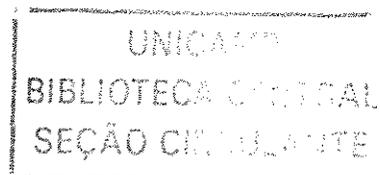


UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE ARTES

**Classificação semi-automática do repertório vocal  
do Sabiá-laranjeira *Turdus rufiventris***

Nilson Evilásio de Souza Filho

CAMPINAS – 2004



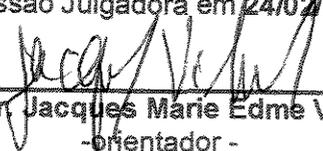
UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE ARTES

Mestrado em Multimeios

**Classificação semi-automática do repertório vocal  
do Sabiá-laranjeira *Turdus rufiventris***

Nilson Evilásio de Souza Filho

Este exemplar é a redação final da  
Dissertação defendida pelo Sr. Nilson  
Evilásio de Souza Filho e aprovado pela  
Comissão Julgadora em 24/02/2005.

  
Prof. Dr. Jacques Marie Edme Vieliard  
-orientador -

Dissertação apresentada ao curso de  
Mestrado em Multimeios do Instituto de  
Artes da UNICAMP como requisito  
parcial para obtenção do grau de Mestre em  
Multimeios e Ciências sob a orientação do  
Prof. Dr. Jacques Marie Edme Vieliard.

CAMPINAS – 2004

UNIDADE BC  
Nº CHAMADA 7/UNICAMP  
So89c

V \_\_\_\_\_ EX \_\_\_\_\_  
TOMBO BC/ 65314  
PROC 16-86-05  
C \_\_\_\_\_ D \*  
PREÇO 14,00  
DATA 17-8-05  
Vº CPD \_\_\_\_\_

Bibid 361409

FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO IA. - UNICAMP

So89c

Souza Filho, Nilson Evilásio de. *haver mudança no título*

Classificação e análise semi-automática do repertório vocal do Sabiá-laranjeira Turdus rufiventris. / Nilson Evilásio de Souza Filho. – Campinas, SP: [s.n.], 2004.

Orientador: Jacques Marie Edme Vielliard.

Dissertação(mestrado) – Universidade Estadual de Campinas. Instituto de Artes.

- informação*
1. Teoria da comunicação. 2. Análise de Fourier.
  3. Processamento de imagem. I. Vielliard, Jacques Marie Edme. II. Universidade Estadual de Campinas. Instituto de Artes. III. Título.

CAMPINAS – 2004

### Agradecimentos

Agradeço ao Programa de Pós-Graduação em Multimeios do Instituto de Artes da Universidade Estadual de Campinas e o atual Coordenador do Programa Prof. Dr. Fernando Passos, a Prof. Dra. Maria Luisa as Silva, ao meu Orientador Prof. Dr. Jacques Vielliard e o Laboratório de Bioacústica da UNICAMP pela oportunidade.

Ao Prof. Dr. Marcos César Danhoni Neves do Dep. De Física da Universidade Estadual de Maringá e ao meu amigo Maycol Alencar pela ajuda com o Matlab.

E especialmente à minha família pelo apoio nesses longos anos.

200517646

Dedico ao meu sábio avô  
Abelegy Alves

Hören Sie denn nicht das entsetzlich Schrein  
ringsum das man gewöhnlich die Stille heißt ?

Prólogo do Filme *O Enigma de Kaspar Hauser* de  
Werner Herzog, 1974

## RESUMO

Baseando-se em estudos sobre o repertório vocal da espécie *Turdus rufiventris* (Aves, Passeriformes, Turdinae), o objetivo nesta pesquisa é elaborar um método para classificação semi-automática de notas (unidades contínuas de som) baseado em cálculos de correlação aplicados a imagens, cuja técnica é conhecida como “Template Matching”. Como se trata de imagens espectrográficas, o método foi chamado de SITM (“Spectrographic Image Template Matching”). Essa técnica de comparação de imagens está sendo incorporada como uma forma compacta de análise bioacústica, facilitando o cálculo de parâmetros relacionados à teoria da comunicação. A avaliação de desempenho do SITM foi realizada com a utilização de gravações do canto de três indivíduos da espécie *T. rufiventris* obtidas na natureza, e depositadas no Arquivo Sonoro Neotropical (ASN) da Universidade Estadual de Campinas – UNICAMP. Os resultados demonstraram que o SITM reconhece corretamente, com uma confiança de 96%, de 77% a 97% das notas do indivíduo 08, de 64% a 82% das notas do indivíduo 15 e de 74% a 97% das notas do indivíduo 26. Esses resultados apresentaram intervalos de confiança significativos, entretanto não o suficiente para que o SITM possa ser utilizado em laboratório. Como os erros ocorridos no processo da pesquisa podem ser corrigidos no futuro, o SITM pode se tornar um método muito eficaz. Otimizando o algoritmo de classificação e utilizando-se novos recursos é possível criar um *software* científico moderno na área de bioacústica.

## SUMMARY

Based on studies of vocal repertoire of the Rufous-bellied Thrush *Turdus rufiventris* (Aves, Passeriformes, Turdinae), the objective in this research is to elaborate a method for semi-automatic classification of the songs notes (continuous units of sound). This research is based on calculations of correlation applied to images; through the technique know as Template Matching. As it deals with spectrographic images, the method was called SITM (Spectrographic Image Template Matching). This technique of comparison of images is begging incorporated into a compact form of bioacoustic analysis, facilitating the calculation of parameters related to the theory of communication. The evaluation of performance of SITM was carried out using songs of three individuals of *T. rufiventris*, recorded in the wild and deposited in the Neotropical Sound Archive (ASN) of the State University of Campinas (UNICAMP), Brazil. The results had demonstrated that SITM recognizes correctly, with a 96% confidence, 77 to 97% of the notes of individual 08, 64% to 82% of the notes of individual 15 and 74 to 97% of the notes of individual 26. These results are significantly reliable, but not enough to be used in laboratory. However, the errors that occurred in this process can be corrected in the future, when SITM can become a very efficient method. Optimizing the classification algorithm and using new tools, it will be possible to create modern scientific software in the biocoustic area.

# ÍNDICE

<b>INTRODUÇÃO</b> .....	01
<b>OBJETIVOS</b> .....	05
<b>METODOLOGIA</b> .....	06
Pré-processamento .....	06
Reconhecimento e classificação .....	07
A inferência estatística .....	09
O <i>software</i> Matlab.....	10
<b>CONCEITOS E APLICAÇÕES DA TEORIA DA COMUNICAÇÃO</b>	
<b>Sinais e sistemas de comunicação</b> .....	12
O sinal sonoro.....	15
<b>Teoria de Fourier</b> .....	19
Representações dependentes do sinais.....	20
A Transformada Rápida de Fourier (FFT).....	21
Análise de tempo-frequência.....	21
Espectrograma.....	23
Tipos de janelas.....	24
<b>Processamento digital de imagens (PDI)</b> .....	28
Conceito de imagem digital.....	29
Morfologia matemática.....	30
Erosão e Dilatação.....	31
<b>Reconhecimento de padrões</b> .....	33
Reconhecimento de padrões auditivos.....	34
Correlação.....	36
Template Matching.....	38
<b>Teoria da comunicação</b> .....	41
Modelo de comunicação de Shannon.....	44
<b>Bioacústica e comunicação sonora</b> .....	47
O Sabiá-laranjeira <i>Turdus rufiventris</i> .....	49
<b>RESULTADOS</b> .....	50
<b>DISCUSSÃO</b> .....	55
<b>CONCLUSÃO</b> .....	56
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	57
<b>APÊNCICE – Pseudocódigo Matlab do SITM</b> .....	60

## Lista de Tabelas

<i>Tabela 1 – Principais Fenômenos Correspondentes do Som.</i>	16
<i>Tabela 2 – Relação entre propriedades de um sinal e a representação de Fourier apropriada</i>	21

## Lista de Figuras

<i>Figura 1 – Espectrograma obtido através do Matlab.</i>	7
<i>Figura 2 – Ilustração da técnica SITM verificando a similaridade entre a 2ª e a 11ª nota (NCC = 76,4%, nota B) da primeira frase de um canto.</i>	8
<i>Figura 3 – Sinal sonoro de voz correspondente à frase “Boa Tarde”.</i>	18
<i>Figura 4 – (a) um sinal <math>x_{(t)}</math> visualizado através de uma janela temporal de extensão limitada. (b) Uma janela retangular e uma janela de Hanning. (c) a STFT de <math>x_{(t)}</math> linear ao sinal.</i>	22
<i>Figura 5 – O espectrograma do sinal sonoro da Figura 3.</i>	23
<i>Figura 6 – (a) Janela retangular; (b) Janela de Bartlett; (c) Janela de Hamming; (d) Janela Blackmann-Harris; todas com <math>N= 35</math>.</i>	27
<i>Figura 7 – (a) Uma simples imagem digital de duas bandas e (b) o correspondente arranjo de números.</i>	29
<i>Figura 8 – (a) Representação gráfica tridimensional da função sinc e (b) imagem correspondente em nível de cinza. (Figura de difração por um orifício retangular)</i>	29
<i>Figura 9 – Efeitos da erosão e dilatação binárias usando um elemento estruturante esférico de raio R numa imagem X.</i>	31
<i>Figura 10 – (a) Imagem original, (b) imagem erodida e (c) imagem dilatada.</i>	32
<i>Figura 11 – (a) Uma função qualquer <math>g_{(t)}</math>. (b) Outra função qualquer <math>h_{(t)}</math>. (c) A correlação entre as duas funções, <math>g_{(t)}</math> e <math>h_{(t)}</math>, limitadas no tempo.</i>	37
<i>Figura 12 – (a) Uma imagem em que o molde (letra “M”) foi correlatado. (b) Localização e ocorrência do molde.</i>	38
<i>Figura 13 – Arranjo para obter a correlação de <math>f_{(x,y)}</math> e <math>h_{(x,y)}</math> no ponto <math>(s, t)</math>.</i>	39
<i>Figura 14 – Modelo de Comunicação de Shannon e Weaver.</i>	44
<i>Figura 15 – O Sabiá-laranjeira e o espectrograma de um canto típico da espécie.</i>	49
<i>Figura 16 – Frases do indivíduo 08.</i>	52

