal de excitação*, mas somente dos polos do filtro.

Deste modo, parece razoável tentar modelar sons percussivos como a saída de um filtro IIR excitado por um sinal de curta duração. O filtro ressonante gera a parte senoidal amortecida do sinal, enquanto que a excitação reproduz o seu ataque.

De um ponto de vista físico, este modelo pode ser interpretado da seguinte maneira: a estrutura vibrante (por exemplo, as cordas e o corpo do violão, as teclas de um vibrafone) é representada pelo filtro ressonante, e a excitação física (por exemplo, o dedo do violonista, a baqueta do vibrafone) é representada pelo sinal de excitação. Assim como o corpo vibrante do instrumento, o filtro ressonante controla as frequências e fatores de amortecimento das senóides na parte quasi-estacionária. Da mesma maneira que a excitação física, o sinal de excitação de curta duração controla as amplitudes e fases e é relevante para o conteúdo não senoidal do ataque do sinal.

É interessante notar que do modo como foi desenvolvido o modelo, pode-se criar vários efeitos apenas modificando a função de excitação em uma tentativa de modelar os vários estilos com que é tocado um mesmo instrumento. Sabe-se que quando um instrumento percussivo é tocado com mais força ou mais rápido há um maior conteúdo de altas frequências no ataque. Modificando o sinal de excitação pode-se obter esta característica facilmente. Portanto esta técnica permite simular vários estilos de execução,

e pode ser utilizada juntamente com um teclado sensível à pressão exercida pelos dedos, por exemplo.

3.5.2. Síntese aditiva baseada em análise.

No modelo de síntese aditiva, um sinal $X(n)$ pode ser aproximado por

$$X(n) \approx \sum_{i=0}^M y_i(n) \tilde{\psi}_i(\theta_i(n)) \quad (3.5.2.1)$$

onde $\tilde{\psi}_i$ é uma função base periódica, y_i e θ_i são as funções de amplitude e fase, respectivamente, associadas a $\tilde{\psi}_i$.

No caso de serem usadas senóides como funções base, tem-se a chamada síntese de Fourier. Usando este modelo, pode-se representar o sinal da seguinte maneira:

$$X(n) = \sum_{k=1}^M A_k(n) \sin\{nT[k\omega + 2\pi F_k(n)]\} \quad (3.5.2.2)$$

onde $X(n)$ é o sinal no instante nT , n é o número da amostra (índice de tempo), T é o tempo entre amostras consecutivas, ω é a frequência fundamental da nota em radianos, k é o número da harmônica, $A_k(n)$ é a amplitude da harmônica k no instante nT (assume-se que varia vagarosamente no tempo), M é o número de harmônicas, e $F_k(n)$ é o desvio de frequência da harmônica k no instante nT . (assume-se que varia vagarosamente no tempo).

Deste modo, a forma de onda é representada por uma soma de senóides harmonicamente relacionadas com variações lentas de amplitude e frequência. (É importante salientar que estas funções senoidais podem ser substituídas por outras funções, gerando assim novas técnicas de síntese.)

O propósito da técnica de análise é estimar estas amplitudes e frequências independentemente para cada harmônica. A equação (3.5.2.2) é o modelo tanto da forma de onda como da técnica de síntese para a classe de sinais harmônicos.

Se a equação (3.5.2.2) é vista como um modelo de uma forma de onda, então $X(n)$ é uma função de dados amostrados que pode ser decomposta em M harmônicas. Olhando-a como uma técnica de síntese, representa a n -ésima amostra de saída da forma de onda sendo sintetizada, e aquela forma de onda pode ser calculada como uma soma de M senóides. Especificando valores para os parâmetros envolvidos (ω , M , T , $A_k(n)$ e $F_k(n)$), especifica-se a forma de onda inteira para todos os instantes.

Estes parâmetros podem ser criados pelo compositor, ou obtidos através da análise dos sons musicais existentes. Para o caso da síntese de Fourier, Grey & Moorer (1977) propuseram um método denominado análise heteródina, que consiste em vários filtros sintonizados nas harmônicas do sinal sob análise. Na saída destes seriam recolhidas as informações das variações de amplitude de cada harmônica, como mostrado na figura 6.

Figura 6: Sistema de análise heteródino.

Estes dados podem ser utilizados diretamente para sintetizar tons pelo uso de (3.5.2.2). O grande problema desta técnica é o fato de que estas funções de amplitude e frequência representam uma quantidade enorme de dados. Estima-se que para uma reprodução de alta fidelidade, sejam necessárias de 20 a 80 harmônicas (Haken, 1992). É então interessante encontrar meios de reduzir a quantidade de dados envolvidos.

Existem muitos meios de se reduzir os dados. Por exemplo pode-se usar a codificação de Huffman ou a modulação delta ou vários outros meios (Moorer, 1977), mas estes violariam um dos critérios adotados para os dados de análise intermediária, que é o de que estes precisam ter uma forma intuitiva atraente que possa ser facilmente modificada pelo compositor. Esta necessidade resulta do fato de o objetivo com as técnicas de síntese não é apenas criar somente réplicas de sons tradicionais. Os dados de análise precisam ser colocados em uma forma onde o efeito de mudanças no timbre resultante seja o mais aparente possível, ou seja, mudanças pequenas nos dados deveriam produzir pequenas mudanças no timbre resultante, enquanto que mudanças mais radicais deveriam fazer-se sentir como uma mudança mais drástica no som gerado.

Os fatores principais que influem na carga computacional para um modelo que emprega a técnica de síntese aditiva são:

- geração de valores para as funções de amplitude;
- multiplicação dos valores destas amplitudes pelos das funções base.

A diminuição de funções base para a síntese alivia a carga computacional nos dois sentidos. Deste modo, as técnicas estatísticas de Karhunen-Loève parecem ser bastante úteis no modelo de síntese aditiva. A razão disso é que pela forma com que foi desenvolvida, a transformada de KL é a transformada ótima no sentido de um mínimo erro quadrático médio de truncamento. Isto quer dizer que é possível realizar a síntese com menos funções base, mantendo ainda a qualidade dos sons gerados.

Uma outra forma de reduzir a quantidade de dados envolvidos na síntese, e que tem sido bem sucedida é a de adequar cada função de amplitude e frequência com funções lineares por partes (Grey & Moorer, 1977). Na figura 7 tem-se uma representação de um tom de clarineta, mostrando as variações temporais de cada harmônica separadamente. Os eixos X, Y e Z representam o tempo, amplitude e frequência, respectivamente. O tempo corre da esquerda para a direita no eixo X, e no eixo Z, a harmônica de mais baixa frequência

(fundamental) está ao fundo. A escala do eixo Y é linear. Na figura 8 tem-se a mesma representação, onde foi aplicada a técnica de linearização por partes.

Figura 7: Tom de clarineta digitalizado e separado em harmônicas.

Figura 8: Tom de clarineta linearizado por partes.

Experimentos psico-acústicos realizados por Grey & Moorer (1977) mostram que os tons sintéticos, mesmo aqueles representados por funções lineares por partes, foram considerados como sendo bem próximos dos originais.

Ainda com a preocupação de diminuir a quantidade de dados envolvida na síntese aditiva de Fourier, Haken (1992) apresentou um artifício interessante para o caso de várias notas sendo sintetizadas ao mesmo tempo (acordes, por exemplo): quando são tocadas várias notas simultaneamente, existe um efeito denominado mascaramento auditivo, onde as parciais com amplitudes maiores mascaram aquelas com amplitudes menores. Deste modo não é necessário reproduzir todas as parciais das notas que estão sendo tocadas. Assim, consegue-se uma redução razoável na carga computacional, sem alterar a qualidade do som sintetizado. A determinação de quais parciais a serem eliminadas foi feita da seguinte maneira: dividiu-se espectro de frequências em 7 regiões (0 a 125 Hz, 125 a 250 Hz, 250 a 500 Hz, 500 a 1000 Hz, 1 a 2kHz, 2 a 4 kHz e 4 a 10kHz). Dentro de cada região, são preservadas as parciais com as maiores amplitudes, de modo que as parciais com baixas amplitudes são eliminadas apenas se existirem parciais com amplitudes maiores na mesma região. Deste modo, garante-

se que todas as regiões estariam presentes na síntese final. Pode-se ver então que esta técnica é uma forma de garantir que a carga computacional seja independente (ou quase) do número de notas sendo sintetizadas no mesmo instante.

Para o trabalho presente, como o objetivo é apenas sintetizar sons musicais isolados, esta última técnica de redução de dados não foi considerada. Contudo, tanto as técnicas estatísticas de Karhunen-Loève quanto a linearização por partes das envoltórias de amplitude e frequência foram utilizadas neste trabalho.

3.6. Novas técnicas de síntese.

Todas as técnicas descritas até agora apresentam algum tipo de problema. As técnicas baseadas em modelos matemáticos simples como a de modulação em frequência provaram que apesar da aparente complexidade do espectro gerado, a previsibilidade matemática dos resultados provoca uma certa familiaridade e um certo cansaço dos timbres produzidos. Com a síntese subtrativa acontece o mesmo problema, com o agravante de produzirem sons muito menos naturais. Mesmo a técnica aditiva tem o limitante na geração de sons inarmônicos. De modo a incrementar o realismo e a naturalidade dos sons gerados, foram propostas várias alternativas, algumas das quais serão descritas a seguir.

3.6.1. PARSHL.

Em 1987, Julius Smith e Xavier Serra³ propuseram uma nova técnica para trabalhar com sons inarmônicos com variações de frequência, utilizando uma tecnologia de localização de regiões de pico emprestada das aplicações navais de processamento de sinais e da técnica de *phase vocoder*. a implementação desta técnica resultou em um programa denominado

³CCRMA, *Research overview*, CCRMA, Stanford, 1992 *apud* RUSCHIONI, Ruggero Andrea (1993) Música digital: histórico e perspectivas. In: WORKSHOP SOBRE COMPUTAÇÃO DE ALTO DESEMPENHO PARA PROCESSAMENTO DE SINAIS. São Carlos, 1993. *Anais*. São Carlos, 1993. pp. 178-191.

PARSHL. Neste, são computadas uma série de janelas, com a diferença que ao invés de proporcionar a saída direta da fase e da magnitude de cada banda, a FFT procura regiões de pico. Os picos maiores são localizados de janela em janela. A principal diferença na análise consiste na substituição das fases derivadas de cada banda de FFT pela interpolação das magnitudes entre os picos das diversas bandas. Isto proporciona uma melhor análise para sons com vibrato.

3.6.2. SMS.

Com o intuito de trabalhar sons com componentes de ruído, a mesma dupla de pesquisadores propôs uma nova técnica denominada *Spectral modeling synthesis* (SMD)⁴. Neste modelo, o som é modelado por uma parte *determinística*, formada por componentes senoidais, de maneira similar à de Fourier, e por outra *estocástica*, descrita pela probabilidade da densidade da amplitude versus frequência. Quando o sinal é assumido como estocástico, não é necessário preservar a fase instantânea nem o detalhe individual exato das janelas de FFT.

3.6.3. Transformada *wavelet*.

A transformada *wavelet* decompõe os sinais em ondas mais simples do que os senos e cossenos da técnica de Fourier. As *wavelets* elementares são funções de dois parâmetros obtidas por deslocamentos no tempo e também por dilatações (ou compressões) que atuam tanto no domínio do tempo como no da frequência. Pelo fato de estar localizada tanto no domínio do tempo como no da frequência, é possível a análise de sons naturais com grande precisão.

⁴ SERRA, Xavier & SMITH, Julius O. Spectral modeling synthesis. *Computer Music Journal*, v. 14, . 4. MIT Press. 1990. *apud* RUSCHIONI, Ruggero Andrea (1993) Música digital: histórico e perspectivas. In: WORKSHOP SOBRE COMPUTAÇÃO DE ALTO DESEMPENHO PARA PROCESSAMENTO DE SINAIS. São Carlos, 1993. *Anais*. São Carlos. 1993. pp. 178-191.

As *wavelets* também permitem a modificação dos sinais de inúmeros modos. Alterando os coeficientes da transformada, pode-se alterar a velocidade do som sem afetar a frequência. Outra coisa interessante que pode ser feita é cruzar a informação de dois sinais e sintetizar um terceiro, utilizando a informação de módulo de um e a informação de fase de outro.

3.6.4. Síntese pelo uso de autômatos lineares (LASy).

De um modo geral, os autômatos lineares podem ser vistos como uma matriz de células em um determinado estado a um determinado momento. Ao toque de um relógio imaginário, cada célula computa o seu novo estado de acordo com o seu estado anterior e com o das células vizinhas. As regras de transição que determinam o estado inicial e o novo estado de cada célula da matriz são os elementos necessários para calcular a evolução do autômato geração após geração. Pelo fato de os autômatos lineares serem um processo temporal, sua transposição musical parece adequada. Jacques Chareyron⁵ propôs um algoritmo baseada nesta técnica, onde encara uma forma de onda gravada na memória como um autômato linear, onde cada elemento, isto é cada amostragem, é vista como uma célula de um autômato. O valor de um elemento da forma de onda é o estado que corresponde à célula do autômato. Geração após geração, o estado de cada célula é computado de acordo com a regra de transição, obtendo assim um sinal auto-modificante. A vantagem desta técnica é a sua simplicidade, propiciando uma síntese em tempo real sem o uso de hardware específico. Pelo fato de o resultado sonoro ser bastante inusitado, o som gerado apresenta texturas harmônicas bastante ricas e em constante evolução. A grande desvantagem é a total imprevisibilidade dos autômatos. A criação de regras de transição que resultem em sons musicalmente úteis, à semelhança das técnicas de FM, depende de grande esforço de experimentação.

⁵ CHAREYRON, Jacques (1990) Digital synthesis of self modifying waveforms by means of linear automata. *Computer Music Journal*, v 14, n. 4. MIT Press *apud* RUSCHIONI, Ruggero Andrea (1993) Música digital: histórico e perspectivas. In: WORKSHOP SOBRE COMPUTAÇÃO DE ALTO DESEMPENHO PARA PROCESSAMENTO DE SINAIS. São Carlos, 1993. *Anais*. São Carlos, 1993. pp. 178-191.

3.6.5. FOF.

O modelo FOF (*Fonction d'onde formantique*) usa um esquema que cria um espectro contendo ressonâncias sem o uso de filtros. Este simula a fala disparando osciladores múltiplos em série harmônica por pequenos e curtos picos no começo de cada período da fundamental. O usuário especifica a frequência central, a largura de banda, a frequência e a amplitude da fundamental, e outras variáveis específicas do sistema.

Na técnica de predição linear um som vocal seria obtido passando uma função de excitação por um filtro programado para simular a resposta do trato vocal. Deste modo, a forma de onda seria descrita como a resposta a impulso do filtro. Este filtro pode ser aproximado por um conjunto de filtros de segunda ordem em paralelo, um para cada formante. A resposta a impulso do filtro total seria a soma das respostas a impulso de cada filtro componente. No modelo FOF, substitui-se cada filtro por um oscilador digital.

Esta técnica oferece ao músico uma simplicidade conceitual que não existia no método de predição linear, onde era necessário construir os algoritmos dos filtros digitais. Outra vantagem é que os parâmetros a serem especificados tem uma relação direta com os atributos acústicos do som.

3.6.6. VOSIM.

A técnica VOSIM (VOice SIMulation) permite a produção de timbres ricos através do controle direto da estrutura dos formantes do som. A forma de onda VOSIM básica consiste de uma série de pulsos seguidos por um curto intervalo de tempo, no qual o sinal tem o valor 0.

3.7. Conclusões.

A técnica de síntese aditiva com funções base senoidais (síntese aditiva de Fourier) é tradicionalmente considerada como a de mais alta qualidade. Entretanto, a carga computacional imposta por esta técnica é muito alta, o que faz com que a sua implementação em tempo real tenha um custo bastante elevado.

Substituído o conjunto de funções base de Fourier por outro menor baseado nas técnicas estatística de Karhunen-Loève, consegue-se diminuir a carga computacional, diminuindo também os custos para operação em tempo real.

4. A TRANSFORMADA DE KARHUNEN-LOÈVE.

4.1. Introdução.

A transformada de Karhunen-Loève é uma transformada ótima em relação ao erro quadrático médio de truncamento (Ahmed & Rao, 1975). Por isto, sua forma é bastante útil para a compressão de dados. Para verificar tal propriedade, é interessante observar como ela é obtida. Nas seções seguintes será mostrada uma maneira de derivar esta transformada e suas propriedades.

4.2. Derivação da transformada.

Se um sinal discreto é composto de N valores amostrados, então ele pode ser visto como sendo um ponto em um espaço N -dimensional. Cada valor amostrado é então uma componente do vetor de dados X , o qual representa o sinal neste espaço. Para uma representação mais eficiente, obtém-se uma transformação ortogonal de X a qual resulta em $Y = TX$, onde Y e T denotam o vetor de transformação e a matriz de transformação, respectivamente. O objetivo é selecionar um subconjunto de M componentes de Y , onde M é substancialmente menor que N para representar o sinal com o menor erro possível. As $(N - M)$ componentes restantes podem então ser descartadas sem introduzir erro representativo, quando o sinal é reconstruído usando as M componentes restantes de Y . As transformações ortogonais precisam portanto ser comparadas em relação a algum critério de erro. Um critério normalmente usado é o do erro quadrático médio.

Seja $\mathbf{T}$ uma transformação ortogonal dada por

$$\mathbf{T}' = [\phi_1 \phi_2 \dots \phi_N] \quad (4.2.1)$$

onde $\mathbf{T}'$ denota a transposta de $\mathbf{T}$, e os ϕ_i são vetores coluna de N elementos. Por conveniência, assume-se que os vetores base $\{\phi_m\}$ são reais e ortonormais, isto é,

$$\phi_i' \phi_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \quad (4.2.2)$$

Para cada vetor X pertencente a um dado conjunto de vetores de dados, obtém-se

$$Y = TX \quad (4.2.3)$$

onde $X' = [x_1 \ x_2 \ \dots \ x_n]$ e $Y' = [y_1 \ y_2 \ \dots \ y_N]$.

Das equações (4.2.1) e (4.2.2) segue que $\mathbf{T}'\mathbf{T} = \mathbf{I}$ e portanto

$$X = \mathbf{T}' Y = [\phi_1 \phi_2 \dots \phi_N] Y$$

a qual leva a

$$X = y_1 \phi_1 + y_2 \phi_2 + \dots + y_N \phi_N = \sum_{i=1}^N y_i \phi_i \quad (4.2.4)$$

O objetivo é conservar um subconjunto $\{y_1, y_2, \dots, y_M\}$ das componentes de Y e ainda estimar X . Isto pode ser feito substituindo as $(N-M)$ componentes restantes de Y pré-selecionando constantes b_i para obter

$$\hat{X}(M) = \sum_{i=1}^M y_i \phi_i + \sum_{i=M+1}^N b_i \phi_i \quad (4.2.5)$$

onde $\hat{X}(M)$ denota a estimativa de X .

O erro introduzido por desprezar os $(N-M)$ termos pode ser representado como

$$\Delta X = X - \hat{X}(M)$$

onde ΔX é o vetor de erro, isto é

$$\Delta X = X - \sum_{i=1}^M y_i \phi_i - \sum_{i=M+1}^N b_i \phi_i \quad (4.2.6)$$

Das equações (4.2.4) e (4.2.6) segue que

$$\Delta X = \sum_{i=M+1}^N (y_i - b_i) \phi_i \quad (4.2.7)$$

Então o erro quadrático médio $\varepsilon(M)$ é dado por

$$\varepsilon(M) = E\{\|\Delta X\|^2\} = E\{(\Delta X)' (\Delta X)\} \quad (4.2.8)$$

Substituindo (4.2.7) em (4.2.8),

$$\varepsilon(M) = E\left\{ \sum_{i=M+1}^N \sum_{j=M+1}^N (y_i - b_i)(y_j - b_j) \phi_i' \phi_j \right\}$$

a qual, pela eq. (4.2.2) se simplifica para gerar

$$\varepsilon(M) = \sum_{i=M+1}^N E\{(y_i - b_i)^2\} \quad (4.2.9)$$

Da equação (4.2.9) segue que para cada escolha de ϕ_i e b_i , obtém-se um valor para $\varepsilon(M)$. O objetivo é então fazer a escolha que minimize $\varepsilon(M)$. O processo para escolher os b_i e os ϕ_i ótimos é realizado em duas etapas.

Passo 1. Os b_i ótimos são obtidos como segue:

$$\frac{\partial}{\partial b_i} E\{(y_i - b_i)^2\} = 2[E\{y_i\} - b_i] = 0$$

o qual gera

$$b_i = E\{y_i\} \quad (4.2.10)$$

Agora, das equações (4.2.2) e (4.2.4) obtém-se

$$y_i = \phi_i' X \quad (4.2.11)$$

Então tem-se

$$b_i = \phi_i' E\{X\} = \phi_i' \bar{X}, \text{ onde } \bar{X} = E\{X\}.$$

Novamente, desde que a quantidade $(y_i - b_i)$ na equação (4.2.9) é um escalar, pode-se escrever $\varepsilon(M)$ como

$$\varepsilon(M) = \sum_{i=M+1}^N E \left\{ (y_i - b_i)(y_i - b_i)' \right\} \quad (4.2.12)$$

Substituindo $y_i = \phi_i' X$ e $b_i = \phi_i' \bar{X}$ na equação (4.2.12) resulta em

$$\varepsilon(M) = \sum_{i=M+1}^N \phi_i' E \left\{ (X - \bar{X})(X - \bar{X})' \right\} \phi_i$$

Considerando $\sum_x = E \left\{ (X - \bar{X})(X - \bar{X})' \right\}$ como a matriz de covariância de X , tem-se

$$\varepsilon(M) = \sum_{i=M+1}^N \phi_i' \sum_x \phi_i \quad (4.2.13)$$

Passo 2. Para obter os ϕ_i ótimos é necessário não somente minimizar $\varepsilon(M)$ em relação a ϕ_i , mas também satisfazer à condição $\phi_i' \phi_i = 1$. Usando o método de multiplicadores de Lagrange (ver Apêndice 1) minimiza-se então a seguinte expressão:

$$J(M) = \varepsilon(M) - \sum_{i=M+1}^N \beta_i [\phi_i' \phi_i - 1] = \sum_{i=M+1}^N \left\{ \phi_i' \sum_x \phi_i - \beta_i [\phi_i' \phi_i - 1] \right\} \quad (4.2.14)$$

em relação a ϕ_i , onde β_i denota o multiplicador de Lagrange.

Agora, pode ser mostrado que

$$\nabla \phi_i \left[\phi_i' \sum_x \phi_i \right] = 2 \sum_x \phi_i$$

e

$$\nabla \phi_i [\phi_i' \phi_i] = 2\phi_i$$

Então da equação (4.2.14) tem-se

$$\nabla \phi_i [J(M)] = 2 \sum_x \phi_i - 2\beta_i \phi_i = 0$$

o qual gera

$$\sum_x \phi_i = \beta_i \phi_i \quad (4.2.15)$$

Por definição, a equação (4.2.15) implica que ϕ_i são os *autovetores* da matriz de covariância Σ_x e β_i são os i -ésimos *autovalores* correspondentes. É conveniente ordenar os autovetores de modo que os primeiros sejam aqueles cujos autovalores correspondentes sejam maiores. Deste modo, ao truncar a representação de um sinal, estarão sendo desprezados os autovetores menos representativos, ou seja, com os menores autovalores.

Denotando β_i por λ_i , e substituindo a equação (4.2.15) na equação (4.2.13), o *mínimo* erro quadrático médio é dado por

$$\varepsilon_{\min}(M) = \sum_{i=M+1}^N \lambda_i \quad (4.2.16)$$

A expansão definida na equação (4.2.4) é então em termos dos autovetores da matriz de covariância, e é chamada de *expansão de Karhunen-Loève*.

4.3. Propriedades da transformada de Karhunen-Loève.

Na literatura estatística, o problema de minimizar $\varepsilon(M)$ é chamado de *análise de componente principal*. Abaixo tem-se duas conclusões importantes pertinentes à análise acima:

- i. A transformada de Karhunen-Loève é a transformada *ótima* para representação de sinais em relação ao critério do erro quadrático médio.
- ii. Desde que $Y = TX$, a matriz de covariância no domínio da transformada Σ_y é dada por

$$\Sigma_y = T \Sigma_x T^{-1} = T \Sigma_x T'$$

Entretanto, desde que T é composta pelos autovetores da matriz de covariância de Σ_x , segue que

$$\Sigma_y = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) \quad (4.3.1)$$

onde λ_i , $i = 1, 2, \dots, N$ são os autovalores de Σ_x .

Desde que Σ_y é uma *matriz diagonal*, pode-se concluir que as componentes do vetor de transformação y_i na equação (4.2.11) são *descorrelacionados*.

4.4. Conclusões.

No capítulo anterior foi visto que o método que tem alcançado a melhor qualidade de síntese é a técnica aditiva com funções base de Fourier. Entretanto são necessárias de 20 a 80 senóides (Haken (1992)) para uma síntese de alta qualidade. A

carga computacional imposta por tal técnica tem sido o maior limitante para a utilização desta técnica para aplicações em tempo real.

A transformada de Karhunen-Loève pode ser um instrumento útil para a síntese de sons musicais, devido à sua característica de ser a transformada ótima em relação ao erro quadrático médio de truncamento. Com isto, seria possível sintetizar sons musicais com poucas funções base, mantendo ainda uma boa qualidade. Assim, a carga computacional, é diminuída, fazendo com que o hardware necessário para a sua implementação em tempo real seja minimizado, baixando o seu custo.

5. SISTEMA EXPERIMENTAL.

5.1. Introdução.

O modelo escolhido para este trabalho foi o de síntese aditiva baseada em análise com funções base de Karhunen-Loève. Isto significa que os parâmetros que irão dirigir a síntese são obtidos através da análise de sons musicais já existentes. Para este fim, foi montado um sistema para analisar e sintetizar estes sons. Este é constituído de diversos programas para microcomputadores do tipo PC, desenvolvidos em linguagem Pascal, com utilização da filosofia de programação estruturada.

Antes de serem analisados, os sinais devem passar por uma série de pré-processamentos para que assumam uma forma mais adequada para a análise. A finalidade disso é aumentar a concentração de energia nas primeiras funções base, fazendo com que seja necessário um número menor delas para a reconstrução dos sinais. Os pré-processamentos incluem normalização em frequência, separação dos sinais em classes, e alinhamento em fase. Com a normalização em frequência, surge a necessidade de medir a frequência fundamental dos sinais sob análise, e então foi desenvolvido um sistema para realizar esta tarefa.

Após os pré-processamentos, o próximo passo é calcular as funções base para a transformada de Karhunen-Loève, que são dadas pelos autovetores da matriz de covariância dos sinais. Para gerar a matriz de covariância, tomou-se um segmento de 103 pontos de cada sinal no início da parte de sustentação.

Para completar o sistema de análise, é necessário agora inserir as variações temporais de amplitude e fase das funções base. Isto é feito com um sistema de detecção de envoltória.

Este consiste em determinar, em pontos espaçados no tempo, os valores das amplitudes e fases que melhor aproximem o sinal sintetizado do original. Nos pontos intermediários, as amplitudes são obtidas através de linearização por partes e as fases por uma aproximação quadrática por partes.

Com estes dados é possível reconstruir o sinal original, aplicando-se a transformada inversa de Karhunen-Loève, utilizando os coeficientes e fases obtidos pelo sistema de detecção de envoltória, e os autovalores calculados anteriormente. O último passo é alterar novamente a frequência fundamental do sinal, fazendo com que volte ao seu valor original.

O diagrama em blocos abaixo mostra as etapas descritas acima.:

Figura 9: Diagrama em blocos do sistema de análise/síntese.

A seguir será detalhado e explicado o funcionamento e o porquê de cada bloco deste diagrama.

5.2. Aquisição de dados.

A aquisição dos sons dos instrumentos musicais foi feita utilizando-se um microfone conectado a um conversor A/D de uma placa de áudio. O equipamento usado para a coleta de sinais foram um microfone dinâmico SHURE, modelo SM-58, um equalizador gráfico STANER de 2/3 de oitava para limitá-los em 15kHz, e uma placa de áudio

SOUNDBLASTER 16 da Creative Labs. Os sinais foram digitalizados com uma faixa dinâmica de 8 bits e frequência de amostragem de 44.1 kHz.

Os sinais capturados são armazenados no formato WAV, o qual é constituído de duas partes: um cabeçalho contendo informações sobre o sinal, e o sinal propriamente dito. As informações contidas no cabeçalho incluem a taxa de amostragem (11.025, 22.05, ou 44.1 kHz), faixa dinâmica (8 ou 16 bits), número de canais (mono ou estéreo) e também o tamanho do sinal em bytes. A seguir vem o sinal, o qual é armazenado de forma linear (sem compressão), como mostrado na figura abaixo:

Frequência de amostragem, nº de canais e faixa dinâmica	Tamanho do sinal	Sinal no formato inteiro de 8 bits
40 bytes	4 bytes	N bytes

Figura 10: Estrutura do formato de áudio WAV.

Os sinais capturados pela placa A/D são armazenados em 8 bits. Neste formato, as amplitudes apresentam valores inteiros na faixa de 0 a 255. Estes dados irão passar por vários processamentos intermediários, tais como correlação, normalização em frequência, alinhamento em fase, entre outros. Estes incluem operações de divisão, radiciação, exponenciação e outras mais, impossíveis de serem realizadas com números inteiros. Uma das tarefas executadas por este programa é fazer a conversão para o formato real.

Outro problema detectado é que a faixa de amplitudes (0 a 255) gerou problemas de estouro da capacidade de processamento nos cálculos intermediários. Isto foi resolvido através da normalização da amplitude dos sinais entre -1 e +1.