## Abreviaturas

ASN – Arquivo Sonoro Neotropical

DTFS – Discrete Time Fourier Series

DTFT – Discrete Time Fourier Transform

FFT – Fast Fourier Transform

FS – Fourier Series

FT – Fourier Transform

IC – Intervalo de Confiança

IFFT – Inverse Fast Fourier Transform

NCC – Normalized Cross-correlation

PDI – Processamento Digital de Sinais

SPCC – Spectrographic Cross-correlation

SITM – Spectrographic Image Template Matching

STFT – Short Time Fourier Transform

## INTRODUÇÃO

A importância das onomatopéias nas línguas indígenas (BERLIN e O'NEILL 1981) é um forte indicativo de que o interesse do homem pelos cantos e gritos de animais é antigo. Atualmente o som emitido pelos animais e, em particular, pelos pássaros tem sido uma das especialidades da Bioacústica. Esse ramo da ciência, que estuda os diversos aspectos da comunicação sonora, beneficiando-se da tecnologia de gravações e análise dos sons de forma a conservar e definir em termos de parâmetros físicos o próprio sinal de comunicação, tem fornecido uma ajuda importante para a ornitologia de campo. Graças a ela foi possível entender vários aspectos da comunicação sonora em aves (VIELLIARD 1987).

Como campo científico moderno, a Bioacústica desenvolveu-se de maneira relativamente recente na Ornitologia, uma das tradicionais áreas do conhecimento, com contribuições fundamentais no estabelecimento dos conceitos básicos da evolução, da biogeografia, da taxonomia, da etologia e da biologia da conservação. Como todo domínio do conhecimento, a Ornitologia depende, para se desenvolver, de dois requisitos: uma base conceitual coerente e uma ferramenta tecnicamente adequada (VIELLIARD 2000).

A proposta desta pesquisa é contribuir para esse campo científico, focalizando um assunto que vem sendo objeto de interesse de alguns pesquisadores de renome internacional. Baseando-se nesses estudos sobre o repertório vocal da espécie *Turdus rufiventris* (Aves Passeriformes, Turdinae), o objetivo é elaborar uma forma de classificação semi-automática das notas (unidades contínuas de som) através de medidas de similaridade de suas imagens espectrográficas. Essa técnica de comparação de imagens está sendo incorporada como uma forma compacta de análise bioacústica, facilitando o cálculo de parâmetros relacionados à teoria da comunicação.

Uma das etapas mais importantes da pesquisa bioacústica é o trabalho de campo. Este consiste basicamente na localização da espécie e em gravações e observações relacionadas ao seu comportamento na natureza. Comparar e classificar sons animais são tarefas comuns

em bioacústica. Cientistas executam comparações detalhadas de sonogramas e procuram relações entre a estrutura do som no plano tempo-frequência e uma infinidade de variáveis de contexto extrínseco. O objetivo pode ser identificar relações entre espécies de um mesmo *habitat* e entre um *habitat* e outro, analisar o contexto social em que os sons são emitidos, indagar se determinados cantos gravados são de um mesmo indivíduo ou grupo social e se indivíduos ou populações de um mesmo espaço geográfico são estruturalmente similares a outros indivíduos, grupos ou populações. O interesse pode ser também identificar as propriedades estruturais isoladas que os animais manifestam em seus comportamentos e/ou classificar perceptualmente os sons recebidos, relatando estruturas de sinal e mecanismos utilizados na emissão de um som específico.

Um ponto crucial nesse tipo de estudo é o método empregado na comparação e categorização estrutural de um conjunto de sons ou sinais. A comparação visual de sonogramas por um observador humano treinado é um método amplamente usado para classificar espectrogramas em grupos. Embora o critério visual raramente seja especificado, o fato de múltiplos observadores convergirem para esquemas similares de classificação sugere que a similaridade estrutural majoritária do canto é identificada.

Os critérios mais freqüentes para selecionar as medições são facilidade e rapidez da medida, significância em estudos prioritários ou relevância para novas conjecturas. Assim, Clark et al. (1987) desenvolveram um método de comparação de sons denominado SPCC (*Spectrographic Cross-correlation*), no qual os espectrogramas de dois sons são correlatados e o valor do pico da correlação é considerado como uma medida da similaridade sonora.

Propõe-se neste trabalho um método para classificação de notas baseado em cálculos de correlação aplicados a imagens. Essa técnica é conhecida como *Template Matching (TM)*. Como se trata de imagens espectrográficas, o método foi chamado de SITM (*Spectrographic Image Template Matching*). A proposta, ao criar o SITM, é utilizar no processo de classificação todas as características estruturais discerníveis de dois

espectrogramas, em vez de comparações limitadas de um predeterminado, e possivelmente incompleto, conjunto de observações.

Como já se afirmou, uma das etapas mais importantes da pesquisa bioacústica é o trabalho de campo. No que diz respeito ao objeto desta pesquisa, o trabalho de campo foi realizado ao longo de vários anos por diversos pesquisadores. Em 1997, Maria Luisa da Silva apresentou ao programa de mestrado em Neurociências e Comportamento do Instituto de Psicologia da USP a “Descrição do repertório vocal do Sabiá-laranjeira *Turdus rufiventris* (Aves Passeriformes, Turdinae)”. Posteriormente, considerando que um conhecimento mais amplo na comunicação sonora animal pode ser obtido através do entendimento da estrutura e organização dos seus sinais, a autora analisou o mesmo caso da complexa comunicação do *T. rufiventris*, utilizando a Teoria da Comunicação proposta por Shannon (SILVA 2001).

Os resultados gerados nos dois trabalhos de Maria Luisa da Silva motivaram o desenvolvimento do SITM e são eles que possibilitam a avaliação do método por consistirem num rico material para comparação de dados. Embora não seja a principal meta, por ser uma consequência direta do reconhecimento e do algoritmo de classificação, a entropia de Shannon foi incluída no SITM. A idéia básica é desenvolver um método compacto que utilize o reconhecimento de padrões de imagens para classificação das notas e que a estrutura e organização da emissão das notas seja pré-determinada segundo a Entropia de Shannon, a qual é calculada conforme a seqüência de notas de uma frase vai sendo reconhecida.

A classificação semi-automática de notas prevê uma boa precisão. O processamento simultâneo de um grande número de dados gera resultados que jamais um observador humano poderia atingir com tanta rapidez. Resumidamente, a pretensão do método é facilitar esse tipo de análise e possibilitar que o mesmo seja aplicado no estudo de outras espécies, de forma a se atingir resultados num período de tempo mais curto.

Por ser um estudo de âmbito inter/multidisciplinar, foi feita uma breve revisão bibliográfica dos principais assuntos que embasam a pesquisa, como Sinais e Sistemas de Comunicação, Teoria de Fourier, Processamento Digital de Imagens, Reconhecimento de Padrões, Teoria da Comunicação, Comunicação Sonora e Bioacústica.

A seqüência lógica dos capítulos segue a idéia de que um canto, como um sinal de comunicação, é melhor visualizado num espectrograma, o qual é uma imagem criada a partir de conceitos da Teoria de Fourier para fornecer detalhadamente informações sobre esse sinal. Por isso, são utilizadas técnicas de processamento digital de imagens para o reconhecimento de padrões de sinais sonoros, para que os mesmos possam ser classificados de forma que possibilitem uma estimativa da organização da comunicação sonora da espécie *T. rufiventris*. Também foram incluídas algumas informações sobre o *software* Matlab<sup>®</sup>, a ferramenta indispensável que aproximou a biologia, a matemática, a física e a engenharia com a arte da programação. A metodologia é explicada conforme a rotina foi desenvolvida e os resultados são discutidos em termos de intervalos de confiança da inferência estatística.

A seqüência dos capítulos poderia assumir várias formas, mas a escolhida foi aquela que demonstrou ser a mais adequada para um leitor iniciante. A intenção é que, embora muito técnico, ele seja direto, claro, útil e, às vezes, até agradável. Embora sintética, a revisão provavelmente servirá como um bom material bibliográfico, não só para a área de comunicação sonora e bioacústica, mas também para a aplicação dos métodos sugeridos e utilizados em outras áreas.

## OBJETIVOS

A partir de estudos já existentes sobre o repertório vocal da espécie *Turdus rufiventris*, pretende-se elaborar um método computacional com uma forma compacta de análise bioacústica, estabelecendo um critério de comparação de notas através de comparações de imagens espectrográficas. Também faz parte do método utilizar esse critério para a classificação semi-automática dos sons que compõem um canto e facilitar a estimativa de parâmetros relacionados à teoria da comunicação.

## METODOLOGIA

O procedimento completo da pesquisa pode ser dividido em três etapas principais: a *Pesquisa de Campo*, o *Pré-processamento* e o *Reconhecimento* e, em consequência, a *Classificação*, a qual, por sua vez, possibilita os cálculos da *Entropia de Shannon*.

A amostra utilizada no estudo da comunicação sonora de *T. rufiventris* é representada por gravações de qualidade profissional, a maioria produzida em Nagra E com microfone ultradirecional Senheiser MKH 816, outras em DAT (SILVA, 1997). As gravações, executadas na natureza, são todas devidamente depositadas no Arquivo Sonoro Neotropical (ASN) da Universidade Estadual de Campinas – UNICAMP. O código internacional de referência contém as iniciais do coletor, o número referente à fita em que foi feita a gravação e o número do “cut” em ordem cronológica (KETTLE e VIELLIARD, 1991).

O ASN possui um total de 75 gravações de cantos de *T. rufiventris*. Foram utilizadas gravações de três indivíduos para avaliar o desempenho do SITM. Os cantos dos indivíduos 08, 15 e 26 foram escolhidos devido à boa qualidade técnica das gravações e às características específicas do canto, como a quantidade e variedade de notas nas frases. O desempenho do programa foi avaliado segundo intervalos de confiança da inferência estatística.