Também foi incluída uma opção para conversão dos sinais em formato ASCII, de modo a permitir a exportação destes dados para outros programas tais como planilhas de cálculo, editores de texto e programas gráficos.

5.3. Visualização gráfica dos sinais.

A identificação das regiões de ataque, sustentação e decaimento é facilitada com a visualização da forma de onda do sinal. Também as características peculiares de cada tom podem ser melhor entendidas quando visualizadas graficamente na tela do computador. Para este fim foi desenvolvido um programa para exibir a forma de onda do sinal na tela do computador. Este mostra os sinais em duas escalas diferentes: uma exibe o sinal inteiro na tela, de modo que se possa observar a envoltória do sinal, e a outra mostra-o em detalhes, onde podem ser observadas partes das formas de onda dos tons coletados.

5.4. Determinação da frequência fundamental do sinal.

Os sinais capturados devem ser normalizados em frequência. Antes porém de pensar em normalizar a frequência fundamental das funções amostra é necessário conhecê-la, ou seja, medi-la. Uma maneira de medir o período fundamental de um sinal é tomar dois picos consecutivos e medir a distância entre eles. Sabendo-se a frequência de amostragem do sinal, o número de amostras dentro deste intervalo fornece o seu período fundamental.

Para melhorar a precisão da medida, Schafer & Rabiner (1973) propuseram a aplicação da autocorrelação ao sinal. Esta função apresenta duas propriedades úteis à detecção do período fundamental: apresenta o mesmo período do sinal, e tem uma forma de onda mais aguda. Assim a detecção dos picos é facilitada.

Existe então o problema de determinar quais seriam estes picos consecutivos. Uma das formas de se fazer isto é detetar, na região de sustentação do tom, por exemplo, o ponto de máximo valor absoluto do sinal. Este é com certeza um pico. O pico seguinte seria o próximo ponto de máximo valor absoluto, assumindo-se que o sinal decai em amplitude com o tempo. A idéia parece funcionar bem, mas na prática foram detetados alguns problemas.

Devido a flutuações de amplitude do sinal, nem sempre os picos tomados desta maneira representam um período. Isto é ilustrado na figura 11.

De acordo com esta figura, pode-se notar que o período fundamental é dado pelo intervalo de tempo compreendido entre os pontos A e B ou B e C. Porém, devido a flutuações na amplitude do sinal, um sistema de detecção de picos assumiria o período do sinal como sendo o intervalo AC. A forma encontrada para contornar este problema foi multiplicar o sinal por uma exponencial negativa de modo a forçar um comportamento monotonicamente decrescente. Isto é mostrado na figura 12.

Figura 11. Exemplo de erro na detecção do período fundamental devido a variações na amplitude do sinal.

Figura 12: Sinal original multiplicado por $y(n) = \exp(-Sn)$, onde $S = 0.004$.

Figura 13: Ilustração de problemas de detecção de pitch devido a irregularidades na forma de onda.

Da figura 12 pode-se ver que agora o detetor entende o período fundamental como sendo o intervalo entre os pontos A e B, como seria o correto. O único problema ainda não resolvido é que o índice S da exponencial (neste caso 0.004) deve variar conforme o caso, isto é, conforme a forma de onda de cada sinal particular. Nos testes realizados, alguns sinais apresentaram um comportamento monotonicamente decrescente para $S = 0.001$. Já outros só apresentaram este comportamento com valores de S acima de 0.007. Com isto, é difícil fazer um detetor automático de pitch desta maneira. A solução encontrada foi ajustar manualmente o índice S . Poderia-se perguntar porque não se usa um valor elevado para S para evitar este ajuste manual. A razão é que as formas de ondas encontradas na prática assumem os mais variados formatos, e nem sempre um valor muito alto para S determina uma detecção correta. Muitas vezes, para S muito elevado, o sinal fica tão deformado que o detetor não consegue mais discernir quais pontos representam os picos.

Outro problema decorrente da forma de onda do sinal é melhor entendido olhando-se a figura 13. Neste caso, um detetor de picos tomaria como sendo o período fundamental o intervalo de tempo entre os pontos A e D quando, na verdade, uma melhor aproximação seria o intervalo entre os pontos A e C ou B e D. A aplicação de uma exponencial negativa ao sinal mostrou ser ineficiente na prática, pois para corrigir estas distorções, o valor de S necessário deve ser tão grande que os picos já não podem mais ser detetados. Uma alternativa razoável é realizar a medição através dos pontos de inversão do sinal, ou seja, nos pontos onde este passa por zero, como mostrado na figura 14.

Nesta, pode-se ver que fazendo a detecção nos pontos A e B, o problema da irregularidade da forma de onda é eliminado. O programa desenvolvido segue então os seguintes passos:

- i) multiplicação do sinal por uma exponencial negativa;
- ii) detecção do máximo global do sinal;
- iii) detecção do próximo máximo global;
- iv) detecção das passagens por zero mais próximas destes máximos globais.

Poderia-se perguntar porque não fazer a detecção do período fundamental diretamente através do passo iv). A razão é melhor compreendida olhando-se para a figura 15. Se a detecção do período fundamental for feita somente através dos pontos de cruzamento, o detetor entenderá como sendo o período fundamental do sinal o intervalo AB, quando na verdade o correto seria AE, ou BF, ou CG, ou ainda, DH. Localizando-se inicialmente os pontos de máximo do sinal (pontos P1 e P2 na figura), o período fundamental fica sendo determinado univocamente como sendo o intervalo CG, ou DH, dependendo do sentido em que se faz a busca pelo cruzamento.

Figura 14: Alternativa para detecção do período fundamental na passagem por zero.

Figura 15: Problemas na detecção do período fundamental somente através dos pontos de cruzamento.

Apesar de um tanto trabalhoso, o algoritmo proposto funcionou bastante bem para as mais variadas formas de onda encontradas. Poderia ser utilizado um detetor automático como aquele proposto por Schafer & Rabiner (1973), mas pelas características particulares das formas de onda encontradas, a medição do período fundamental por meios gráficos apresentou melhores resultados.

5.5. Normalização em frequência.

Os tons coletados de instrumentos diferentes são tocados geralmente em diferentes notas e oitavas, portanto têm frequências fundamentais diferentes. Além disso, estes tons são em geral não periódicos. O valor $x_{\xi}(n+K)$ pode diferir aprecialmente de $x_{\xi}(n)$ devido a variações naturais na envoltória de amplitude e frequência do tom original através de seu tempo de vida. Torna-se então necessário fazer algum tipo de normalização no período dos sinais. Esta normalização pode ser realizada com uma mudança na taxa de amostragem usando os métodos tradicionais de interpolação e decimação descritos por Schafer e Rabiner (1973). O processo consiste de duas etapas:

- i. incrementa-se a taxa de amostragem de um fator L , interpolando-se $(L-1)$ pontos entre cada um dos pontos da sequência original. A sequência resultante $y_1(n)$ tem um período de amostragem $T_1 = T / L$.
- ii. decrementa-se a taxa de amostragem de um fator M , tomando um em cada M pontos de $y_1(n)$. Obtém-se assim a sequência $y(n)$ desejada, e cujo período de amostragem é dado por $T' = MT_1 = (M/L)T$.

Para aumentar a taxa de amostragem do sinal, foi escolhida a interpolação polinomial de Lagrange. Neste caso tem-se que a forma correspondente para a interpolação com espaçamento equidistante de $(L-1)$ amostras entre cada amostra do sinal original, isto é, para mudar o período de amostragem de T para $T' = T/L$, é:

$$y(n) = \sum_{k=-(Q-2)/2}^{Q/2} A_k^Q (n/L) x(k), \quad Q \text{ par} \quad (5.5.1a)$$

$$y(n) = \sum_{k=-(Q-1)/2}^{(Q-1)/2} A_k^Q (n/L) x(k), \quad Q \text{ ímpar} \quad (5.5.1b)$$

onde: Q é o número de amostras utilizadas para o cálculo;

$x(k)$ são amostras do sinal original;

L é o fator em que se deseja alterar o período de amostragem

A são os coeficientes de Lagrange, dados por

$$A_k^Q(t) = \frac{(-1)^{k+Q/2}}{\left(\left(\frac{Q-2}{2}\right) + k\right)! \left(\frac{Q}{2} - k\right)! (t - k)} \prod_{i=0}^Q \left(t + \frac{Q}{2} - i\right), \quad Q \text{ par} \quad (5.5.2a)$$

$$A_k^Q(t) = \frac{(-1)^{k+(Q-1)/2}}{\left(\left(\frac{Q-1}{2}\right) + k\right)! \left(\frac{Q-1}{2} - k\right)! (t - k)} \prod_{i=0}^{Q-1} \left(t + \left(\frac{Q-1}{2}\right) - i\right), \quad Q \text{ Ímpar} \quad (5.5.2b)$$

Uma propriedade útil para o cálculo dos coeficientes é:

$$A_k^Q(t) = \begin{cases} 0, & t \text{ inteiro, e } t \neq k \\ 1, & t = k \end{cases} \quad (5.5.3)$$

Isto resulta do fato de que o polinômio de interpolação passa através de Q amostras do sinal original, isto é

$$y(kL) = x(k), \text{ para } k \text{ inteiro} \quad (5.5.4)$$

As expressões (5.5.1a) e (5.5.1b) podem ser avaliadas para qualquer valor de n tal que

$$-\left(\frac{Q-2}{2}\right)L < n < \frac{Q}{2}L, \quad Q \text{ par} \quad (5.5.5a)$$

$$-\left(\frac{Q-1}{2}\right)L < n < \left(\frac{Q-1}{2}\right)L, \quad Q \text{ ímpar} \quad (5.5.5b)$$

É fácil ver que o valor de Q influi tanto na qualidade da interpolação quanto no tempo de processamento. Foram realizados testes com vários valores de Q , e foi então escolhido um valor $Q = 10$, que resulta numa boa interpolação, em um período de tempo não muito longo.

Para prevenir o *aliasing*, pode ser necessária uma redução na largura de banda se o número de amostras em um segmento precisar ser reduzido durante a normalização do período.

O comprimento T_d de uma função amostra normalizada determina as dimensões da matriz de covariância. Existe um compromisso entre o tamanho da matriz de covariância e a necessidade de preservar a largura de banda nos dados originais tanto quanto possível. No caso deste trabalho, foi escolhido um valor $T_d = 103$ amostras.

Portanto, para normalizar os sinais com um período fundamental original de T_o amostras para um período $T_d = 103$ amostras, faz-se primeiro uma interpolação de $(T_o - 1)$ pontos entre cada ponto do sinal original, e depois faz-se uma decimação, selecionando um em cada T_d pontos do sinal interpolado.

Pelo fato de os sinais processados serem grandes (da ordem de centenas de Kbytes), este procedimento, além de demorado, gera sinais enormes no processo de interpolação, se o valor de T_o é grande. Para contornar este problema, realizou-se a interpolação juntamente com o processo de decimação, calculando-se apenas o valor do sinal interpolado que seria selecionado no processo de decimação.

5.6. Classificação dos sinais.

A concentração de energia nas primeiras funções base depende fortemente do grau de similaridade entre as funções amostra: quanto maior a similaridade, maior a concentração de energia. O agrupamento dos sinais em classes de acordo com a sua

similaridade faz com que se possa reconstruí-los com menos funções base, desde que o erro de truncamento torna-se menor com a maior concentração de energia.

Uma maneira de se medir a similaridade entre dois sinais é realizar a correlação cruzada entre eles:

$$\gamma_{xy}(\tau) = \sum_{n=1}^N x(n)y(n + \tau) \quad (5.6.1)$$

onde: x e y são os sinais analisados;

γ_{xy} é a correlação cruzada;

τ é o deslocamento relativo entre os sinais.

N é o tamanho da janela em que são analisados os sinais.

Como γ_{xy} depende de τ , foi feita uma varredura em τ , e foi escolhido o valor máximo assumido por γ_{xy} como a medida de similaridade entre os sinais.

Com o valor da correlação entre os sinais em mãos, pode-se realizar a separação dos sinais em classes de similaridade. Escolhendo-se um limiar, todos os sinais com correlação igual ou acima deste seriam agrupados em uma mesma classe. Os limiares escolhidos são mostrados na seção 6.2.1.

5.7. Alinhamento em fase.

Um dos problemas com a transformada de KL é que esta é fortemente dependente da fase relativa entre os sinais que constituem a matriz de covariância. Christensen & Hirschman (1979) mostraram que o alinhamento em fase das funções amostra tende a concentrar a maior parte da energia nos primeiros autovetores. O alinhamento força características similares das funções amostra a ocorrerem aproximadamente ao mesmo tempo. Por exemplo, se cada função amostra tem um pico o qual contém uma larga

fração da energia do sinal, então estas funções amostra são alinhadas de modo que estes picos ocorram ao mesmo tempo.

A idéia básica é ajustar as fases das funções amostra de modo a otimizar o alinhamento, no sentido de uma correlação cruzada, com a primeira função base $\phi_1(t)$. A função de correlação entre dois sinais x e y é dada pela expressão

$$\gamma_{xy}(\tau) = \sum_{n=0}^N x(n)y(n+\tau) \quad (5.7.1)$$

onde: γ_{xy} é a correlação cruzada;

τ é o deslocamento relativo entre os sinais;

N é o tamanho da janela em que é feita a análise.

Calcula-se inicialmente $\phi_1^{(1)}(t)$, a função base principal de KL para os dados de entrada iniciais. Um modo fácil de calcular $\phi_1(t)$ é utilizar o chamado método das potências (ver apêndice 2). Cada função amostra $x_i(t)$, é deslocada ao longo do eixo de tempo até que seja alcançada a máxima correlação com $\phi_1^{(1)}(t)$. Os dados deslocados tornam-se a entrada para a próxima iteração. Calcula-se então $\phi_1^{(2)}(t)$, e repete-se o processo. O algoritmo para quando não ocorrer mais nenhuma mudança de fase em nenhuma das funções amostra.

Resta ainda o problema de escolher o tamanho N da janela. Para se ter um retrato fiel do comportamento do sinal, N foi escolhido de modo a conter pelo menos um período do sinal (dado que os sinais foram normalizados para terem um período fundamental de 103 amostras). Com estas considerações, chegou-se a $N = 103$.

Nas figuras 16 e 17 é mostrado um exemplo de como funciona o alinhamento em fase. Na figura 16 tem-se três sinais: uma onda quadrada, uma triangular e uma senóide com fases aleatórias. Na figura 17 são mostrados os mesmos sinais alinhados em fase.

Figura 16: Ondas quadrada, senoidal e triangular com fases aleatórias.

Figura 17: Ondas quadrada, senoidal e triangular alinhadas em fase.

5.8 Cálculo da base de Karhunen-Loève.

Uma vez que os sinais foram pré-processados (normalizados em frequência e alinhados em fase) pode-se calcular a base de KL para este conjunto de funções amostra. De cada sinal foram tomados 103 pontos na região de sustentação do tom, que pelo fato de estarem normalizados em frequência, representam aproximadamente um período do sinal.

O primeiro passo é calcular a matriz de covariância, que é dada pela expressão

$$R_x = \frac{1}{M} \sum_{\xi} x_{\xi} x_{\xi}^t \quad (5.8.1)$$

onde os x_{ξ} são as funções amostra armazenadas em vetores coluna.

A seguir calcula-se os autovalores λ_i e os autovetores ϕ_i de R_x . Uma das maneiras de se fazer isto é inicialmente calcular os autovalores como as raízes do polinômio característico de R_x :

$$\det (R_x - \lambda I) = 0 \quad (5.8.2)$$

Os autovetores podem ser obtidos através de

$$R_x \phi_i = \lambda_i \phi_i \quad (5.8.3)$$

Apesar deste procedimento fornecer os resultados desejados, a carga computacional imposta é extremamente alta, tornando-o muito lento, especialmente dada a ordem de grandeza da matriz de covariância (103 x 103). Uma das maneiras mais eficientes de realizar esta tarefa é utilizar o algoritmo QL em conjunção com o procedimento de Householder para tridiagonalização da matriz de covariância (ver apêndices 3 e 4).

Estes dados serão utilizados pelos estágios seguintes para a detecção da envoltória e posterior reconstrução dos sinais.

5.9 Detecção de envoltória.

Os sons musicais são sinais variantes no tempo. São estas variações que determinam o seu realismo e naturalidade. O sistema de análise recupera um período do sinal na região de sustentação do tom através da expressão

$$x_\xi = \sum_{i=1}^N y_{\xi_i} \tilde{\phi}_i \quad (5.9.1)$$

onde $\tilde{\phi}_i$ é agora uma função base periódica, y_i é a função de amplitude, e N é o número de funções base utilizadas na reconstrução. Os coeficientes y_i são obtidos a partir da expressão

$$y_i = x_{\xi}^T \tilde{\Phi}_i, i = 0, 1, \dots, N-1. \quad (5.9.2)$$

O objetivo é sintetizar tons inteiros e nestes, tanto a frequência como a amplitude variam com o tempo. Moorer (1977) sintetizou com sucesso vários tons com a técnica de síntese aditiva com funções base senoidais (síntese de Fourier), utilizando envoltórias independentes para cada harmônica, e aplicou uma aproximação linearizada por partes em cada uma delas. Stapleton & Bass (1986) aplicaram esta mesma técnica na síntese de sons com funções base de KL, e o algoritmo proposto por estes pesquisadores foi adotado no presente trabalho.

Um som musical variante no tempo pode ser aproximado por:

$$x(n) \approx \sum_{i=0}^{N-1} y_i(n) \tilde{\Phi}_i(\theta_i(n)) \quad (5.9.3)$$

Pode-se ver que (5.9.3) é uma generalização de (5.9.1), que incorpora as características variantes no tempo dos sons musicais. O próximo passo é então procurar uma maneira de generalizar (5.9.2).

Muitas pesquisas têm sido dedicadas à análise de Fourier de sinais variantes no tempo. A transformada rápida de Fourier de um sinal de comprimento finito $x(k)$ é definida por

$$z_i(k) = \sum_{i=0}^{N-1} x(i) h_k(i) e^{-j2\pi i k / K}, k = 0, 1, \dots, K-1, K < N \quad (5.9.4)$$

onde $h(k)$ é uma função janela. Esta função janela seleciona o segmento de $x(k)$ a ser transformado. Pode-se ver que (5.9.4) é uma generalização variante no tempo de (5.9.2) para uma base de Fourier. Com a escolha apropriada de $h(k)$, o segmento pode ser recuperado exatamente de sua transformada rápida.

Para aplicações de síntese de música, é mais eficiente separar cada função complexa $z_i(k)$ em uma função magnitude $y_i(k)$ e uma função fase $\theta_i(k)$ da qual pode ser obtida a função frequência $\omega_i(k)$. Em particular, um método eficiente de gerar a função fase é integrar (acumular) uma função de frequência linear por partes, a qual por sua vez é obtida pela acumulação de um conjunto de inclinações. Assim só é preciso armazenar o valor inicial da função de frequência, um conjunto de inclinações, e os pontos onde as inclinações mudam de modo a reconstruir a função de aproximação.

Para determinar a funções de frequência e fase, seja a figura 18, mostrando uma função de frequência linear por partes:

Figura 18: Função de frequência linear.

Uma função de frequência linear pode ser escrita como

$$\omega(t) = at + b \quad (5.9.5)$$

Para $t = 0$, $\omega(t) = \omega_0$. Deste modo, (5.9.5) pode ser reescrita como

$$\omega(t) = at + \omega_0 \quad (5.9.6)$$

Para $t = T$, $\omega(t) = \omega_1$. Com isto

$$\omega_1 = aT + \omega_0 \Rightarrow a = \frac{\omega_1 - \omega_0}{T} \quad (5.9.7)$$

Chamando a quantidade α na equação (5.9.7) de $\Delta\omega$:

$$\omega(t) = \Delta\omega t + \omega_0 \quad (5.9.8)$$

Utilizando agora a notação discreta para o tempo, pode-se reescrever (5.9.8) como

$$\omega(n) = \omega(n-1) + \Delta\omega(n_j) T_s, \quad n = 1, 2, \dots, N-1 \quad (5.9.9)$$

onde $\Delta\omega(n_j)$ é a inclinação entre os pontos n_j e n_{j-1} , N é o número de pontos dentro do intervalo e T_s é o período de amostragem do sinal.

Pode-se escrever uma função fase, resultante da integração de (5.9.8) como

$$\theta(t) = \alpha + \beta t + \gamma t^2 \quad (5.9.10)$$

Realizando uma diferenciação em (5.9.10) chega-se a

$$\frac{d}{dt}\theta(t) = \beta + 2\gamma t = \omega(t) \quad (5.9.11)$$

Comparando as eqs. (5.9.8) e (5.9.11) tem-se que

$$\begin{aligned} \beta &= \omega_0 \\ 2\gamma &= \Delta\omega \Rightarrow \gamma = \frac{\Delta\omega}{2} \\ \alpha &= \theta(0) \end{aligned}$$

Com isto, pode-se reescrever (5.9.10) como

$$\theta(t) = \theta(0) + \omega_0 t + \frac{\Delta\omega}{2} t^2 \quad (5.9.12)$$

Reescrevendo (5.9.12) em unidades de tempo discretas, tem-se

$$\theta(n) = \theta(0) + \omega_0 n T_s + \frac{\Delta\omega}{2} n^2 T_s^2 \quad (5.9.13)$$

Para $t = (n-1)T_s$ tem-se

$$\theta(n-1) = \theta(0) + \omega_0 n T_s - \omega_0 T_s + \frac{\Delta\omega}{2} n^2 T_s^2 + \frac{\Delta\omega}{2} n^2 T_s^2 - \Delta\omega n T_s^2 \Rightarrow$$

$$\theta(n-1) = \theta(n) - \omega(n)T_s + \frac{\Delta\omega}{2} T_s^2 \quad (5.9.14)$$

Rearranjando (5.9.14) de modo a isolar $\theta(n)$, tem-se

$$\theta(n) = \theta(n-1) + \omega(n)T_s - \frac{\Delta\omega}{2} T_s^2 \quad (5.9.15)$$

As funções amplitude, também lineares por partes, são acumuladas como em (5.9.9):

$$y(n) = y(n-1) + \Delta y(n_j) T_s, \quad n = 1, 2, \dots, N-1 \quad (5.9.16)$$

Durante o ataque, os pontos podem ser colocados mais próximos para rastrear mais de perto as variações rápidas no tom. De forma similar, podem ser mais espaçados durante o regime de sustentação e manter ainda uma síntese de alta qualidade.

O procedimento adotado é similar à técnica descrita acima. Será utilizada a aproximação linearizada por partes para as funções para uma maior eficiência. Entretanto, existem duas diferenças específicas devido à natureza das funções base de KL.

A primeira diferença é que a síntese de KL precisa ser restrita a uma síntese harmônica, ou seja, todas as funções base precisam ter a mesma fase. Permitindo variações de frequência independentes no meio de um conjunto de funções base de KL ocorre uma cisão harmônica. Isto ocorre quando a frequência instantânea de uma função base é ligeiramente diferente da frequência instantânea de outra. Isto produz um som menos focalizado, como se vários instrumentos estivessem sendo tocados ao mesmo tempo. Na figura 19 tem-se uma idéia deste efeito.

Figura 19: Espectro depois da cisão harmônica.

A menos que haja a intenção de usar isto apropriadamente como um efeito especial, a síntese de KL deve ser restrita a uma síntese harmônica. Apesar de os tons musicais não serem perfeitamente harmônicos, pode-se realizar uma análise e síntese harmônica e obter bons resultados. Por esta razão, $\theta(i)$ é independente de i em (5.9.3).

A segunda diferença é que as funções base de KL precisam ser convenientemente alinhadas com o tom em cada ponto antes de se calcular a correlação. As funções base de KL são derivadas de um conjunto de funções amostra alinhadas em fase. As

características principais do conjunto alinhado são extraídas pelas funções base, de modo que grande parte da energia aparece nas poucas primeiras funções base. Para tirar vantagem desta concentração de energia, o segmento a ser sintetizado deve ter a mesma relação de fase com as funções base como o conjunto original (caso contrário o segmento não é representado apropriadamente pelo conjunto). Então, para que a correlação produza coeficientes úteis, as funções base como um grupo precisam ser alinhadas com o segmento a ser aproximado, preservando a relação de fase com o conjunto original.

Com estas restrições em mente, a função de correlação generalizada pode ser determinada como:

$$y_i(k_j) = \sum_{k=k_j}^{k_j+K} x_\xi(k) \tilde{\phi}_i(\theta(k)) \quad (5.9.17)$$

Em cada ponto k_j , um período local (K amostras) do tom é correlacionado com cada função base, e esta correlação é utilizada para alinhar a primeira função base com o tom. Isto determina um ponto na função fase $\theta(k)$ comum a todas as funções base. A correlação em (5.9.17) é realizada para cada função base, produzindo um ponto na função amplitude associada. O algoritmo utilizado para selecionar os pontos, alinhar o tom e funções base, e calcular a correlação é esboçado abaixo.

1) Obtém-se uma estimativa do período fundamental K do tom. A frequência nominal do tom é definida como sendo

$$\hat{\omega} = \frac{2\pi}{K T_s} \quad (5.9.18)$$

onde T_s é o período de amostragem.

2) Inicialmente, evita-se os transientes rápidos característicos do regime de ataque e realiza-se um alinhamento de fase inicial da primeira função base com o tom perto do início do regime de sustentação. Esta posição é denotada como k_0 . A fase inicial da função base $\theta(k_0)$ é aquela que maximiza a relação

$$\left| \sum_{k_0}^{k_0+K} x(k) \tilde{\phi}_0(\theta(k_0)) + \hat{w}(k - k_0) T_s \right| \quad (5.9.19)$$

Embora a intenção seja reconstruir o tom usando uma função frequência linear por partes, a aproximação constante por partes sobre um período produz erro pequeno para este propósito imediato. O alinhamento é feito com a primeira função base desde que é estatisticamente a mais importante (tem o autovalor mais alto). Desde que a fase é compartilhada, isto alinha as funções base restantes também.

3) No próximo passo será iniciado um loop para calcular a função fase para o resto do tom. O parâmetro P é o número aproximado de períodos pulados entre os pontos onde o sinal é analisado. Este parâmetro pode ser incrementado quando o tom está estável e decrementado se o tom flutua. Em cada ponto, serão testados valores de fase abaixo e acima de um valor estimado de modo a encontrar o valor correto da fase. A faixa de fases de prova pode também ser incrementada ou decrementada dependendo da estabilidade do tom. Inicialmente, pula-se um período para obter o próximo ponto

$$k_1 = k_0 + PK, \quad P = 1 \quad (5.9.20)$$

A faixa de testes inicial utilizada foi de $-\pi/2$ a $+\pi/2$

4) Estima-se a fase verdadeira em k_j assumindo que a frequência seja constante entre os pontos k_{j-1} e k_j

$$\hat{\theta}(k_j) = \theta(k_{j-1}) + \hat{\omega} KP T_s \quad (5.9.21)$$

Dentro da faixa corrente de fases de prova, testa-se valores de fase $\theta^*(k_j)$ menores e maiores que para achar o maior valor absoluto de

$$y_0(k_j) = \sum_{k=k_j}^{k_j+K} x(k) \tilde{\phi}_0(\theta^*(k_j) + \hat{\omega} \cdot (k - k_j) T_s) \quad (5.9.22)$$

Se não é encontrado um máximo, aumenta-se a faixa corrente de fases de teste até que um máximo absoluto seja detetado. Neste caso, o aumento foi de $2\pi/Td$ para cada tentativa frustrada de detecção do máximo absoluto.

5) Utiliza-se este melhor valor de fase para determinar os coeficientes de amplitude para cada uma das funções base

$$y_i(k_j) = \sum_{k=k_j}^{k_j+K} x(k) \tilde{\phi}_i(\theta(k_j) + \hat{\omega} \cdot (k - k_j) T_s), \quad i = 1, 2, \dots, N-1 \quad (5.9.23)$$

6) A inclinação da amplitude é dada por

$$\Delta y(k_j) T_s = \frac{y(k_j) - y(k_{j-1})}{KP} \quad (5.9.24)$$

Como mencionado, a fase é uma função do tempo quadrática por partes se assume-se que a frequência é linear por partes. A inclinação da frequência é calculada rearranjando (5.9.15):

$$\Delta\omega(k_j) T_s^2 = \frac{2(\theta(k_j) - \theta(k_{j-1}) - \omega(k_{j-1}) T_s KP)}{(KP)^2} \quad (5.9.25)$$

Então a frequência é calculada de acordo com

$$\omega(k_j) T_s = \omega(k_{j-1}) T_s + \Delta\omega(k_j) T_s^2 KP \quad (5.9.26)$$

Este valor de frequência é usado em (5.9.25) no próximo passo através do loop (No primeiro passo, assume-se $\omega(k_0) = \hat{\omega}$).

7) Para ajustar o tamanho do passo (P) e a faixa de fases de prova foi utilizada a seguinte estratégia. Se a estimativa para a fase foi bem perto da correta (no caso, foi tolerada uma diferença de $8\pi/Td$), aumenta-se o tamanho do passo de uma unidade, a menos que este exceda um limite escolhido. Caso contrário, decrementa-se o tamanho do passo a menos que este caia abaixo de um piso escolhido. O limite escolhido foi de quatro períodos, e o piso, meio período. Se o limite foi alcançado e a estimativa foi próxima, decrementa-se a faixa de fases de teste para o próximo passo. Se o piso foi alcançado e a suposição foi muito distante, incrementa-se a faixa para o próximo passo. No caso, a faixa de fases foi incrementada ou decrementada de um valor $2\pi/Td$ de cada vez.