### **Pré-processamento**

Os sons escolhidos foram equalizados, digitalizados, editados e armazenados em um CD. O *software* Matlab é capaz de ler e tocar arquivos com extensão *.wav* mono, com 44,1 kHz, 16 bit e formato de compactação PCM. O *script* foi desenvolvido para ler um som de cada vez e mostrar o respectivo espectrograma utilizando uma janela  $w(t)$  de Hann e FFT de tamanhos iguais a de 1024 bandas.

## Reconhecimento e classificação

A seguir é exemplificada a aplicação do SITM a uma frase do Sabiá-laranjeira. Ao dar entrada com um arquivo sonoro, primeiramente o programa verifica se há uma base de dados, depois é iniciada a análise. Surgirá o espectrograma, como ilustra a figura.1.

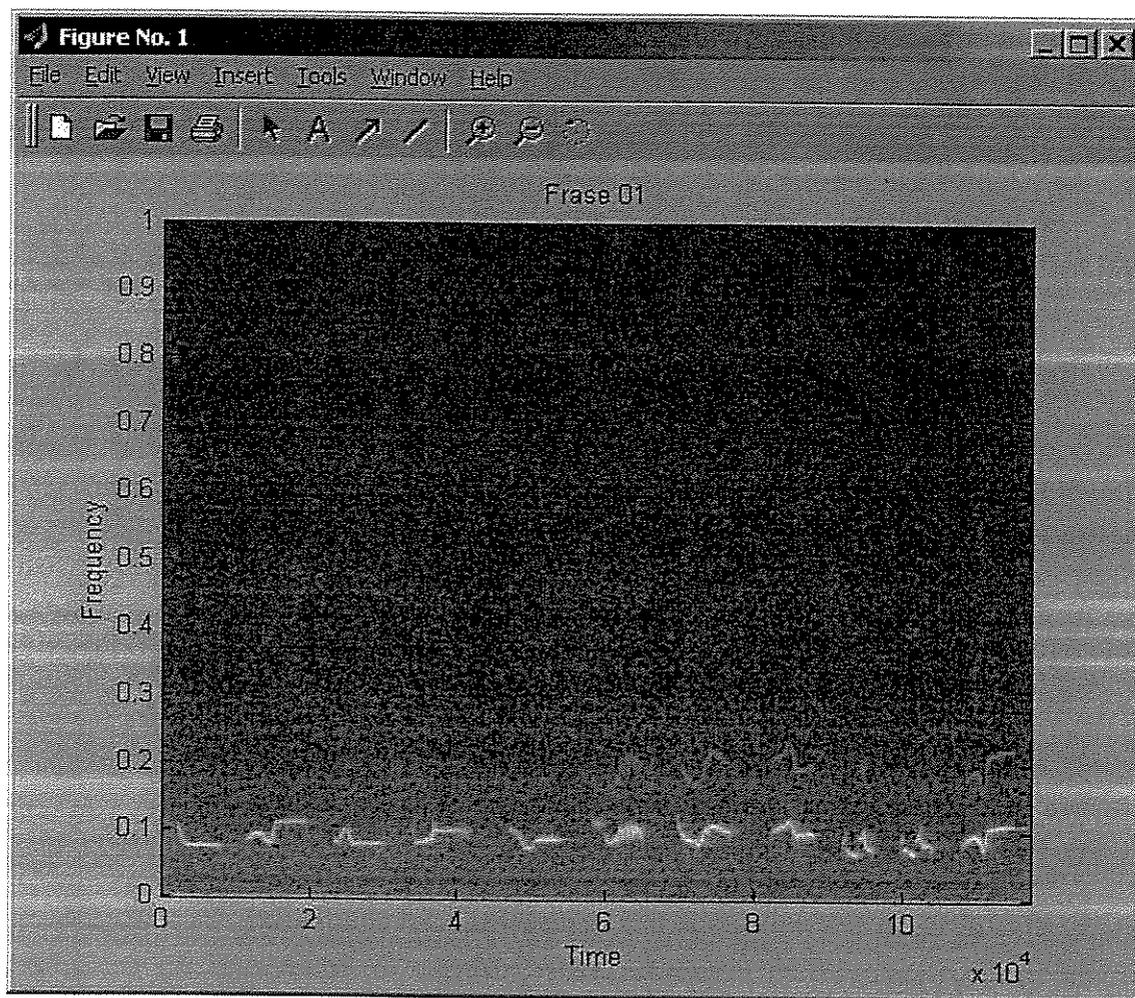


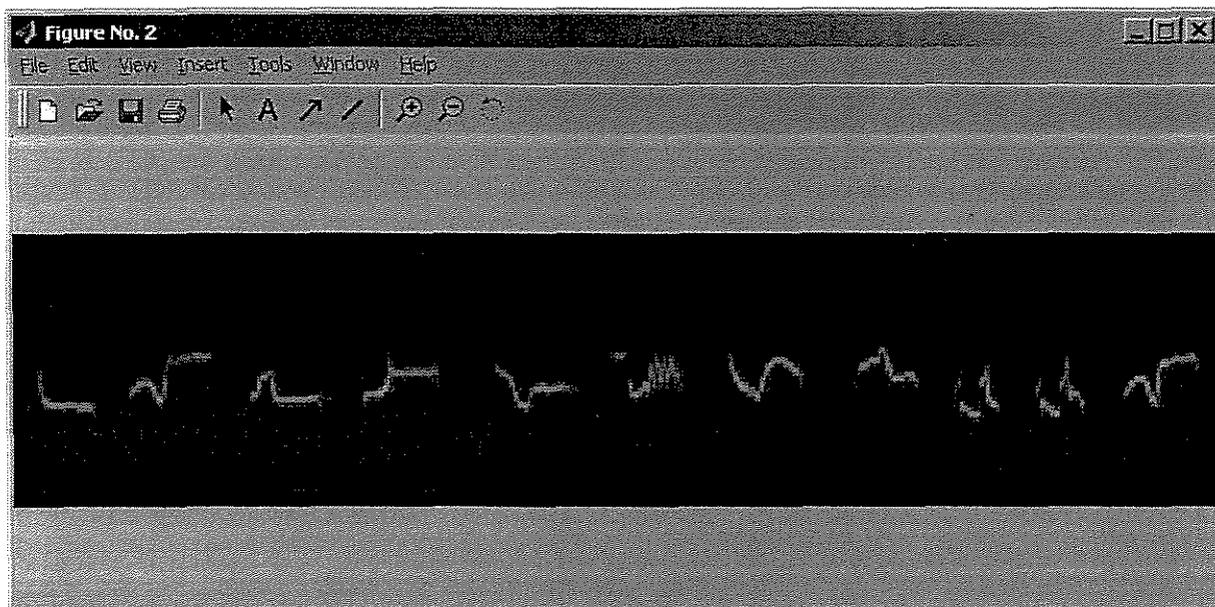
Figura 1 – Espectrograma obtido através do Matlab.

Em seguida, a parte mais importante do espectrograma, a faixa de frequência que contém todas as notas da frase, deve ser selecionada com o *mouse* para que seja

transformada numa imagem de extensão .jpg em escala de cinza, passando por um rápido tratamento gráfico no qual é utilizado um elemento estruturante esférico.

É importante ressaltar que tanto o espectrograma quanto o tratamento gráfico das imagens podem sofrer suas devidas alterações, antes de iniciar uma análise, conforme a qualidade do arquivo de som.

Ao inserir o número de notas que compõe a frase, como pede na continuação do *Command Window*, ou no *prompt* do Matlab, é necessário seleccionar nota por nota, novamente com o auxílio do *mouse*.



**Figura 2 – Ilustração da técnica SITM verificando a similaridade entre a 2ª e a 11ª nota (NCC = 76,4%, nota B) da primeira frase de um canto.**

O critério adotado para a classificação de notas é baseado no grau de correlação (*NCC* em %). Para que duas notas sejam consideradas as mesmas, o grau de correlação deve ser maior ou igual a 76%.

Observe-se, na figura 2, que a segunda e a décima primeira nota apresentam pequenas diferenças estruturais. Entretanto, auditivamente elas são idênticas. Essa foi a principal característica considerada nos testes para determinar o valor mínimo do grau de correlação (de 76%) entre notas, o valor mínimo para que duas notas sejam consideradas como sendo a mesma.

A primeira nota que aparece no repertório é classificada como nota **A**. Se a segunda nota tiver  $NCC < 76\%$  em relação à primeira (nota **A**) ela é classificada como nota **B** e assim por diante vai-se nomeando cada nota diferente com letras do alfabeto e construindo uma base de dados composta pelas diferentes notas. Conforme as notas vão sendo classificadas, as frequências relativas de cada uma dentro de cada frase é computada; assim, a Entropia de Shannon é facilmente determinada.

Após a seleção de todas as notas de uma frase, o retorno dado pelo programa é a seqüência das notas (frase\_atual) conforme elas foram nomeadas, as diferentes notas que compõem as frases (base\_do\_canto) e a Entropia de Shannon da atual frase.

### **Inferência estatística**

Usualmente, é impraticável observar toda a população, seja pelo custo caríssimo seja por dificuldades diversas. Examina-se então uma amostra. Se essa amostra for bastante significativa, os resultados obtidos poderão ser generalizados para toda a população.

Toda conclusão tirada por uma amostragem, quando generalizada para a população, virá acompanhada de um grau de incerteza ou risco. Toda população possui um parâmetro  $\theta$ , então, a partir de uma amostra, pode-se obter uma estimação do parâmetro  $\theta$ . Uma maneira de expressar a precisão da estimação é estabelecer limites, que com certa probabilidade incluem o verdadeiro valor do parâmetro da população.

Esses limites são chamados “limites de confiança”: determinam um Intervalo de Confiança, no qual deverá estar o verdadeiro valor do parâmetro. A estimação por intervalos consiste na fixação de dois valores tais que  $(1-\alpha)$  seja a probabilidade de que o intervalo contenha o verdadeiro valor do parâmetro, sendo  $\alpha$  o nível de incerteza ou grau de desconfiança e  $(1-\alpha)$  o coeficiente de confiança ou nível de confiabilidade (MORETTIN 1995).

### **O software Matlab**

O Matlab é uma linguagem de programação de alto-desempenho para computação técnica, integrando computação, visualização e programação num ambiente amigável, na qual os problemas e respectivas soluções são formulados recorrendo à familiar notação matemática. Sua primeira versão, escrita na Universidade do Novo México e na Universidade de Stanford, no final da década de 1970, destinava-se a cursos de teoria matricial, álgebra linear e análise numérica, nos quais os estudantes podiam utilizar pacotes de sub-rotinas em FORTRAN para a manipulação de matrizes, sem a necessidade de reescrevê-las (HANSELMAN e LITTLEFIELD 1999).

O nome Matlab deriva de *Matrix Laboratory*. Em ambientes universitários, este software representa a ferramenta padrão na aprendizagem de Matemática, Engenharia e Ciência em geral. Na indústria, é a ferramenta mais empregada nas atividades de I&D de linha avançada e análise de dados.

Hoje, sua capacidade estende-se muito além do “Laboratório de Matrizes” original. O Matlab tornou-se um sistema interativo e uma linguagem de programação para computação técnica e científica em geral, integrando a capacidade de fazer cálculos, a visualização gráfica e a fácil programação. Os usos típicos incluem:

- Cálculos matemáticos;
- Desenvolvimento de algoritmos;
- Modelagem, simulação e confecção de protótipos;
- Análise, exploração e visualização de dados;
- Gráficos científicos e de engenharia;
- Desenvolvimento de Aplicações, incluindo GUI's (*Graphical User Interface*).

O Matlab é um sistema interativo cujo elemento de dados básico é uma matriz que não requer dimensionamento. Isso permite solucionar muitos problemas computacionais, principalmente os que envolvem formulações matriciais ou vetoriais, em uma fração de tempo que seria necessário para escrever um programa em uma linguagem C ou FORTRAN.

É tanto um ambiente quanto uma linguagem de programação e um de seus aspectos mais poderosos é o fato de sua linguagem permitir construir suas próprias ferramentas reutilizáveis.

Pode-se facilmente criar funções e programas especiais próprios (conhecidos como arquivos M). À medida que se escrevem mais e mais funções relacionadas entre si em diretórios especiais para lidar com certos problemas, naturalmente agrupam-se, por conveniência, funções relacionadas entre si em diretórios especiais formando as chamadas "toolbox" (caixa de ferramentas). Essas ferramentas são mais do que simples coleções de funções úteis; expressando os esforços de alguns dos maiores pesquisadores do mundo em diversas áreas, elas são uma coleção de funções do Matlab (chamadas "*m-files*"), constituindo uma extensão do ambiente Matlab na resolução de classes específicas de problemas. De grande importância para a grande maioria dos utilizadores, essas caixas de ferramentas permitem a aprendizagem e aplicação de tecnologias especializadas.

Áreas para as quais estão disponíveis "*toolboxes*" incluem processamento de sinal e de imagem, aquisição de dados, sistemas de controle, redes neurais, lógica difusa, simulação e etc.

## CONCEITOS E APLICAÇÕES DA TEORIA DA COMUNICAÇÃO

### Sinais e sistemas de comunicação

Um sinal é formalmente definido como uma função de uma ou mais variáveis, a qual veicula informações sobre a natureza de um fenômeno físico. Quando a função depende de uma única variável, diz-se que o sinal é *unidimensional*. Quando a função depende de duas ou mais variáveis, diz-se que o sinal é *multidimensional*. Um sinal sonoro é um exemplo de sinal unidimensional cuja amplitude varia com o tempo. Uma imagem é um sinal multidimensional, cujas duas dimensões são representadas pelas coordenadas horizontal e vertical da imagem.

Os sinais podem ser classificados de várias maneiras. Uma delas baseia-se no tempo de emissão. Assim, diz-se que um sinal  $x(t)$  é um *sinal de tempo contínuo* se ele for definido para todo tempo  $t$ . Um sinal de tempo discreto é definido somente em instantes isolados de tempo. Os sinais podem ser classificados também de acordo com sua paridade ou simetria em relação a ordenada como pares (simétricos,  $x(-t) = x(t)$ ) ou ímpares (anti-simétricos,  $x(-t) = -x(t)$ ). Também podem ser classificados em *sinais periódicos*, se o sinal for uma função que satisfaça a condição  $x(t) = x(t + T)$ , sendo o período  $T$  uma constante positiva.

O sinal não-estacionário é aquele cujas características intrínsecas variam com o tempo, como os sinais da fala, nos quais os sons que os constituem são classificados em fonemas sonoros e fonemas surdos. Diz-se que um sinal é aleatório se ele for imprevisível. Apesar de não ser possível fazer uma previsão precisa de um sinal aleatório, ele pode ser descrito em termos de suas propriedades estatísticas (COHEN, 1995).

Um sinal caótico é definido como sendo um sinal de aparência aleatória gerado por um sistema de equações diferenciais de diferenças não-lineares com parâmetros fixos e que, diferentemente do ruído, é ligeiramente previsível, mas não tão previsível quanto num processo puramente determinístico (OPPENHEIM e SCHAFER 1975).

A entidade que manipula/processa um ou mais sinais para realizar uma função, produzindo, assim, novos sinais, é denominada *sistema*. Existem vários tipos de sistema: entre eles podemos citar o sensoriamento remoto, sistemas auditivos, processamento de sinais biomédicos, sistemas de controle, sistema sensorial etc. Neste trabalho, o interesse concentra-se no sistema de comunicação sonora.

Há três elementos básicos em todo sistema de comunicação: o *transmissor*, o *canal* e o *receptor*. O transmissor está localizado num ponto do espaço e o receptor em algum outro ponto, separado do transmissor; o canal é o meio físico que os liga. Cada um destes três elementos pode ser visto como um sistema com sinais próprios. A função do transmissor é converter o sinal de mensagem produzido por uma fonte de informação em uma forma apropriada ao canal. Devido às características físicas deste canal, o sinal transmitido pode ser distorcido à medida que se propaga por ele. A função do receptor é operar no sinal recebido a fim de reconstruir uma forma reconhecível (isto é, produzir uma estimativa) do sinal da mensagem original e entregá-lo ao usuário final (OPPENHEIM e SCHAFFER 1975).

Os detalhes das operações executadas no transmissor e no receptor dependem do sistema de comunicação utilizado, o qual pode ser *analógico* ou *digital*. Em um sistema de comunicação analógico (que manipula sinais contínuos no tempo), o transmissor consiste em *modulador* e o receptor, num *demodulador*. Modulação é o processo de converter o sinal da mensagem em uma forma compatível com as características de transmissão de canal. Comumente, o sinal transmitido é representado como uma variação de amplitude, fase ou frequência de uma onda portadora senoidal. Neste caso, fala-se de modulação de amplitude, fase ou frequência, respectivamente. Correspondentemente, através do uso da demodulação de amplitude, de fase ou de frequência, uma estimativa do sinal da mensagem original é produzida na saída do receptor. Cada uma destas técnicas de modulação/demodulação analógica tem suas próprias vantagens e desvantagens.

Em contrapartida, um sistema de comunicação digital (que processa sinais discretos no tempo) é consideravelmente mais complexo. Se o sinal da mensagem for analógico, como a maioria dos sinais sonoros, o transmissor executará as seguintes operações para convertê-lo em forma digital:

- *Amostragem*: converte-se o sinal da mensagem em uma seqüência de números, na qual cada um deles representa a amplitude do sinal da mensagem em um instante de tempo particular.
- *Quantização*: representa-se cada número produzido pelo amostrador no nível mais próximo escolhido dentre um número finito de níveis de amplitude discreta.
- *Codificação*: representa-se cada amostra quantizada por meio de uma palavra ou um código composto de um número finito de símbolos.

Depois da combinação de amostragem e quantização, obtém-se uma representação do sinal da mensagem que é *discreta* tanto em termos de tempo como de amplitude, ou seja, realiza-se a *digitalização* do sinal.

Diferentemente das operações de amostragem e codificação, a quantização é completamente irreversível. Porém, usando-se um quantizador com um número suficiente grande de níveis de amplitude distintos, esta perda pode se tornar pequena e não perceptível para todos os efeitos práticos. À medida que o número de níveis de amplitude discreta se eleva, o tamanho da palavra de código também deve elevar-se (OPPENHEIM e SCHAFER 1975).

O transmissor pode envolver operações adicionais como compactação de dados e codificação de canal. O propósito da compactação de dados é reduzir o número de bits por amostra necessários para a transmissão e remover informações redundantes do sinal da mensagem de forma a promover uma utilização eficiente do canal. A codificação de um canal, por sua vez, envolve a inserção de elementos redundantes na palavra de código (por

exemplo, símbolos extras) de uma maneira controlada; isto é feito para oferecer proteção contra o ruído e sinais de interferência capturados durante a transmissão através do canal. Finalmente, o sinal codificado é modulado numa onda portadora (usualmente senoidal) para ser transmitido pelo canal. No receptor, essas operações são executadas em ordem inversa. Uma estimativa do sinal da mensagem original é assim produzida e entregue ao usuário final.

As classificações de sinais, assim como os sinais elementares utilizados, as operações básicas em sinais, o teorema de Niquist e as propriedades dos sistemas são bem definidos e facilmente encontrados na literatura. Os pioneiros na abordagem de processamento de sinais são A. V. Oppenheim e R. V. Schafer.

Nesta pesquisa, sinais unidimensionais (sons) foram interpretados como sinais multidimensionais (imagens). Devido ao maior número de ferramentas que um sistema digital oferece para o processamento do sinal, eles foram manipulados em sistemas de comunicação digital como forma de aproximação de um sistema de comunicação analógico (comunicação sonora em aves).

## **O sinal sonoro**

As ondas podem ser de natureza eletromagnética ou mecânica. As ondas mecânicas são originadas pela deformação de uma região de um meio material elástico. Se a direção de propagação de uma onda for perpendicular à direção de vibração ela é denominada onda *transversal*. Se a direção de propagação de uma onda for a mesma que a direção de vibração ela é denominada *longitudinal*.