8) Volta-se ao passo 4) até que o final do tom seja alcançado. Quando o final do tom é alcançado, retorna-se para a posição k_0 e reseta-se a fase para $\theta(k_0)$. O processo é reiniciado com um passo $P = -1$, e retorna-se ao passo 4) usando passos negativos até que o começo do tom seja alcançado. Neste caso, como as variações do sinal são mais bruscas, optou-se por um teto no valor de P de dois períodos, e o limiar para a sua incrementação ou decrementação foi baixado para $6\pi/Td$.

Nas figuras 20, 21 e 22 abaixo tem-se os gráficos mostrando as saídas do sistema de detecção de envoltória para um tom de violão. Na figura 20 é mostrada a variação da frequência com o tempo, na figura 21, a variação da fase, e na figura 22, a evolução dos coeficientes para os seis primeiros autovetores.

Figura 20: Envoltória de frequência para um tom de violão.

Figura 21: Envoltória de fase para um tom de violão.

Figura 22: Envoltória de amplitude para as 6 primeiras funções base de um tom de violão

5.10. Reconstrução dos sinais

O sistema de análise tem por saídas os autovetores da base de KL e seus respectivos autovalores. Estes dados, juntamente com o sinal original, são utilizados pelo sistema de detecção de envoltória para determinar, em pontos espaçados no tempo, quais os valores de fase, frequência e dos coeficientes dos autovetores que melhor aproximam o sinal original. O sistema de reconstrução toma estes dados e realiza uma interpolação para determinar os parâmetros nos pontos restantes, antes de realizar a reconstrução propriamente dita. O processo ocorre da seguinte maneira:

Em pontos espaçados no tempo, o sistema de detecção de envoltória fornece os valores de:

- coeficientes y_i e a inclinação dos coeficientes $\Delta y(k_i)$ para os i autovetores selecionados para a síntese;
- a frequência ω_i e a inclinação da frequência $\Delta \omega(k_i)$ do sinal no ponto;
- a fase θ_i dos autovetores para a síntese;
- o número de pontos N contidos no intervalo de reconstrução.

A partir destes dados é possível reconstruir o sinal original. Inicialmente, deve ser lembrado que para este trabalho foi adotada uma aproximação linear por partes para as funções de amplitude e frequência dos autovetores, e uma aproximação quadrática por partes para a função de fase.

Com isto, dentro do intervalo, as amplitudes dos autovetores podem ser encontradas por (5.9.7), as frequências por (5.9.5) e as fases por (5.9.6).

O sinal é reconstruído a partir de:

$$x(k) \approx \sum_{i=0}^{N-1} y_i(k) \tilde{\phi}_i(\theta(k)) \quad (5.10.1)$$

onde $\tilde{\phi}_i$ é o i -ésimo autovetor.

Os resultados obtidos com tal sistema foram muito bons. Os programas de detecção de envoltória e reconstrução de sinais conseguiram reconstruir com êxito tanto as formas de onda dos sinais como as suas variações temporais.

Como um exemplo de saída do programa de reconstrução dos sinais, tem-se nas figuras 23 e 24 os sinais original e reconstruído de um tom de guitarra elétrica. No eixo x tem-se o tempo em segundos e no eixo y , a amplitude normalizada. Para a reconstrução foram utilizadas 5 funções base, que resultou num som sintetizado bastante próximo do original.

Figura 23. Tom de guitarra elétrica.

Figura 24. Tom de guitarra elétrica reconstruído com 5 funções base.

5.11. Volta à frequência original.

Como os sinais foram normalizados em frequência para a análise, os sons sintetizados apresentam também a frequência normalizada. É necessário então alterar o período de amostragem novamente, de modo que voltem a ter a mesma frequência fundamental do sinal original. Para isto utilizou-se o sistema descrito no item 5.5, com $L = 103$ e $M = \text{Tor}$.

Com isto, os sinais são agora reconvertidos para o formato WAV para que possam ser ouvidos através do sistema de som.

6. PARTE EXPERIMENTAL.

Com o sistema inteiramente montado, foram realizados alguns testes para verificar o seu desempenho. O espaço amostral para a análise foi composto de seis tipos de sons diferentes: de flauta, violão, guitarra e violino, sendo este último executado de três maneiras diferentes: com arco, pizzicato e vibrato. Para cada um destes tipos foram coletadas 15 notas (duas oitavas), perfazendo um total de 90 sons de teste.

A avaliação do sistema foi dividida em duas etapas: em um primeiro momento, foram realizados testes para verificar as propriedades descritas na parte teórica, ou seja, a influência dos pré-processamentos de normalização em frequência, alinhamento em fase e classificação dos sinais na concentração de energia nos primeiros autovetores. Verificadas estas propriedades, foi feita uma avaliação do produto final, ou seja, dos sons sintetizados. Para isto foram realizados ensaios de comparação auditiva, onde os ouvintes eram submetidos a dois sons sucessivos, sendo um original e o outro sintetizado, não necessariamente nesta ordem. Depois eram convidados a expressarem suas opiniões quanto ao grau de similaridade entre eles. O resultado final foi obtido através da ponderação da opinião de várias pessoas.

6.1. Testes iniciais.

De modo a reduzir a quantidade de cálculos nesta etapa foram selecionados apenas 30 dos 90 sinais originais. Escolheu-se para estes ensaios sons de flauta e de violão. De cada

instrumento foram selecionadas 15 notas (duas oitavas) indo de G2 a G4, sempre com tons naturais (sem sustenidos ou bemóis).

Inicialmente foram realizados testes apenas com um trecho de cada sinal em sua região de sustentação, de modo a verificar a influência dos pré-processamentos de normalização em frequência, e alinhamento em fase e classificação dos sinais na concentração de energia nos primeiros autovetores. Os resultados foram analisados em termos do erro quadrático médio de truncamento e da porcentagem de energia armazenada no primeiro autovetor.

6.1.1. Verificação da influência da normalização em frequência e do alinhamento em fase.

Nesta seção, todos os 30 sinais de teste foram agrupados em uma única base. De modo a isolar a influência de cada um dos pré-processamentos de normalização em frequência e alinhamento em fase, foram realizados três ensaios: o primeiro com os sinais sem processamento nenhum, o segundo com os sinais normalizados em frequência, e o terceiro com os sinais alinhados em fase.

Comparando o erro quadrático médio de truncamento para cada um destes casos, tem-se o gráfico da figura 25. A partir deste pode-se ver claramente que a concentração de energia nos primeiros autovetores é maior no caso de sinais normalizados em frequência se comparado ao caso de sinais sem pré-processamentos, e maior no caso de sinais alinhados em fase se comparado ao caso de sinais normalizados em frequência. Isto significa que adotando os procedimentos de normalização em frequência e alinhamento em fase consegue-se uma redução no número de funções amostra necessário para uma síntese de boa qualidade.

Esses pré-processamentos traduzem-se em uma maior correlação entre as funções amostra, como pode ser visto nas três figuras abaixo. Na figura 26, são mostradas as funções amostra sem nenhum tipo de pré-processamento; na figura 27, as funções amostra foram normalizadas em frequência, e na figura 28, alinhadas em fase.

Figura 25: Comparação do erro quadrático médio de truncamento para os casos de sinais sem pré-processamentos, normalizados em frequência e alinhados em fase.

Figura 26: Funções amostra da base formada com sinais sem pré-processamentos

Figura 27: Funções amostra da base formada com sinais normalizados em frequência

Figura 28: Funções amostra da base formada por sinais alinhados em fase.

Os resultados obtidos nestes três testes iniciais podem ser resumidos na tabela 1. Pode-se verificar, olhando para esta tabela que a concentração de energia no primeiro autovetor é bastante incrementada quando são aplicados os procedimentos de normalização em frequência e alinhamento em fase.

Tabela 1: Quadro comparativo para a verificação da influência da normalização em frequência e alinhamento em fase no desempenho da transformada.

	Sinais sem pré-processamentos	Sinais normalizados em frequência	Sinais alinhados em fase
% da energia total do sinal no 1º autovetor	20,66	47,16	85,02
Nº de sinais na base	30	30	30

6.1.2. Verificação da influência da classificação dos sinais.

Foi mostrado por Stapleton & Bass (1988) que quanto mais alta a correlação entre os sinais da base de KL, maior a concentração de energia nos primeiros autovetores.

Para que os resultados das correlações possam ser comparados, procedeu-se a uma equalização da energia dos sinais; na janela em que foram tomados, todos os sinais foram normalizados para terem energia unitária. De modo similar, e pelas mesmas razões que o problema de alinhamento em fase, os sinais foram analisados sob uma janela de 103 pontos.

Com os resultados das correlações em mãos, procedeu-se a separação dos mesmos em grupos. Num primeiro grupo foram colocados sinais com correlação maior que 0.9 entre todos os elementos. Dos sinais restantes, separou-se um segundo grupo, cujos elementos apresentaram correlação maior que 0.5. Os sinais restantes foram colocados em um terceiro grupo. Com esta classificação, o primeiro grupo ficou com 15 sinais, o segundo com 9, e o terceiro com 6.

Como era de se esperar, a classe formada por sinais altamente correlacionados ($\gamma_{xy} \geq 0.9$) apresentou os melhores resultados, enquanto que a classe formada por sinais de correlação baixa ($\gamma_{xy} \leq 0.5$) apresentou o menor desempenho. Estas conclusões podem ser vistas no gráfico da figura 29.

O grau de correlação dos sinais nas diferentes bases pode ser analisado visualmente através das figuras 30, 31 e 32. Na figura 30, tem-se os sinais da base formada com funções amostra de correlação baixa ($\gamma_{xy} \leq 0.5$), na figura 31, os da base formada com sinais de correlação “moderada” ($0.5 < \gamma_{xy} \leq 0.9$), e na figura 32, os sinais da base com correlação “alta” ($\gamma_{xy} \geq 0.9$).

Figura 29: Comparação do erro quadrático médio de truncamento para verificar a influência da classificação dos sinais na concentração de energia nos primeiros autovetores.

Figura 30: Funções amostra da base formada por sinais com correlação menor que 0.5.

Figura 31: Funções amostra da base formada por sinais com correlação entre 0.5 e 0.9

Figura 32: Funções amostra da base formada por sinais com correlação maior que 0.9.

A classificação dos sinais ajuda na concentração de energia nos primeiros autovetores. Neste caso, a energia armazenada no primeiro autovetor aumentou de 58.42% para sinais com correlação menor que 0.5 para 96.51% nos sinais com correlação acima de 0.9. Estes resultados podem ser vistos na tabela 2.

Tabela 2: Quadro comparativo para a verificação da influência da correlação entre as funções amostra no desempenho da transformada.

	$\gamma_{xy} < 0.5$	$0.5 \leq \gamma_{xy} < 0.9$	$\gamma_{xy} \geq 0.9$
% da energia total do sinal no 1º autovetor	58.42	83.05	96.51
Nº de sinais na base	6	9	15

6.1.3. Apreciação dos resultados.

A partir dos resultados apresentados nas seções anteriores, pode-se concluir que realmente os pré-processamentos de normalização em frequência, alinhamento em fase e classificação dos sinais auxiliam na concentração de energia nos primeiros autovetores.

Como uma ilustração prática disto, é interessante verificar o desempenho do sistema na prática, ou seja, na reconstrução de um sinal real. Para isto, tomou-se um tom de violão e realizou-se a sua reconstrução a partir de várias bases as quais, como nos testes anteriores, continham sinais de flauta e violão. De modo a poder comparar os resultados, realizou-se a reconstrução com o mesmo número de autovetores em todos os casos. Em particular, nas figuras 33 a 36, os sinais foram reconstruídos com 5 autovetores.

Figura 33. Tom de violão reconstruído a partir de base formada por funções amostra sem pré-processamentos.

Figura 34. Tom de violão reconstruído a partir de base formada por funções amostra normalizadas em frequência.

Figura 35. Tom de violão reconstruído a partir de base formada por funções amostra alinhadas em fase

Figura 36. Tom de violão reconstruído a partir de base formada por funções amostra com correlação maior que 0.9.

6.2. Testes de audição.

Uma vez que foi comprovada a eficácia da utilização dos pré-processamentos de normalização em frequência, alinhamento em fase e classificação dos sinais, estes foram incorporados definitivamente ao procedimento de análise. O próximo passo então seria verificar realmente o número de funções base necessário para uma reconstrução de boa qualidade dos sons originais. Como foi visto no capítulo 2, esta avaliação ainda deve ser feita através de testes de comparação subjetiva de similaridade em ensaios de audição, uma vez que ainda não é possível quantizar estes parâmetros.

6.2.1. Material para os testes.

Os 90 sinais do espaço amostral foram normalizados em frequência, classificados e alinhados em fase. No procedimento de classificação foram geradas 5 classes:

1. sinais com $\gamma_{xy} \geq 0.8$ (sinais de flauta e violino pizzicato)
2. sinais com $\gamma_{xy} \geq 0.8$ (sinais de violão e guitarra elétrica)
3. sinais com $0.5 \leq \gamma_{xy} \leq 0.8$ (sinais de violão, violino, flauta, violino pizzicato e violino vibrato)
4. sinais com $0.3 \leq \gamma_{xy} \leq 0.5$ (sinais de violão, violino, violino pizzicato e violino vibrato)
5. sinais com $\gamma_{xy} \leq 0.3$ (sinais de violão, violino, violino pizzicato, violino vibrato, flauta e guitarra)

De cada uma destas foram selecionados pelo menos dois sinais para os testes de audição, gerando um total de 18 tons de teste. Cada um deles foi reconstruído com um número variável de funções base, a fim de verificar a partir de que ponto a síntese deixava de ter qualidade. Este número variou desde 2 até o número total de autovetores na base. Com isto, foram gerados 176 sinais de teste.

Os sons de flauta reconstruídos soaram bastante artificiais de início. A exemplo do trabalho de Stapleton & Bass [1988] adicionou-se a estes um pouco de ruído branco de baixa intensidade, o que os tornou mais naturais.

6.2.2. Procedimentos.

Os testes consistiram em sessões de audição onde os ouvintes escutavam dois sons consecutivos, um original e outro reconstruído, não necessariamente nesta ordem. Depois eram convidados a avaliar a semelhança entre um e outro, atribuindo as seguintes notas:

1. se os sinais fossem diferentes
2. se fossem pouco parecidos
3. se fossem bem parecidos
4. se fossem iguais

Os ouvintes também tinham a liberdade de repetir os sons caso desejassem ouvir melhor. De forma a não induzir resultados, os sinais foram executados de forma aleatória, ou seja, alternando-se os instrumentos, o número de autovetores para a reconstrução e as notas. O número total de pares de sons para a análise foi de 158.

Os testes foram feitos com vários ouvintes, e ao final, foi feita uma média dos resultados. Foi considerado o menor número de funções base que obtivesse avaliação 3 (sinais bastante parecidos) como a quantidade suficiente de autovetores para uma síntese de alta qualidade. Isto gerou os valores da Tabela 3.

Tabela 3: Numero suficiente de autovetores para síntese de alta qualidade para cada uma das classes.