Uma *onda sonora* é uma onda mecânica longitudinal tridimensional, originada por uma fonte capaz de alterar a pressão do meio de propagação. Frequentemente, as zonas de maior pressão, ou compressão (maior densidade das moléculas de ar), alternam-se com zonas de pressão reduzida, ou rarefação (menor densidade das moléculas de ar).

O som é a sensação auditiva criada por uma onda sonora. As ondas sonoras têm propriedades que dependem exclusivamente de suas características físicas. O som possui propriedades relacionadas à capacidade do nosso sistema auditivo. Normalmente, a palavra som é usada para significar o fenômeno físico em si, ocasionando distorções conceituais quanto aos seus atributos físicos e fisiológicos. Por isso as principais qualidades fisiológicas do som devem ser devidamente associadas às características físicas da onda sonora como indica a tabela abaixo:

Tabela 1– Principais Fenômenos Correspondentes do Som.

Acústico / Elétrico	Descrição	Psicoacústico	Descrição
<i>Frequência</i>	Representada em Hertz (Hz), é o número de repetições de uma vibração em 1 segundo (ciclos por segundo).	<i>Altura (Pitch)</i>	É o atributo da percepção auditiva, segundo o qual os tons podem ser classificados numa escala que se estende dos mais graves (freq. baixas) aos mais agudos (freq. altas).
<i>Amplitude (Intensidade)</i>	Representada em, entre outros, W/m <sup>2</sup> . É o fluxo de energia sonora por unidade de área em direção normal à propagação.	<i>Audibilidade (Loudness)</i>	É a impressão auditiva que um observador tem da intensidade de uma fonte sonora.
<i>Forma de Onda</i>	Define perfeitamente a natureza de uma vibração completada num determinado ciclo.	<i>Timbre</i>	É a impressão auditiva que um observador tem da forma de onda de uma fonte sonora.
<i>São fenômenos físicos que podem facilmente ser medidos por instrumentos de medidas.</i>		<i>São fenômenos que são sentidos, mas não podem ser medidos. A percepção desses fenômenos é individual.</i>	

- **Altura** é a percepção da frequência do som. A frequência é expressa em ciclos por segundo ou Hertz. Um som caracterizado por uma única frequência é chamado de tom ou som puro. Na vida cotidiana, porém, praticamente não existem sons puros. Os sons comuns, desde o som mais musical ao maior ruído, apresentam sempre um grande número de frequências. Os sons que consideramos como sendo musicais apresentam um número limitado de frequências: trata-se geralmente de um tom fundamental com diversas frequências harmônicas superpostas. O tom fundamental determina o “período de repetição” do complexo de flutuações da pressão sonora. As frequências harmônicas, ou harmônicos, são tons superpostos, cujas frequências são múltiplos integrais da

freqüência fundamental. Quando um som abrange uma grande variedade de freqüências, ele é considerado como “ruído”; quando todas as freqüências estão presentes com a mesma intensidade, o som é chamado de *ruído branco*. Outros ruídos apresentam espectros diferentes de freqüência, mas o que caracteriza todos estes sons é que o registro das variações temporais da pressão sonora não revela uma periodicidade evidente. O tom de um som é afetado por outras características do som, tais como a amplitude e o timbre.

- **Audibilidade ou Potência do som** pode ser definida como a percepção da amplitude do som. A amplitude de um som pode ser medida com o auxílio de um transdutor eletroacústico. A escala de pressões sonoras capazes de afetar o sistema auditivo é tão ampla que foi necessário empregar uma escala logarítmica para definir o nível de pressão sonora. Para definir este nível, adota-se como ponto de referência um nível de pressão arbitrário, mas que corresponde aproximadamente ao limiar para a audição  $P_0 = 10^{-12} \text{ W/m}^2$ . O nível de pressão sonora NPS, mais conhecido como SPL (*Sound Pressure Level*) de uma dada pressão sonora  $p$ , em decibéis, é definido por:  $\text{NPS} = 20 \log P/P_0$ . O nível mais baixo para a audição é o zero dB absoluto, o mais alto já seria o limiar do pânico ou da dor correspondendo a 120dB.
- **Timbre** é a qualidade característica de um dado som. É um atributo do som que permite ao ouvinte julgar que dois sons de mesma potência e tom são de fontes diferentes. A diferença perceptível entre um Dó médio de um clarinete e o Dó médio de um violino é identificada graças ao timbre. É um atributo complexo e pode ser definido fisicamente como a forma de onda; o componente principal do timbre é a repartição da energia sonora entre harmônicos; as variações temporais da amplitude podem também influir no timbre.

Um ponto de partida para uma outra abordagem na representação gráfica de ondas periódicas baseia-se na análise de uma relação estabelecida entre um movimento oscilatório senoidal e a definição matemática de uma circunferência. O percurso de uma trajetória

circular no sentido anti-horário assume, em cada ciclo, valores entre 0 e  $2\pi$  rad. A quantificação em cada momento do arco percorrido resulta a grandeza designada por fase. Assim, pode-se pensar numa representação gráfica de uma forma de onda apresentando a sua amplitude em função da fase, correspondendo a variação entre 0 e  $2\pi$  rad à duração de um período do sinal.

O transporte de uma informação através de um canal de comunicação se dá por meio da modulação de três características básicas de um sinal sonoro: amplitude, frequência e fase. Por isso, muitos autores denominam a onda sonora como uma onda portadora.

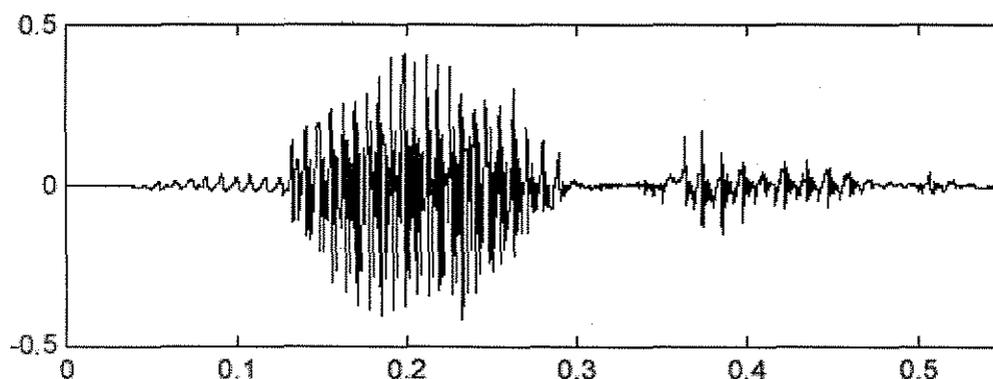


Figura 3 – Sinal sonoro de voz correspondente à frase “Boa Tarde”.

Transdutor é um dispositivo que efetua a conversão de energia de uma forma em outra. Para converter o som em um sinal digital, é necessário, antes, que o som seja convertido em sinal elétrico. Para essa primeira conversão existem os transdutores acústico-elétricos. Um tipo bastante conhecido de transdutor acústico-elétrico é o microfone. Esse transdutor, excitado por ondas sonoras, produz ondas elétricas substancialmente equivalentes.

Uma vez estando o som convertido em sinal elétrico, pode-se, então, convertê-lo em informação digital. Isso é feito capturando-se a amplitude de onda, ou seja, a tensão da corrente elétrica em um determinado tempo  $t$ . No entanto, um registro apenas não basta para caracterizar a informação que está disposta ao longo da onda. De acordo com o *teorema da amostragem*, são necessários diversos registros em diversos momentos diferentes para capturar a informação contínua. Essa periodicidade de amostra é chamada de *taxa de amostragem*.

## Teoria de Fourier

O mais simples sinal que varia com o tempo é o senoidal. Esse tipo de sinal (ou função) é a solução de muitas equações fundamentais e é comum na natureza. Ele é caracterizado por uma amplitude  $A$  e frequência  $\omega_0$  constantes.

A história da teoria de Fourier teve início quando grandes cientistas, como D. Bernoulli (1700-1782), L. Euler (1707-1783) e L. d'Alembert (1717-1783), tentaram encontrar a solução da equação de onda como a que governa uma corda vibrante de equação diferencial do tipo:

$$\frac{\partial^2 \Psi(x,t)}{dx^2} = \alpha \frac{\partial^2 \Psi(x,t)}{dt^2} \quad [01]$$

Primeiramente Bernoulli representou a posição inicial da corda em termos de uma soma infinita de funções senos com diferentes frequências. Depois Euler encontrou uma maneira conveniente para calcular o que hoje são denominados coeficientes de Fourier. Bem mais tarde, Joseph Fourier (1768-1830) apresentou em seu "*Théorie Analytique de la Chaleur*", publicada em 1822, um interessante estudo representando funções como superposições ponderadas de senos e co-senos para a solução da equação de calor, uma equação diferencial do tipo:

$$\frac{\partial^2 \Psi(x,t)}{dx^2} = \alpha \frac{\partial \Psi(x,t)}{dt} \quad [02]$$

Essas representações ficaram conhecidas como série de Fourier (FS – *Fourier Series*) (COSTA e CESAR 2001).

A partir de então, os sinais puderam ser descritos no domínio do tempo  $x(t)$  ou no domínio da frequência  $X(\Omega)$  por meio de transformações lineares denominadas *Transformadas de Fourier* (FT – *Fourier Transform*). Sendo assim, qualquer sinal pode ser graficamente ou matematicamente representado de duas maneiras: no domínio do tempo e no da frequência.

### **Representações dependentes do sinal**

Em termos de quantidades físicas reais, a amplitude pode representar uma medida, como a da pressão exercida pela vibração do ar, no caso do som, ou a de uma voltagem em algum ponto de um circuito elétrico. No domínio da frequência, as amplitudes de um sinal são representadas como uma função da frequência.

A mudança de representação de um sinal do domínio do tempo para o domínio da frequência, que mostra como diferentes componentes de frequência contribuem para a composição do sinal, é conhecida como análise espectral. A FT é a função matemática que converte um sinal do domínio do tempo contínuo em espectro.