Bases	Correlação entre os sinais	Número de funções base para reconstrução
1	$\gamma_{xy} \geq 0.8$	4
2	$\gamma_{xy} \geq 0.8$	5
3	$0.5 \leq \gamma_{xy} \leq 0.8$	8
4	$0.3 \leq \gamma_{xy} \leq 0.5$	5
5	$\gamma_{xy} \leq 0.3$	7

6.2.3. Apreciação dos resultados.

Segundo os dados fornecidos pela tabela acima, pode-se ver que os resultados obtidos confirmam os resultados pretendidos de síntese de boa qualidade com poucas

funções base. O número máximo de funções requerida para uma síntese de boa qualidade foi de 8, no caso da base 3, e o mínimo, de apenas 4 para a base 1 (sinais altamente correlacionados). Deste modo, com um total de 29 funções base conseguiu-se reconstruir 90 sons musicais.

Os sinais de violino e violino com vibrato foram os que apresentaram a melhor qualidade de síntese, sendo em alguns casos indistinguíveis dos originais. Já os sons percussivos (violão, guitarra e violino pizzicato) não tiveram a mesma performance. Em alguns casos houve a presença de um zumbido ao final da parte de decaimento, como reportado por Stapleton & Bass (1988).

Para ilustrar este ponto sejam as figuras 37 a 40 abaixo. Nestas tem-se o espectro dos sinais original e reconstruído de duas regiões distintas de um tom de guitarra. No eixo x tem-se a frequência em hertz, e no eixo y, a amplitude relativa de cada harmônica.

Os gráficos das figuras 37 e 38 mostram os espectros dos sinais original e reconstruído na região de ataque dos mesmos. Pode-se verificar que o conteúdo harmônico dos dois sinais é bem parecido nesta região.

Analisando-se agora os gráficos das figuras 39 e 40, nota-se que para o sinal reconstruído, há uma presença maior nas componentes de média/alta frequências, comparando-se com o sinal original. Isto explica o chiado ouvido durante a execução.

Figura 37: Espectro de um tom de guitarra durante o ataque.

Figura 38: Espectro do tom reconstruído, durante o ataque.

Figura 39: Espectro de um tom de guitarra no final da região de decaimento.

Figura 40: Espectro do tom reconstruído no final da região de decaimento.

Este último problema foi analisado mais de perto, e uma das razões apontadas para isto foi que os sons percussivos apresentam uma variação espectral muito grande durante o seu tempo de vida. As harmônicas mais altas são mais amortecidas do que as mais baixas no decorrer do tempo. Como as funções base geradas pela transformada de KL apresentam espectros mais amplos (não são limitados a apenas uma harmônica como no caso da Transformada de Fourier), esta variação harmônica parece não ser reproduzida fielmente. Nas figuras abaixo tem-se o espectro dos quatro últimos autovetores utilizados na reconstrução do tom de guitarra acima.

Figura 41. Espectro do décimo autovetor da base onde foi reconstruído o tom de guitarra.

Figura 42. Espectro do décimo primeiro autovetor da base onde foi reconstruído o tom de guitarra.

Figura 43. Espectro do décimo segundo autovetor da base onde foi reconstruído o tom de guitarra.

Figura 44. Espectro do décimo terceiro autovetor da base onde foi reconstruído o tom de guitarra.

Com isto, chega-se à conclusão de que os autovetores mais altos (com os autovalores menores) são os responsáveis pelo zumbido ouvido ao final do tom. A solução encontrada para minimizar este efeito foi reconstruir a região de ataque com um número grande de autovetores, e a região de decaimento com um número menor.

Outra conclusão que pode ser tirada disto é que, pelo fato de as funções base geradas pela transformada de KL serem intrinsecamente dependentes das funções amostra, não se consegue reproduzir adequadamente formas de onda muito diferentes daquelas das funções amostra analisadas. Desta maneira, sons que apresentam formas de onda que variam muito durante o seu tempo de vida (como é o caso dos sons

percussivos) devem ser tratados com bastante cuidado. Uma das maneiras de contornar este problema seria tomar as funções amostra não logo após o ataque, mas sim na metade do tom (para sons percussivos). Isto faria com que a parte analisada do sinal contivesse aspectos tanto do ataque quanto do decaimento.

Adotando os dois procedimentos acima, conseguiu-se uma grande melhora no som sintetizado.

7. CONCLUSÕES.

Neste trabalho foi implementado e testado um modelo de síntese digital de sons musicais usando o modelo de síntese aditiva baseada em análise com funções base de Karhunen-Loève.

Tradicionalmente considera-se a técnica de síntese aditiva como a que produz sons de mais alta qualidade. O método mais comumente empregado com esta técnica é a síntese aditiva de Fourier, com funções base senoidais. Entretanto, a carga computacional imposta por este método tem limitado o seu uso para aplicações em tempo real.

Foi demonstrado que um conjunto de funções base de Karhunen-Loève é aquele que apresenta o mínimo erro médio quadrático de truncamento para o conjunto de funções amostra da qual foi derivado. Ainda, a base de Karhunen-Loève apresenta todas as propriedades de qualquer conjunto de funções base ortogonais. Isto permite a representação das funções amostra com menos funções base do que qualquer outro conjunto de funções base ortogonais.

Os testes realizados mostraram que, comparada à síntese aditiva de Fourier, a técnica de Karhunen-Loève apresenta um desempenho melhor no que diz respeito ao compromisso entre a qualidade dos sons sintetizados e o esforço computacional requerido. Com 29 funções base conseguiu-se representar 90 sons musicais, sendo que este número poderia ser bem maior se fossem tomadas mais notas dos mesmos instrumento (por exemplo com sustenidos e bemóis).

A transformada foi adaptada para reproduzir não apenas segmentos relativamente curtos, mas tons musicais inteiros, variantes no tempo. Para isto foram formadas bases

com um período de cada sinal tomados na sua região de sustentação. Este trecho representaria as características do som sendo sintetizado. De modo a otimizar o desempenho e reduzir ainda mais o número de funções base necessárias à síntese foram aplicadas algumas técnicas de processamento digital de sinais tais como normalização em frequência e alinhamento em fase.

Os sons foram agrupados em classes antes de serem analisados, o que resultou numa melhor concentração de energia nas funções base, reduzindo assim a quantidade destas para a síntese. Com isto consegue-se alguma economia no hardware necessário para operação em tempo real.

Os sinais de violino e violino com vibrato foram os que apresentaram a melhor qualidade de síntese, sendo em alguns casos indistinguíveis dos originais. Isto é bastante positivo, desde que a técnica aditiva tradicionalmente não lida bem com sons que apresentam vibrato.

O sistema apresentou alguns problemas na reconstrução de sons percussivos (sons onde a parte de sustentação está ausente), apresentando um zumbido ao final da região de decaimento. Uma das causas apontadas para este baixo desempenho foi que o sistema de análise não consegue reproduzir de maneira fiel a grande variação harmônica sofrida por estes sinais durante o seu tempo de vida. Isto se deve principalmente à presença dos autovetores mais altos (com autovalores menores). Dois procedimentos foram incorporados ao sistema de síntese para melhorar a qualidade dos sons sintetizados:

- a região de ataque foi sintetizada com um número maior de funções base do que a região de decaimento. Isto faz com que se preserve o ataque e minimize o zumbido na região de decaimento.
- ao invés de pegar funções amostra logo após a região de ataque, foram tomadas amostras do sinal no meio da região de decaimento. Isto garante que as funções base tenham características das duas regiões.

Com a técnica de síntese aditiva baseada em análise elimina-se o procedimento cansativo de tentativa-e-erro para a derivação dos parâmetros de síntese. O método utilizado para a detecção das envoltórias de fase e amplitude das funções base funcionou bastante bem, preservando adequadamente os transientes iniciais da porção de ataque dos sons que, como visto anteriormente, desempenham um papel essencial na caracterização destes sinais.

Existem duas desvantagens inerentes à utilização das funções de Karhunen-Loève para a síntese de sons musicais:

- A síntese com KL é limitada a uma síntese harmônica, exceto para efeitos especiais, o que faz imaginar que a síntese de sons inarmônicos seja dificultada;
- O *aliasing* é melhor tratado com a síntese de Fourier, onde as harmônicas problemáticas são simplesmente eliminadas, enquanto que com KL é necessário limitar os sinais em banda.

As vantagens inerentes à técnica aditiva, que são a flexibilidade e o controle sobre as funções base utilizadas na síntese são preservadas com a técnica de KL.

Como sugestões para trabalhos futuros, poderia-se comparar os resultados da técnica aqui desenvolvida com outros modelos (subtrativos, não-lineares, etc.) de modo a fornecer um panorama melhor sobre quais técnicas funcionam melhor em que situações. Também poderia-se empregar esta técnica em síntese de voz. Testes preliminares com o sistema desenvolvido mostraram bons resultados. O algoritmo de detecção de envoltória poderia ser refinado, através de uma aproximação quadrática por partes para os coeficientes y_i dos autovetores, já que esta acompanha melhor as curvas não lineares. É preciso verificar, entretanto, qual o custo computacional de tal refinamento no tempo de reconstrução do sinal, para averiguar a sua viabilidade para aplicações em tempo real.

7. BIBLIOGRAFIA.

- AHMED, N. & RAO, K. R. (1975) *Orthogonal transforms for digital signal processing*. pp 180-224. Springer-Verlag.
- ANSARI, Rachid & LIU, Bede. (1983) Efficient sampling rate alteration using recursive (IIR) digital filters. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. ASSP-31, n. 6. pp.1366-1373, December. BRUIJN, A. de. (1978) Timbre classification of complex tones. *Acustica*. v. 40, n. 2, pp. 108-114. June.
- CHRISTENSEN, Ronald A. & HIRSCHMAN, Alan D. (1979) Automatic phase alignment for the Karhunen-Loève expansion. *IEEE Transactions on Biomedical Engineering*. v. BME-26, n. 2. pp. 94-99. February.
- CHOWNING, J. M. (1973) The synthesis of complex audio spectra by means of frequency modulation. *Journal of Audio Engineering Society*. v. 21, n 7, pp. 526-534. September.
- GOEDDEL, T. W. and BASS, S.C. (1984) High quality synthesis of musical voices in discrete time. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. ASSP-32, n. 3. pp. 623-633. June.
- GREY, J. M. (1977) Multidimensional scaling of musical timbres. *Journal of Acoustical Society of America*, v. 61, n. 5. pp. 1270-1277. May.

- GREY, J. M. (1978) Timbre discrimination in musical patterns. *Journal of Acoustical Society of America*, v. 64, n. 2. pp. 467-472. August.
- GREY, J. M. & GORDON, J. W. (1978) Perceptual effects of spectral modifications of musical timbres. *Journal of Acoustical Society of America*, v. 63, n. 5. pp. 1493-1500. May.
- GREY, J. M. and MOORER, J. (1977) Perceptual evaluations of synthesized instrument tones. *Journal of Acoustical Society of America*, v. 62, n. 2. pp. 454-462. August.
- HAKEN, Lippold. (1992) Computational methods for real-time Fourier synthesis. *IEEE Transactions on Signal Processing*, v. 40, n. 9. pp. 2327-2329. September.
- IVERSON, Paul and KRUMHANSL, Carol L. (1993) Isolating the dynamic attributes of musical timbre. *Journal of Acoustical Society of America*, v. 94, n. 5. pp. 2595-2603. November.
- JUSTICE, James H. (1979) Analytic signal processing in music computation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. ASSP-27, n. 6. pp. 670-684. December.
- LAROCHE, Jean & MEILLIER, Jean-Louis. (1994) Multichannel excitation/filter modeling of percussive sounds with application to the piano. *IEEE Transactions on Speech and Audio Processing*, v. 2, n. 2. pp. 329-344. April.
- MAKHOUL, J. (1975) Linear prediction: A tutorial review. *Proceedings of the IEEE*, vol 63, pp. 561-580. April.

- MARKEL, J. D. & GRAY Jr., A. H. (1974) A linear prediction vocoder simulation based on the autocorrelation method. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. ASSP-22, pp.124-134. April.
- MITSUHASHI, Yasuhiro. (1982) Piecewise interpolation technique for audio signal synthesis. *Journal of Audio Engineering Society*. v. 30, n. 4. April.
- MOORER, James A. (1977) Signal processing aspects of computer music - a survey. *Proceedings of the IEEE*. v. 8, n. 8. pp. 1108-1136. August.
- POLLARD, H. F. and JANSSON, E. V. (1982) Analysis and assesment of musical starting transients. *Acustica*, v. 51, n. 5. pp. 249-262. October.
- POLLARD, H. F. (1988) Feature analysis of musical sounds. *Acustica*, v. 65, n. 5. pp. 232-244. April.
- PREIS, A. (1984) An attempt to describe the parameter determining the timbre of steady-state harmonic complex tones. *Acustica*, v. 55, n. 1. pp.1-13. April.
- PRESS, Willian H., FLANNERY, Brian P., TEUKOLSKY, Saul A., VETTERLING, Willian T. (1987) *Numerical recipes - the art of cientific computing*. Cambridge University Press. Cambridge.
- RABINER, Lawrence R. *et al.* (1972) Terminology in digital signal processing. *IEEE Transactions on Audio and Eletroacoustics*, v. AU-20, n. 5, pp. 322-336. December.

RABINER, Lawrence R. (1977) On the use of autocorrelation analysis for pitch detection. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. ASSP-25, n. 1. pp. 24-33. February.

RISSET, J. C. and MATHEWS, V. (1969) Analysis of musical instrument tones. *Physics Today*. v. 22, n. 2. pp. 26-30. February.

RUSCHIONI, Ruggero Andrea (1993) Música digital: histórico e perspectivas. In: WORKSHOP SOBRE COMPUTAÇÃO DE ALTO DESEMPENHO PARA PROCESSAMENTO DE SINAIS. São Carlos, 1993. *Anais*. São Carlos. 1993. pp. 178-191.

SCHAFER, Ronald W., RABINER, Lawrence R. A. (1973) Digital signal processing approach to interpolation. *Proceedings of the IEEE*, v. 61, n. 6. pp. 692-702. June.

SCHAFER, R. W., RABINER, L. R., and HERRMANN, O. (1975) FIR digital filter banks for speech analysis. *The Bell System Technical Journal*, v. 54, n. 3. pp. 531-544. March.

STANFORD UNIVERSITY. (1994) *Music synthesis approaches sound quality of real instruments*. Newsgroups: su.news-service.press-releases.

STAPLETON, John C. & BASS, Steven Craig. (1986) Synthesis of musical tones based on the Karhunen-Loève transform. *IEEE Transactions on Acoustic, Speech and Signal Processing*, v. 36, n. 3. pp. March.

STOLL, G. and PARNCUTT, R. (1987) Harmonic relationship in similarity judgements of nonsimultaneous complex tones. *Acustica*. v. 63, n. . pp. 111-119.

APÊNDICE 1

Método dos multiplicadores de Lagrange

Uma condição necessária para que uma função contínua e diferenciável $f(x_1, x_2, \dots, x_n)$ de n variáveis $x_1, x_2, \dots, x_n$ tenha um máximo ou um mínimo em um ponto interior a uma região é

$$df = \frac{df}{dx_1} + \frac{df}{dx_2} + \dots + \frac{df}{dx_n} = 0 \quad (1)$$

naquele ponto, para todos os valores permitidos de $dx_1, dx_2, \dots, dx_n$.