A representação no domínio da frequência consiste em duas partes. O espectro de magnitude  $|X(e^{j\omega})|$ , que contém informações sobre a magnitude de cada componente de frequência do sinal inteiro, e o espectro de fase  $\arg |X(e^{j\omega})|$  (BRIGHAM, 1988).

Quando o sinal é representado como uma seqüência de amostras discretas, é usada a versão da transformada de Fourier conhecida por DTFT (*Discrete Time Fourier Transform*). Os sinais periódicos contínuos no tempo são representados como Séries de Fourier (FS). Se o sinal for discreto no tempo, a representação será a Série de Fourier de Tempo Discreto (DTFS – *Discrete Time Fourier Series*).

Tabela 2 – Relação entre propriedades de um sinal e a representação de Fourier apropriada

Propriedades	<i>Periódica</i>	<i>Não Periódica</i>
<i>Contínuo</i>	FS	FT
<i>Discreto</i>	DTFS	DTFT

### A transformada rápida de Fourier

A DTFT é a única representação de Fourier que pode ser avaliada e manipulada numericamente em um computador, sendo muitas vezes usada para aproximar numericamente as outras três representações. A transformada rápida de Fourier (FFT – *Fast Fourier Transform*) pode ser compreendida como uma forma numérica de computar a DTFT de um sinal ou de uma função. A maioria dos sistemas que realizam o cálculo da FFT utilizam o clássico algoritmo de Cooley – Tukey criado em 1969 (BRIGHAM 1988).

As definições e as propriedades da FFT podem ser estendidas a funções definidas num intervalo contínuo ou discreto. Também podemos estender essas definições e outras propriedades para duas dimensões. Uma propriedade particularmente importante da transformada de Fourier é a da correlação.

### Análise de tempo-freqüência

A teoria de Fourier é válida somente para sinais estacionários. Para a análise de sinal não-estacionário, o método preferido é uma descrição do sinal que envolva simultaneamente tempo e freqüência. Como o nome implica, a análise de tempo-freqüência associa um sinal temporal (uma função unidimensional do tempo) com uma imagem (uma função bidimensional de tempo e freqüência) que exhibe os componentes espectrais do sinal como uma função do tempo.

Em termos conceituais, pode-se imaginar essa representação como um mapeamento espectral variável no tempo do sinal. Ela é análoga a uma partitura musical, na qual o

tempo e a frequência representam os dois eixos principais. Os valores da representação de tempo-frequência do sinal fornecem uma indicação dos tempos específicos nos quais certos componentes espectrais do sinal são observados (COHEN 1995).

A transformada breve de Fourier (STFT – *Short Time Fourier Transform*) é uma extensão natural da FT, por isso muitas das propriedades da FT são transportadas para a STFT.

Suponha-se que  $x(t)$  denote um sinal que seja, *a priori*, considerado “estacionário” e que é visualizado através de uma janela temporal  $w(t)$  de extensão limitada (fig.4a). A STFT de  $x(t)$ , denotada por  $X(\tau, \omega)$ , é definida como a transformada de Fourier do sinal janelado (fig.4c); ou seja:

$$X(\tau, \omega) = \int_{-\infty}^{+\infty} x(t)w^*(t - \tau) e^{-j(2\pi f)t} dt \quad [03]$$

Muitas formas de janelas são usadas na prática. Tipicamente, elas são simétricas, unimodais e, em geral, têm valor complexo. O parâmetro  $\tau$  representa a posição central da janela (fig.4b).

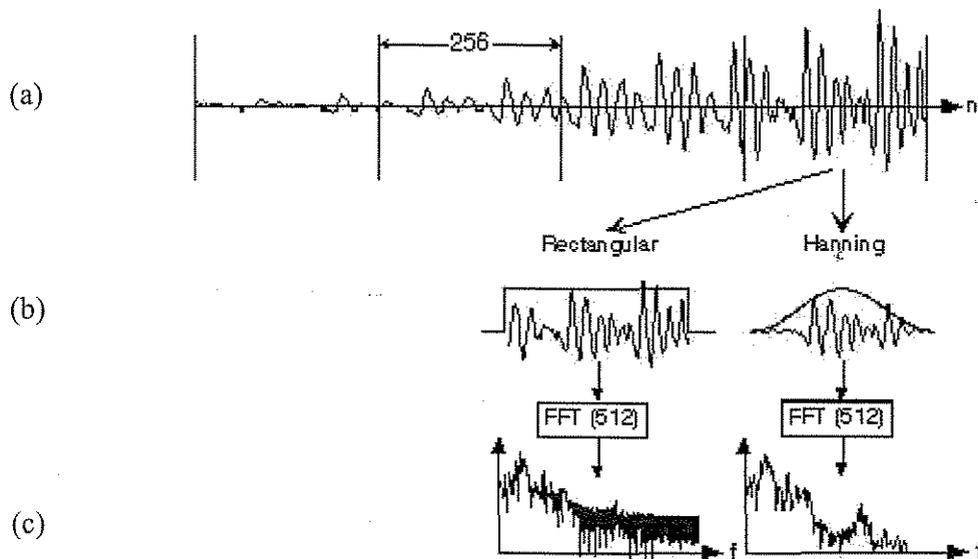


Figura 4 – (a) um sinal  $x(t)$  visualizado através de uma janela temporal de extensão limitada. (b) Uma janela retangular e uma janela de Hanning. (c) a STFT de  $x(t)$  linear ao sinal.

## Espectrograma

O módulo ao quadrado da STFT de um sinal  $x(t)$  é chamado de espectrograma do sinal. O espectrograma representa uma extensão simples, mas poderosa, da clássica teoria de Fourier. Em termos físicos, o espectrograma fornece uma medida da energia do sinal no plano de tempo-frequência.

O espectrograma é uma imagem que representa visualmente a variação da frequência de um sinal no decorrer do tempo e, simultaneamente, fornece precisas informações da energia do sinal (através de cores ou de níveis de cinza) descrevendo perfeitamente todos os detalhes do som em questão.

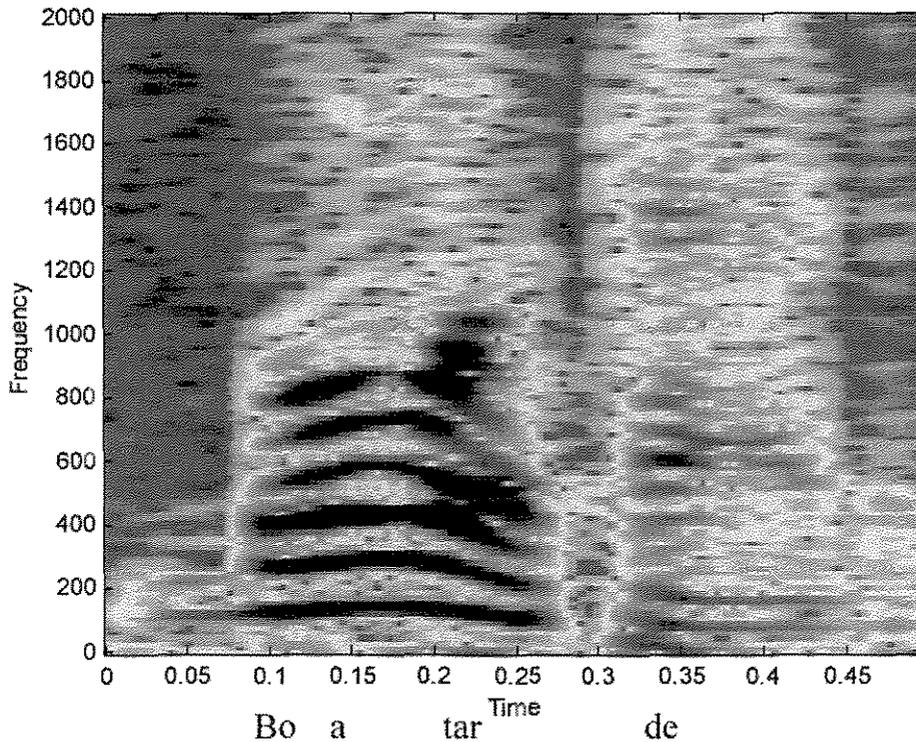


Figura 5 – O espectrograma do sinal sonoro da Figura 3.

A cor de um padrão particular nos espectrogramas é indicativa da energia de sinal nesse padrão. O código de cores em energia decrescente é o vermelho, o amarelo e depois o

azul. Em bioacústica, usualmente, utiliza-se o código de cores em energia decrescente preto e branco, porém, o padrão usual em processamento digital de imagens (PDI) é o inverso.

Uma grande preocupação ao usar a STFT é a questão da resolução de tempo-frequência. Especificamente, dado um par de sinais puramente senoidais cujas frequências angulares sejam espaçadas com uma separação de  $\Delta\omega$  rad/s, o menor valor de  $\Delta\omega$  para o qual os dois sinais podem ser resolvidos ou distinguidos é chamado resolução de frequência. A correspondente duração da janela  $w(t)$  é chamada resolução de tempo, denotada por  $\Delta\tau$ . A resolução de frequência  $\Delta\omega$  e a resolução de tempo  $\Delta\tau$  são inversamente relacionadas por  $\Delta\tau\Delta\omega \geq \frac{1}{2}$  que é uma manifestação de dualidade da STFT, herdada da FT (BEECHER 1988).

Esta relação é chamada princípio da incerteza, uma analogia com um termo usado em mecânica quântica. A condição especial de igualdade é satisfeita pelo uso de uma janela gaussiana. Conseqüentemente, a capacidade de resolução de tempo-frequência da STFT é fixa ao longo do plano inteiro de tempo-frequência.

A função da janela da STFT é extrair uma parte do sinal para análise e assegurar que a secção que está sendo analisada seja aproximadamente estacionária. Para que a janela realize a sua função é necessário esclarecer duas questões: o tipo de janela e a sua dimensão ou duração.

### **Tipos de janelas**

O espectro de cada janela revela parâmetros que caracterizam cada uma delas para uma determinada análise. Esses parâmetros são o *lóbulo central* e os *lóbulos laterais* (ou lóbulo principal e lóbulos secundários). A *largura* do lóbulo central é importante para separar frequências muito próximas. A *amplitude* dos lóbulos laterais (relativo à amplitude do lóbulo central) é importante no controle do grau de influência de uma componente nas

outras componentes vizinhas do sinal – *Spectral Leakage* – e está relacionado com as abruptas variações na amplitude do sinal do início ao fim da janela (COHEN 1995).

A equação da janela retangular é dada por:

$$w[n] = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{Para outros valores de } n \end{cases} \quad [04]$$

A janela retangular pode ser entendida como a amostragem do sinal de entrada, tomada sobre N pontos, cada um destes com pesos iguais a 1 (OPPENHEIM e SCHAFER 1975). Nota-se um lóbulo secundário 13 dB abaixo do lóbulo principal, indicando um espalhamento espectral significativo para a janela retangular. A janela retangular nada mais é que um termo para designar o fato de que um sinal contínuo no tempo deve ser amostrado durante um intervalo de tempo finito, desprezando-se os valores do sinal fora deste intervalo. Os pontos assim amostrados são ponderados uniformemente com uma função unitária, que é a janela retangular, de forma a manter seus valores originais inalterados.

A equação da janela triangular é dada por:

$$w[n] = \begin{cases} 2n/N & n=0,1,2,\dots,N-1 \\ a(N-n) & n=(N/2)+1,(N/2)+2,\dots,N-1 \end{cases} \quad [05]$$

Com a utilização desta janela, o lóbulo secundário fica 26 dB abaixo do lóbulo principal e tem a largura em frequência duplicada.

A janela de Hanning, mais comumente chamada de janela<sup>6</sup> de Hann, oferece grandes melhorias quando comparadas com as anteriores. A função de Hann é definida como:

$$w[n] = \begin{cases} 0,5[1 - \cos(2\pi n / N)] & n = 0,1,2,\dots,N-1 \\ 0 & \text{demais} \end{cases} \quad [06]$$

No espectro da janela de Hann, o lóbulo secundário está aproximadamente 32 dB abaixo do lóbulo principal.

A janela Hamming é muito semelhante à janela de Hann, porém se utiliza coeficientes diferentes em sua equação para os termos em co-seno:

$$w[n] = 0,54 - 0,46[\cos(2\pi n / N)] \quad [08]$$

É possível observar a resposta em frequência e a curva da função na figura 6c. Nota-se que o lóbulo secundário está aproximadamente 43 dB abaixo do lóbulo principal. As funções de ponderação de Hann e Hamming são exemplos de funções no domínio do tempo com um termo co-senoidal.

$$w[n] = \sum_{m=0}^M (-1)^m b_m \cos(2\pi m n / N) \quad n = 0, 1, 2, \dots, M \quad [09]$$

Estas funções podem ser expandidas como um somatório de termos em co-senos, onde a somatória dos termos  $b_m$  é unitária. A função de *Blackman* é uma expressão de três termos onde  $b_0 = 0,42$ ,  $b_1 = 0,5$  e  $b_2 = 0,08$ . *Harris* usa um conjunto de janelas com três ou quatro termos diferentes de zero, que resultam em um nível mínimo para os lóbulos laterais.

As curvas da equação e resposta em frequência da *Janela de Blackman-Harris*, para 3 ou 4 coeficientes, estão demonstradas na figura 4(d), onde foi adotado o caso de 4 termos para atenuação de 92 dB do lóbulo secundário.

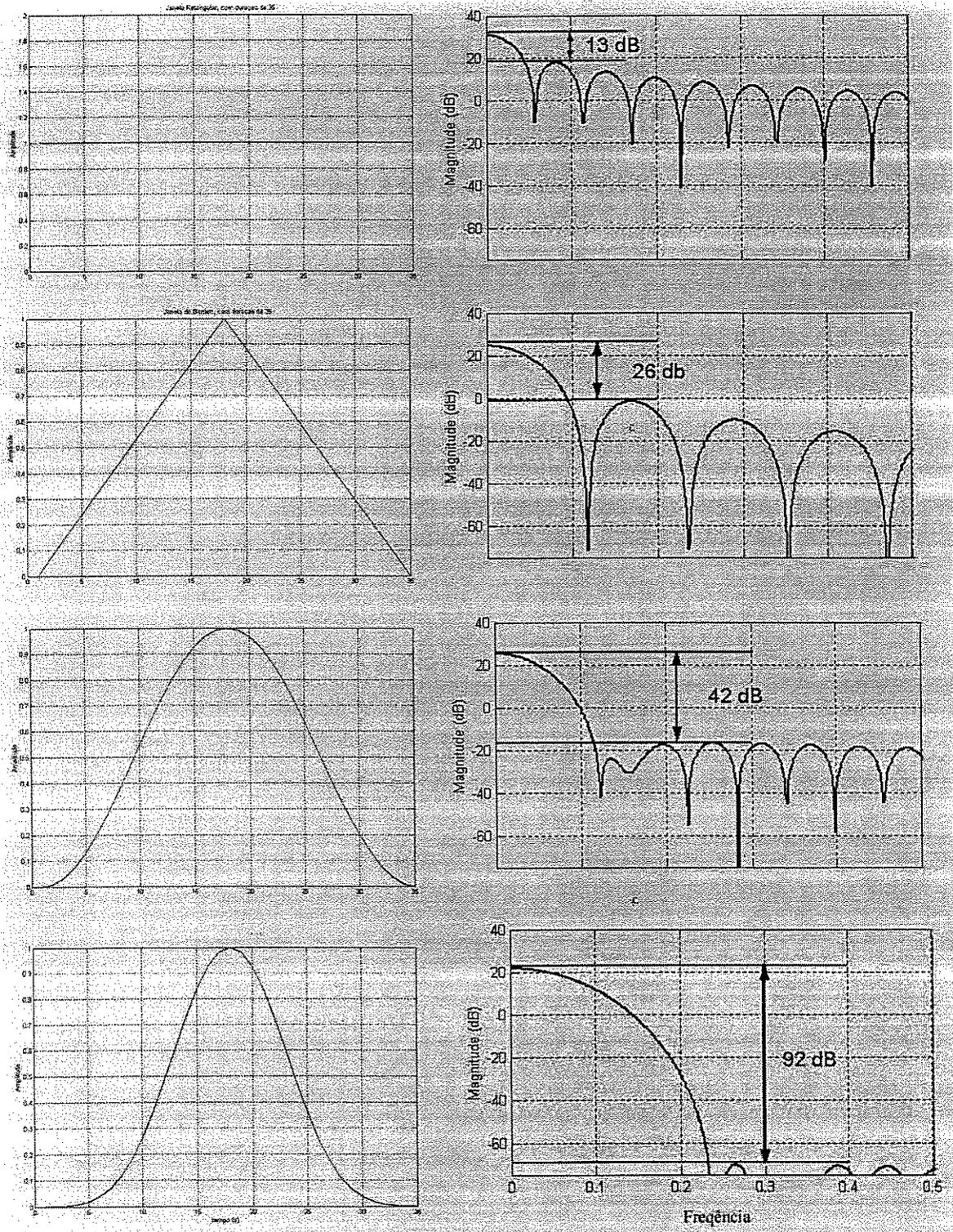


Figura 6 – (a) Janela retangular; (b) Janela de Bartlett; (c) Janela de Hamming; (d) Janela Blackmann-Harris; todas com  $N=35$ .

## **Processamento digital de imagens**

O sentido da visão pode ser considerado o meio mais eficiente de que o ser humano dispõe para captar as informações originadas no ambiente que o cerca. Algumas aplicações e alguns tipos de imagens requerem uma interação visual bastante intensa. A capacidade humana para processar e interpretar imensas quantidades de dados de natureza visual motiva o desenvolvimento de técnicas e dispositivos que estendem ainda mais essa capacidade e essa sensibilidade.

O conhecimento de como reage o sistema visual humano e de algumas técnicas disponíveis para melhor adequar a imagem à aplicação é importante para desenvolver soluções computacionais automatizáveis. Nesta ótica, a área de Processamento e Análise de Imagens permite viabilizar um grande número de aplicações, tanto no domínio do aprimoramento de informações picturais para interpretação humana quanto no domínio da análise automática, por computador, de informações extraídas de uma imagem ou cena.

Entre os seus objetivos principais, destacam-se a implementação e o desenvolvimento de técnicas e metodologias voltadas para a representação e o processamento de imagens. Nesta última década, o processamento digital de imagens está entre as tecnologias computacionais que mais rapidamente estão evoluindo, sendo estudado cada vez mais por diversas áreas como: engenharia, física, ciência da computação e matemática. Trata-se de uma área emergente, cuja potencialidade de aplicação tem dado origem a projetos e estudos de casos em sistemas de processamento de imagens, voltados para problemas específicos de outras áreas como engenharia, medicina, telecomunicações.

## Conceito de Imagem Digital

Uma imagem é uma figura que dê uma representação visual de um objeto ou cena. No computador, a imagem é um arranjo bi-dimensional de números. Cada número corresponde a uma pequena área, denominada *pixel*, sendo que os valores de cada número representam as cores de cada *pixel*.

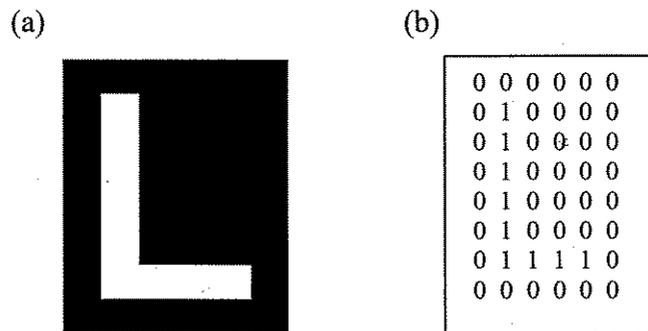


Figura 7 – (a) Uma simples imagem digital de duas bandas e (b) o correspondente arranjo de números.