Em um ponto satisfazendo (1), a função f é dita como sendo estacionária.

Se as n variáveis são todas independentes, as n diferenciais podem ser determinadas arbitrariamente, e segue então que (1) é equivalente às n condições

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0 \quad (2)$$

Se porém as n variáveis não forem independentes, mas relacionadas por N condições da forma

$$\phi_K(x_1, x_2, \dots, x_n) = 0, \quad K = 1, 2, \dots, N$$

então, pelo menos teoricamente, estas N equações podem ser resolvidas para expressar N das variáveis em termos das $(n-N)$ variáveis restantes, e portanto para expressar f e df em termos de $(n-N)$ variáveis independentes e suas diferenciais.

Alternativamente, pode-se obter N relações lineares entre as n diferenciais por diferenciação. Estas condições permitem exprimir N diferenciais como combinações lineares das diferenciais das $(n-N)$ variáveis independentes.

Se (1) é expressa em termos destas diferenciais, os seus coeficientes precisam então desaparecer, fornecendo $(n-N)$ condições para valores estacionários de f , os quais suplementam as N condições de restrição.

Um procedimento mais conveniente neste caso consiste na introdução dos chamados multiplicadores de Lagrange. Para ilustrar o seu uso, seja o problema de obter valores estacionários de $f(x,y,z)$,

$$df \equiv f_x dx + f_y dy + f_z dz = 0 \quad (3)$$

sujeita às duas restrições

$$\phi_1(x,y,z) = 0 \quad (4a)$$

$$\phi_2(x,y,z) = 0 \quad (4b)$$

Desde que as três variáveis x , y e z precisam satisfazer as duas condições auxiliares (4a) e (4b), apenas uma variável pode ser considerada independente. As equações (4a,b) implicam nas relações diferenciais

$$\phi_{1x} dx + \phi_{1y} dy + \phi_{1z} dz = 0 \quad (5a)$$

$$\phi_{2x} dx + \phi_{2y} dy + \phi_{2z} dz = 0 \quad (5b)$$

O procedimento esboçado acima poderia consistir em resolver primeiramente (5a,b) para dx e dy em termos de dz (se isto for possível) e introduzir os resultados em (3), para fornecer um resultado da forma

$$df \equiv (...) dz$$

Desde que dz pode ser determinada arbitrariamente, o desaparecimento da expressão indicada entre parênteses nesta forma é a condição desejada para que f seja estacionária quando (4a,b) são satisfeitas.

Como um procedimento alternativo, multiplica-se primeiramente (5a) e (5b) por quantidades λ_1 e λ_2 , respectivamente. Soma-se então o resultado obtido com (3). Desde que os lados direitos das três equações são todos nulos, pode-se escrever

$$(f_x + \lambda_1 \phi_{1x} + \lambda_2 \phi_{2x}) dx + (f_y + \lambda_1 \phi_{1y} + \lambda_2 \phi_{2y}) dy + (f_z + \lambda_1 \phi_{1z} + \lambda_2 \phi_{2z}) dz = 0 \quad (6)$$

para valores arbitrários de λ_1 e λ_2

Determina-se λ_1 e λ_2 de modo que dois dos parênteses em (6) desapareçam*.

Então o diferencial multiplicando o parêntese restante pode ser determinado arbitrariamente, e portanto este parêntese precisa também desaparecer. Desta maneira:

$$\frac{\partial f}{\partial x} = \lambda_1 \frac{\partial \phi_1}{\partial x} + \lambda_2 \frac{\partial \phi_2}{\partial x} = 0 \quad (7a)$$

$$\frac{\partial f}{\partial y} = \lambda_1 \frac{\partial \phi_1}{\partial y} + \lambda_2 \frac{\partial \phi_2}{\partial y} = 0 \quad (7b)$$

$$\frac{\partial f}{\partial z} = \lambda_1 \frac{\partial \phi_1}{\partial z} + \lambda_2 \frac{\partial \phi_2}{\partial z} = 0 \quad (7c)$$

As equações (7a,b,c) e (4a,b) formam um sistema de 5 equações determinando x , y , z , λ_1 e λ_2 . As quantidades λ_1 e λ_2 são conhecidas como multiplicadores de Lagrange. A

* Pode ser mostrado que se isto não for possível, então as funções ϕ_1 e ϕ_2 devem ser funcionalmente dependentes, de modo que as duas restrições (4a,b) podem ser ou equivalentes ou incompatíveis.

sua introdução frequentemente simplifica a álgebra relevante em problemas do tipo considerado. Em muitas aplicações, eles também têm um significado físico.

É interessante notar que as condições em (7) são as condições para que

$$f + \lambda_1 \phi_1 + \lambda_2 \phi_2$$

seja estacionária quando não existe nenhuma restrição.

O procedimento acima é aplicável sem modificação ao caso geral de n variáveis e $(N < n)$ restrições.

Apêndice 2

Método das potências para o cálculo de ϕ_1

O método das potências é uma técnica para calcular iterativamente a função base principal da expansão de Karhunen-Loève a partir de um conjunto de dados $\{\psi_i(t), i = 1, 2, \dots, N\}$.

As funções base $\phi_i(t)$ satisfazem a seguinte equação integral (assumindo, por simplicidade que $\tau_i = 0$ para todo i):

$$\sum_{i=1}^N \psi_i(t') \int_0^T \psi_i(t) \phi_i(t) dt = \lambda_1 \phi_1(t) \quad (1)$$

A implementação em microcomputadores requer a representação das funções temporais como vetores de N elementos, cada um com n pontos, $\{\psi_i, i = 1, 2, \dots, N\}$. Deste modo a equação (1) pode ser reescrita como

$$\sum_{i=1}^N \vec{\psi}_i (\vec{\psi}_i \cdot \vec{\phi}_1) = \lambda_1 \vec{\phi}_1 \quad (2)$$

ou equivalentemente,

$$A \vec{\phi}_1 = \lambda_1 \vec{\phi}_1 \quad (3)$$

onde A é a matriz de correlação $n \times n$

$$A = \sum_{i=1}^N \vec{\psi}_i \vec{\psi}_i^T$$

A equação (3) pode gerar $\vec{\phi}_1$ de uma maneira iterativa através do algoritmo do método das potências:

$$A\vec{v}(k) = \vec{u}^{(k+1)}$$

$$\text{e} \tag{4}$$

$$\vec{v}^{(k+1)} = \frac{\vec{u}^{(k+1)}}{\|\vec{u}^{(k+1)}\|}, i = 1, 2, \dots$$

Pode ser mostrado que $\vec{v}^{(k+1)}$ aproxima um vetor proporcional a $\vec{\phi}_1$ a medida que k aumenta. A equação (4) requer que se calcule os elementos da matriz A , mas é fácil evitar este problema transformando (2) em um algoritmo iterativo que calcule produtos escalares de vetores ao invés de operações matriciais.

$$\sum_{i=1}^N \vec{\psi}_i (\vec{\psi}_i \cdot \vec{v}^{(k)}) = \vec{u}^{(k+1)} \tag{5}$$

$$\vec{v}^{(k+1)} = \frac{\vec{u}^{(k+1)}}{\|\vec{u}^{(k+1)}\|}, k = 1, 2, \dots \tag{6}$$

e onde $\vec{u}^{(1)} = \vec{v}^{(1)}$ é algum vetor arbitrário. A taxa de convergência depende de $\ln(\lambda_1/\lambda_2)$, onde λ_2 é o autovalor associado ao segundo vetor base principal da expansão de Karhunen-Loève $\vec{\phi}_2$.

APÊNDICE 3

Redução de uma matriz simétrica para a forma tridiagonal.

A estratégia ótima para encontrar autovalores e autovetores é, primeiramente, reduzir a matriz a uma forma mais simples, e só então iniciar um procedimento iterativo. Para matrizes simétricas, a forma preferida é a tridiagonal. O algoritmo de Householder reduz uma matriz simétrica ($n \times n$) A para a forma tridiagonal através de $(n - 2)$ transformações ortogonais. Cada transformação elimina a parte requerida de uma coluna e uma linha inteira correspondente. O ingrediente básico é uma matriz de Householder P , a qual tem a forma:

$$P = I - 2w \cdot w^T \quad (1)$$

onde w é um vetor real com $|w|^2 = 1$. (Na notação presente, o produto matricial de dois vetores, a e b é escrito como $a \cdot b^T$, enquanto o produto escalar dos vetores é escrito como $a^T \cdot b$.) A matriz P é ortogonal pois

$$\begin{aligned} P^2 &= (I - 2w \cdot w^T) \cdot (I - 2w \cdot w^T) \\ &= I - 4w \cdot w^T + 4w(w^T \cdot w) \cdot w^T \\ &= I \end{aligned} \quad (2)$$

Portanto, $P = P^{-1}$. Mas $P^T = P$, provando a ortogonalidade.

$\mathbf{P}$ pode ser reescrito como

$$\mathbf{P} = \mathbf{I} - \frac{\mathbf{u} \cdot \mathbf{u}^T}{H} \quad (3)$$

onde o escalar H é dado por

$$H \equiv \frac{1}{2} |\mathbf{u}|^2 \quad (4)$$

e $\mathbf{u}$ pode ser agora um vetor qualquer. Supondo que $\mathbf{x}$ é o vetor composto pela primeira coluna de $\mathbf{A}$. Escolhe-se

$$\mathbf{u} = \mathbf{x} \mp |\mathbf{x}| \mathbf{e}_1 \quad (5)$$

onde $\mathbf{e}_1$ é o vetor unitário $[1, 0, \dots, 0]^T$, e a escolha de sinais será feita mais tarde. Então

$$\begin{aligned} \mathbf{P} \cdot \mathbf{x} &= \mathbf{x} - \frac{\mathbf{u}}{H} \cdot (\mathbf{x} \mp |\mathbf{x}| \mathbf{e}_1)^T \cdot \mathbf{x} \\ &= \mathbf{x} - \frac{2\mathbf{u} \cdot (|\mathbf{x}|^2 \mp |\mathbf{x}|x_1)}{2|\mathbf{x}|^2 \mp 2|\mathbf{x}|x_1} \end{aligned} \quad (6)$$

$$= \mathbf{x} - \mathbf{u}$$

$$= \pm |\mathbf{x}| \mathbf{e}_1$$

Isto mostra que a matriz $\mathbf{P}$ de Householder age em um dado vetor $\mathbf{x}$ de modo a zerar todos os seus elementos, exceto o primeiro.

Para reduzir uma matriz simétrica A para a forma tridiagonal, escolhe-se o vetor x de modo que a primeira matriz de Householder seja os primeiros $(n - 1)$ elementos da primeira coluna. Então os $(n - 2)$ elementos inferiores serão zerados:

$$\begin{aligned} P_1 \cdot A &= \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & & & & \\ 0 & & & & \\ \dots & & & & \\ 0 & & & & \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & & & & \\ a_{31} & & & & \\ \dots & & & & \\ a_{n1} & & & & \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ k & & & & \\ 0 & & & & \\ \dots & & & & \\ 0 & & & & \end{bmatrix} \quad \text{irrelevante} \end{aligned} \quad (7)$$

As matrizes foram escritas na forma particionada, com $^{(n-1)}P$ denotando uma matriz de Householder com dimensões $(n - 1) \times (n - 1)$. A quantidade k é simplesmente mais ou menos a magnitude do vetor $[a_{21}, \dots, a_{n1}]^T$.

A transformação ortogonal completa é agora

$$A' = P \cdot A \cdot P = \begin{bmatrix} a_{11} & k & 0 & \dots & 0 \\ k & & & & \\ 0 & & & & \\ \dots & & & & \\ 0 & & & & \end{bmatrix} \quad \text{irrelevante} \quad (8)$$

onde foi usado o fato de que $P^T = P$.

Agora escolhe-se o vetor x para a segunda matriz de Householder como sendo os $(n - 2)$ elementos de baixo da segunda coluna, e a partir dela constroi-se

$$\mathbf{P}_2 \equiv \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & & & \\ \dots & \dots & & & \\ 0 & 0 & & & \end{bmatrix} \begin{matrix} \\ \\ (n-2) \mathbf{P}_2 \\ \\ \end{matrix} \quad (9)$$

O bloco de identidade no canto superior direito assegura que a tridiagonalização conseguida no primeiro passo não será destruída por este, enquanto que a matriz de Householder $(n - 2)$ -dimensional ${}^{(n-2)}\mathbf{P}_2$ cria uma linha e uma coluna adicionais da saída tridiagonal. Claramente, uma sequência de $(n - 2)$ transformações deste tipo irá reduzir a matriz $\mathbf{A}$ à forma tridiagonal.

Ao invés de realizar as multiplicações matriciais em $\mathbf{P} \cdot \mathbf{A} \cdot \mathbf{P}$, calcula-se um vetor

$$\mathbf{p} \equiv \frac{\mathbf{A} \cdot \mathbf{u}}{H} \quad (10)$$

Então

$$\mathbf{A} \cdot \mathbf{P} = \mathbf{A} \cdot \left(\mathbf{I} - \frac{\mathbf{u} \cdot \mathbf{u}^T}{H} \right) = \mathbf{A} - \mathbf{p} \cdot \mathbf{u}^T$$

$$\mathbf{A}' = \mathbf{P} \cdot \mathbf{A} \cdot \mathbf{P} = \mathbf{A} - \mathbf{p} \cdot \mathbf{u}^T - \mathbf{u} \cdot \mathbf{p}^T + 2K \mathbf{u} \cdot \mathbf{u}^T$$

onde o escalar K é definido por

$$K = \frac{\mathbf{u}^T \cdot \mathbf{p}}{2H} \quad (11)$$

Se for definido

$$\mathbf{q} \equiv \mathbf{p} - K \mathbf{u} \quad (12)$$

então tem-se

$$\mathbf{A}' = \mathbf{A} - \mathbf{q} \cdot \mathbf{u}^T - \mathbf{u} \cdot \mathbf{q}^T \quad (13)$$

Esta é a fórmula computacionalmente útil.

A rotina para redução de Householder dada acima começa realmente na n -ésima coluna de $\mathbf{A}$, e não na primeira como mostrado acima. Em detalhes, as equações são como segue: No estágio m ($m = 1, 2, \dots, n-2$) o vetor $\mathbf{u}$ tem a forma

$$\mathbf{u}^T = [a_{i1}, a_{i2}, \dots, a_{i,i-2}, a_{i,i-1}, \pm\sqrt{\sigma}, 0, \dots, 0] \quad (14)$$

Aqui

$$i \equiv n - m + 1 = n, n-1, \dots, 3 \quad (15)$$

e a quantidade σ ($|x|^2$ na notação anterior) é

$$\sigma = (a_{i,1})^2 + \dots + (a_{i,i-1})^2 \quad (16)$$

O sinal de σ é escolhido em (14) para ser o mesmo de $a_{i,i-1}$ de modo a minimizar o erro de arredondamento.

As variáveis são então calculadas na seguinte ordem: σ , $\mathbf{u}$, H , $\mathbf{p}$, K , $\mathbf{q}$, $\mathbf{A}'$. Em qualquer estágio m , $\mathbf{A}$ é tridiagonal em suas $(m-1)$ linhas e colunas.

Se os autovetores da matriz tridiagonal final são encontrados, então os autovetores de $\mathbf{A}$ podem ser obtidos aplicando a transformação acumulada

$$\mathbf{Q} = \mathbf{P}_1 \cdot \mathbf{P}_2 \cdots \mathbf{P}_{n-2} \quad (17)$$

a aqueles autovetores. Assim, Q é formada por recursão depois que todos os P 's foram determinados:

$$Q_{n-2} = P_{n-2},$$

$$Q_j = P_j \cdot Q_{j+1}, j = n-3, \dots, 1, \quad (18)$$

$$Q = Q_1$$

APÊNDICE 4

Autovalores e autovetores de uma matriz tridiagonal.

Avaliação do polinômio característico.

Uma vez que a matriz real, simétrica original foi reduzida para a forma tridiagonal, um método possível para determinar seus autovalores é encontrar as raízes de seu polinômio característico $p_n(\lambda)$ diretamente. O polinômio característico de uma matriz tridiagonal pode ser avaliado para qualquer valor de teste de λ através de uma relação de recursão eficiente. Os polinômios de baixa ordem produzidos durante a recursão formam uma sequência de Sturmian que pode ser usada para localizar os autovalores em intervalos do eixo real. Um método de busca de raízes tais como os métodos de bissessão ou de Newton pode ser utilizado para refinar os intervalos. Os autovetores correspondentes podem ser então encontrados por iteração inversa.

Os algoritmos QR e QL.

A idéia básica por trás do algoritmo QR é que qualquer matriz real pode ser decomposta na forma

$$A = Q \cdot R \quad (1)$$

onde Q é ortogonal e R é triangular superior. Para uma matriz geral, a decomposição é construída aplicando-se transformações de Householder para eliminar colunas sucessivas de A abaixo da diagonal.

Seja a matriz formada pelos fatores da equação (1) na ordem inversa

$$A' = R \cdot Q \quad (2)$$

Desde que Q é ortogonal, a equação (1) fornece $R = Q^T \cdot A$. Então a equação (2) torna-se

$$A' = Q^T \cdot A \cdot Q \quad (3)$$

Pode-se ver que A' é uma transformação ortogonal de A .

Não existe nenhuma razão especial para que um dos fatores de A seja triangular superior; pode-se também fazê-lo triangular inferior. Este tipo de decomposição é chamado de algoritmo QL, desde que

$$A = Q \cdot L \quad (4)$$

onde L é triangular inferior.

É interessante lembrar que na redução de Householder para a forma tridiagonal (2), inicia-se na n -ésima (última) coluna da matriz original. Para minimizar o arredondamento, deve-se colocar os elementos maiores no canto inferior direito, se possível. Para diagonalizar a matriz tridiagonal resultante, o algoritmo QL irá ter um arredondamento menor que o QR.

O algoritmo QL consiste em uma série de transformações ortogonais:

$$A_s = Q_s \cdot L_s \quad (5)$$

$$A_{s+1} = L_s \cdot Q_s$$

O teorema a seguir é a base do algoritmo para uma matriz geral A :

i) Se A tem autovalores de valores absolutos $|\lambda_i|$, então $A_s \rightarrow$ [forma triangular inferior] a medida que $s \rightarrow \infty$. Os autovalores aparecem na diagonal em ordem crescente de magnitude absoluta.

ii) Se A tem um autovalor $|\lambda_i|$ de multiplicidade p , $A_s \rightarrow$ [forma triangular inferior] a medida que $s \rightarrow \infty$, exceto para uma matriz de bloco diagonal de ordem p , cujos autovalores $\rightarrow \lambda_i$.

A carga para o algoritmo QL é $O(n^3)$ por iteração para uma matriz geral, o que o torna proibitivo nestes casos. Entretanto a carga computacional é somente $O(n)$ por iteração para uma matriz tridiagonal e $O(n^2)$ para uma matriz de Hessenberg, o que o torna altamente eficiente nestas formas.

Nesta seção será dada atenção apenas ao caso em que A é uma matriz real, simétrica e tridiagonal. Todos os autovalores λ_i são então reais. De acordo com o teorema, se qualquer λ_i tem uma multiplicidade p , então deve existir pelo menos $(p - 1)$ zeros na sub e superdiagonal. Então a matriz pode ser dividida em submatrizes que podem ser diagonalizadas separadamente, e a complicação de blocos diagonais que possam aparecer no caso geral é irrelevante.

Na prova do teorema cotado acima, mostra-se que em geral um elemento na superdiagonal converge a zero por

$$a_{ij}^{(s)} \approx \left(\frac{\lambda_i}{\lambda_j} \right)^s \quad (6)$$

Embora $\lambda_i < \lambda_j$, a convergência pode ser lenta se λ_i for próximo de λ_j . A convergência pode ser acelerada pela técnica de *deslocamento*: Se k é uma constante qualquer, então $A - kI$ tem autovalores $\lambda_i - k$. Se for feita a decomposição

$$\mathbf{A}_s - k_s \mathbf{I} = \mathbf{Q}_s \cdot \mathbf{L}_s \quad (7)$$

de modo que

$$\mathbf{A}_{s+1} = \mathbf{L}_s \cdot \mathbf{Q}_s + k_s \mathbf{I} = \mathbf{Q}_s^T \cdot \mathbf{A}_s \cdot \mathbf{Q}_s \quad (8)$$

então a convergência é determinada pela razão

$$\frac{\lambda_i - k_s}{\lambda_j - k_s} \quad (9)$$

A idéia é escolher o deslocamento k_s em cada estágio para maximizar a taxa de convergência. Uma boa escolha para o deslocamento inicial poderia ser k_s próximo de λ_1 , o menor autovalor. Então a primeira linha dos elementos fora da diagonal poderia tender rapidamente para zero. Entretanto λ_1 não é conhecida *a priori*. Uma estratégia bastante eficaz na prática (embora não exista nenhuma prova de que seja ótima) é calcular os autovalores da submatriz diagonal principal 2×2 de $\mathbf{A}$. Então escolhe-se k_s igual ao autovalor próximo de a_{11} .

De uma maneira geral, supondo que foram encontrados $(r-1)$ autovalores de $\mathbf{A}$. Então pode-se *esvaziar* a matriz riscando as primeiras $(r-1)$ linhas e colunas deixando

$$\mathbf{A} = \begin{bmatrix} 0 & \dots & \dots & & 0 \\ & \dots & & & \\ & & 0 & & \\ \dots & & d_r & e_r & \dots \\ \dots & & e_r & d_{r+1} & \\ & & & \dots & 0 \\ & & & d_{n-1} & e_{n-1} \\ 0 & \dots & 0 & e_{n-1} & d_n \end{bmatrix} \quad (10)$$

Escolhe-se k_s igual ao autovalor da submatriz principal 2×2 que está mais próximo de d_r . Pode-se mostrar que a convergência do algoritmo com esta estratégia é geralmente cúbica (e no pior caso quadrática para autovalores degenerados). Esta convergência rápida é que torna o algoritmo tão atrativo.

Pode-se notar que com o deslocamento, os autovalores não aparecem na diagonal principal necessariamente em ordem crescente de valor absoluto.

Como mencionado anteriormente, a decomposição QL de uma matriz geral é efetuada por uma sequência de transformações de Householder. Para uma matriz tridiagonal, entretanto, é mais eficiente utilizar rotações planas P_{pq} . Usa-se a sequência $P_{12}, P_{23}, \dots, P_{n-1,n}$ para eliminar os elementos $a_{12}, a_{23}, \dots, a_{n-1,n}$. Por simetria, os elementos subdiagonais $a_{21}, a_{32}, \dots, a_{n,n-1}$ serão eliminados também. Então cada Q_s é um produto de rotações de planos:

$$Q_s^T = P_1^{(s)} \cdot P_2^{(s)} \cdot \dots \cdot P_{n-1}^{(s)} \quad (11)$$

onde P_i elimina $a_{i,i+1}$. É importante notar que é Q^T e não Q em (11) pois definiu-se

$$L = Q^T \cdot A.$$

Algoritmo QL com deslocamentos implícitos.

O algoritmo descrito pode ser bastante bem sucedido. Entretanto quando os elementos de A diferem grandemente em ordem de magnitude, a subtração de um k_s grande dos elementos da diagonal pode levar a uma perda de precisão para os autovalores menores. esta dificuldade pode ser evitada pelo algoritmo QL com *deslocamentos implícitos*. O algoritmo QL implícito é matematicamente equivalente ao algoritmo QL original, mas o processamento não requer que $k_s I$ seja realmente subtraído de A .

O algoritmo é baseado no seguinte lema: Se A é uma matriz simétrica não singular e $B = Q^T \cdot A$, onde Q é ortogonal e B é tridiagonal com elementos fora da diagonal positivos, então Q e B são inteiramente determinados quando a última linha de Q^T é especificada.

Prova: Seja q_i^T denotando o i -ésimo vetor linha da matriz Q^T . Então q_i é o i -ésimo vetor linha da matriz Q^T . Então q_i é o i -ésimo vetor coluna da matriz Q . A relação $B \cdot Q^T = Q^T \cdot A$ pode ser escrita como

$$\begin{bmatrix} \beta_1 & \gamma_1 & & & \\ \alpha_2 & \beta_2 & \gamma_2 & & \\ & & \dots & & \\ & & & \alpha_{n-1} & \beta_{n-1} & \gamma_{n-1} \\ & & & & \alpha_n & \beta_n \end{bmatrix} \cdot \begin{bmatrix} q_1^T \\ q_2^T \\ \dots \\ q_{n-1}^T \\ q_n^T \end{bmatrix} = \begin{bmatrix} q_1^T \\ q_2^T \\ \dots \\ q_{n-1}^T \\ q_n^T \end{bmatrix} \cdot A \quad (12)$$

A n -ésima linha desta equação matricial é

$$\alpha_n q_{n-1}^T + \beta_n q_n^T = q_n^T \cdot A \quad (13)$$

Desde que Q é ortogonal,

$$q_n^T \cdot q_m = \delta_{nm} \quad (14)$$

Então se a equação (13) for pós-multiplicada por q_n

$$\beta_n = q_n^T \cdot A \cdot q_n \quad (15)$$

o qual é conhecido desde que q_n é conhecido. Então a equação (13) fornece

$$\alpha_n q_{n-1}^T = z_{n-1}^T \quad (16)$$

onde

$$\mathbf{z}_{n-1}^T \equiv \mathbf{q}_n^T \cdot \mathbf{A} - \beta_n \mathbf{q}_n^T \quad (17)$$

é conhecido. Portanto

$$\alpha_n^2 = \mathbf{z}_{n-1}^T \mathbf{z}_{n-1} \quad (18)$$

ou

$$\alpha_n = |\mathbf{z}_{n-1}| \quad (19)$$

e

$$\mathbf{q}_{n-1}^T = \frac{\mathbf{z}_{n-1}^T}{\alpha_n} \quad (20)$$

(onde α_n é não nulo por hipótese). Similarmente, pode-se mostrar por indução que se forem conhecidos $\mathbf{q}_n, \mathbf{q}_{n-1}, \dots, \mathbf{q}_{n-j}$ e os α 's, β 's e γ 's até o nível $(n - j)$, pode-se determinar as quantidades no nível $n-(j+1)$.

Para aplicar o lema na prática, supõe-se inicialmente que é possível encontrar uma matriz tridiagonal $\bar{\mathbf{A}}_{s+1}$ tal que

$$\bar{\mathbf{A}}_{s+1} = \bar{\mathbf{Q}}_s^T \cdot \bar{\mathbf{A}}_s \cdot \bar{\mathbf{Q}}_s \quad (21)$$

onde $\bar{\mathbf{Q}}_s^T$ é ortogonal e tem a mesma última linha que $\mathbf{Q}_s^T$ no algoritmo QL original. Então

$$\bar{\mathbf{Q}}_s = \mathbf{Q}_s \text{ e } \bar{\mathbf{A}}_{s+1} = \mathbf{A}_{s+1}.$$

Agora, no algoritmo original, da equação (11), pode-se ver que a última linha de $\mathbf{Q}_s^T$ é a mesma da última linha de $\mathbf{P}_{n-1}^{(s)}$. É conveniente lembrar que $\mathbf{P}_{n-1}^{(s)}$ é uma rotação plana para eliminar o elemento $(n-1, n)$ de $\mathbf{A}_s - k_s \mathbf{I}$. Um cálculo simples usando (11) mostra que seus parâmetros

$$c = \frac{d_n - k_s}{\sqrt{e_n^2 + (d_n - k_s)^2}}, \quad s = \frac{-e_{n-1}}{\sqrt{e_n^2 + (d_n - k_s)^2}} \quad (22)$$

A matriz $\mathbf{P}_{n-1}^{(s)} \cdot \mathbf{A}_s \cdot \mathbf{P}_{n-1}^{(s)T}$ é tridiagonal com 2 elementos extras:

$$\begin{bmatrix} \dots & & & & \\ & \times & \times & \times & \\ & & \times & \times & \times & \mathbf{x} \\ & & & \times & \times & \times \\ & & & & \mathbf{x} & \times & \times \end{bmatrix} \quad (23)$$

É necessário então reduzir esta matriz para a forma tridiagonal com uma matriz ortogonal cuja última linha é $[0, 0, \dots, 0, 1]$ de modo que a última linha de $\overline{\mathbf{Q}}_s^T$ fique igual a $\mathbf{P}_{n-1}^{(s)}$. Isto pode ser feito por uma sequência de transformações de Givens* ou Householder. Para a forma especial da matriz (23), o método de Givens é melhor. Faz-se uma rotação no plano $(n-2, n-1)$ de modo a eliminar o elemento $(n-2, n)$. [Por simetria, o elemento $(n, n-2)$ irá ser zerado também.] Isto fornece uma forma tridiagonal exceto pelos elementos extra $(n-3, n-1)$ e $(n-1, n-3)$. Estes são eliminados com uma rotação no plano $(n-3, n-2)$, e assim por diante. Então é necessária uma sequência de $(n-2)$ rotações de Givens. O resultado é tal que

$$\mathbf{Q}_s^T = \overline{\mathbf{Q}}_s^T = \overline{\mathbf{P}}_1^{(s)} \cdot \overline{\mathbf{P}}_2^{(s)} \cdots \overline{\mathbf{P}}_{n-2}^{(s)} \cdot \overline{\mathbf{P}}_{n-1}^{(s)} \quad (24)$$

* O método de Givens é descrito em Press (1987)

onde os $\bar{\mathbf{P}}$'s são as rotações de Givens e $\mathbf{P}_{n-1}$ são as mesmas rotações planas do algoritmo original. Então a equação (21) fornece a próxima iteração de $\mathbf{A}$. É interessante notar que o deslocamento k_s entra implicitamente através dos parâmetros (22).