O conceito de imagem digital torna possível a visualização de uma imagem como um gráfico tridimensional comumente utilizado para representar funções bivariáveis.

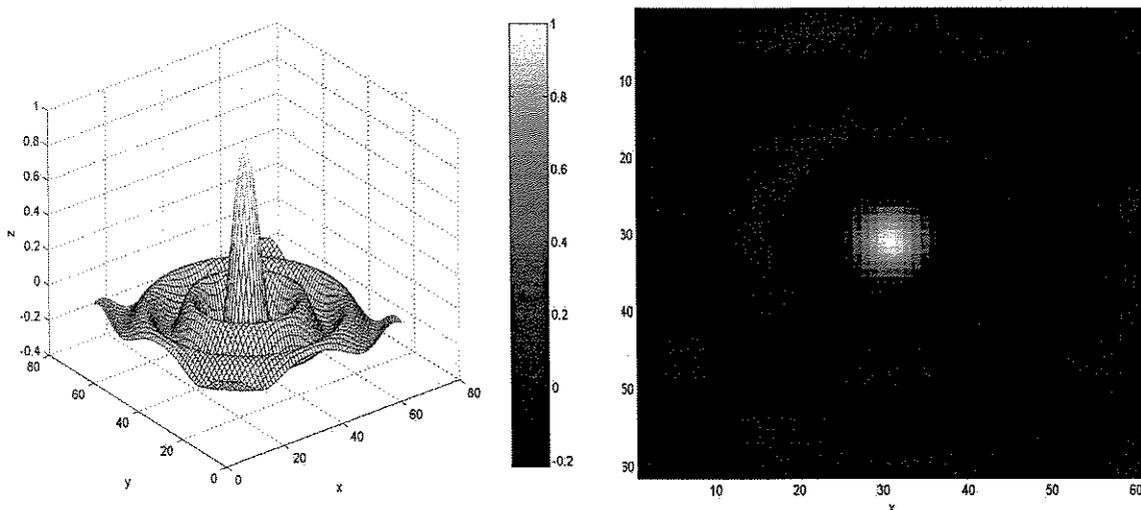


Figura 8 – (a) Representação gráfica tridimensional da função sinc e (b) imagem correspondente em nível de cinza. (Figura de difração por um orifício retangular)

## Morfologia Matemática

As transformações lineares, como a Transformada de Fourier, foram inicialmente definidas para sinais unidimensionais; mais tarde foram estendidas para imagens. Porém, já que as formas nas imagens não são privilegiadas pelas transformações lineares, para explorá-las foi necessário desenvolver uma abordagem denominada morfologia matemática, baseada na teoria dos conjuntos.

A morfologia matemática, sugerida na década de 60 na França, tem uma formulação baseada em reticulados complexos. Um *reticulado complexo* é formado pela coleção de todos os subconjuntos de um conjunto  $E$ , com relação de ordem parcial  $\subseteq$  usual entre os conjuntos e as operações de intersecção, união e complemento. Assim, mapeamentos de imagens podem ser vistos como operadores entre reticulados complexos (GONZALEZ e WOODS 2002).

Os mapeamentos da morfologia binária podem ser expressos através de operações elementares sobre conjuntos e mais dois operadores básicos, conhecidos como erosão e dilatação. A erosão e a dilatação são operadores elementares que formam a base para a construção de transformações mais complexas. Assim, numa cadeia morfológica de processamento de imagens, encontra-se um grande número de operadores encadeados, todos definidos a partir destas funções elementares (GONZALEZ e WOODS 2002).

Todos os operadores da morfologia matemática são caracterizados por subconjuntos denominados *elementos estruturantes*, uma imagem menor de forma conhecida, com características geométricas/topológicas relacionadas com a informação que se pretende extrair da imagem de interesse. A idéia básica é explorar a relação entre estes subconjuntos e o conjunto que define uma imagem binária (GONZALEZ e WOODS 2002).

## Erosão e Dilatação

A erosão de uma imagem  $X$  por um elemento estruturante  $B$  é definida como

$$\varepsilon_B(X) = \{x \in E : B_x \subseteq X\} \quad [10]$$

sendo  $B_x$  o conjunto  $B$  transladado por  $x$ .

A dilatação de uma imagem  $X$  por um elemento estruturante  $B$  é definida como

$$\delta_B(X) = \{x \in E : B_x \cap X \neq \emptyset\} \quad [11]$$

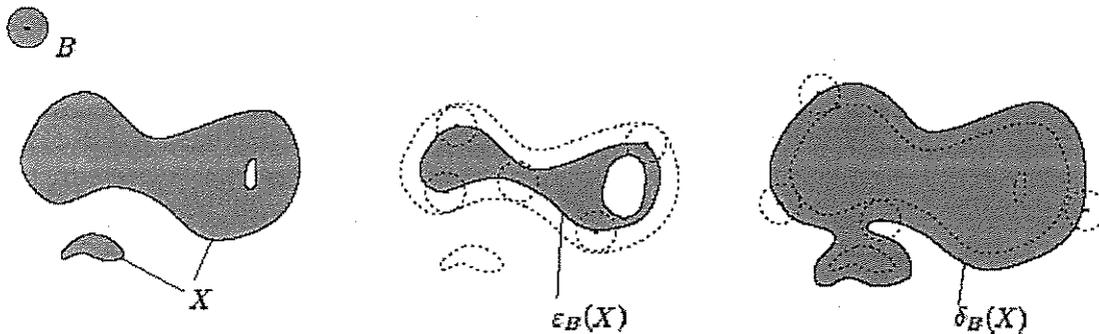


Figura 9 – Efeitos da erosão e dilatação binárias usando um elemento estruturante esférico de raio  $R$  numa imagem  $X$ .

A erosão diminui o tamanho dos objetos contidos na imagem, enquanto a dilatação aumenta. Note-se também que a erosão elimina objetos menores que o elemento estruturante, enquanto a dilatação elimina buracos e espaços menores que o elemento estruturante.

Esta não é apenas uma coincidência; de fato, há uma relação de dualidade entre esses dois operadores, dada por

$$\varepsilon_B(X) = (\delta_B(X^c))^c \quad [12]$$

onde  $c$  representa a operação de complementação usual de conjuntos.

Note-se também que, se a origem está contida em  $B$ , então:

$$\varepsilon_B(X) \subseteq X \subseteq \delta_B(X). \quad [13]$$

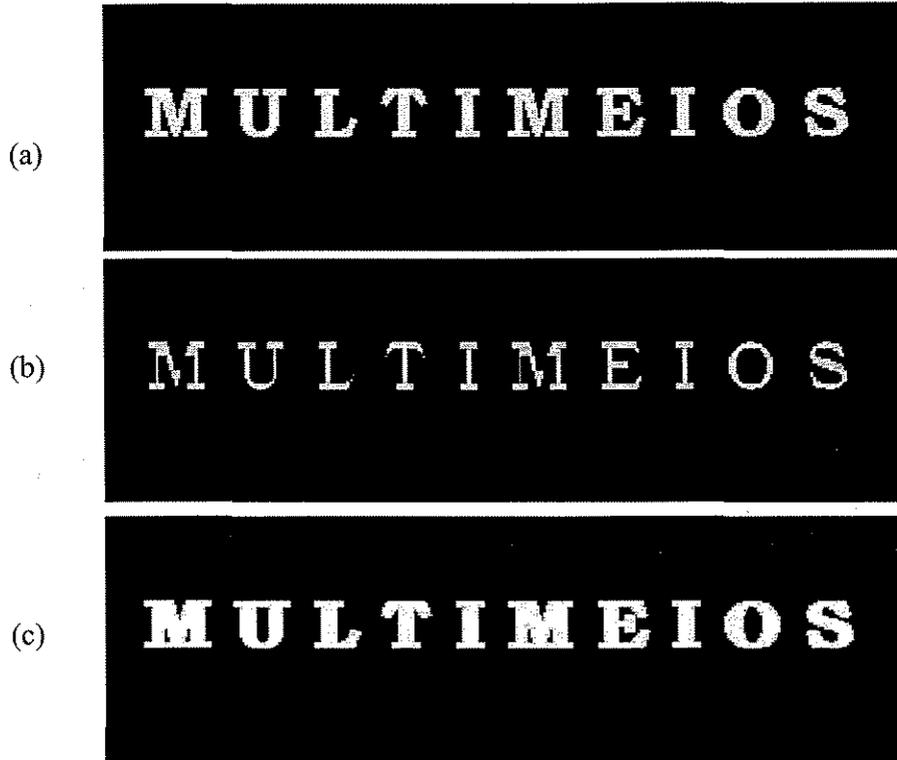


Figura 10 – (a) Imagem original, (b) imagem erodida e (c) imagem dilatada.

O operador  $\rho_B(X) = \delta_B(X) - \varepsilon_B(X)$  é denominado *gradiente morfológico*. O operador  $\rho^+_B(X) = \delta_B(X) - X$  é a borda ou gradiente externo e o operador  $\rho^-_B(X) = X - \delta_B(X)$  é a borda ou gradiente interno. O gradiente morfológico é a união da borda externa com a borda interna e, portanto, pode ser entendido como uma borda mais grossa.

O operador  $\gamma_B(X) = \delta_B(\varepsilon_B(X))$  é denominado abertura de  $X$  por  $B$  e o operador  $\varphi_B = \varepsilon_B(\delta_B(X))$  é denominado fechamento de  $X$  por  $B$ . A abertura elimina objetos menores que o elemento estruturante e partes pequenas do objeto. O fechamento preenche pequenos espaços entre objetos, pequenos buracos e reentrâncias estreitas.

## Reconhecimento de padrões

O reconhecimento de padrões é uma habilidade extremamente desenvolvida nos seres humanos e em alguns animais. Pode-se dizer que padrões são os meios pelos quais o mundo é interpretado. Desde os primeiros computadores registram-se grandes esforços para a elaboração de atitudes e decisões a partir de interpretações do mundo real. Assim, alguns conceitos importantes referentes à área de reconhecimento de padrões serão apresentados a seguir (SCHALKOFF 1992).

- **Padrão:** um padrão é um arranjo ou ordenação em que se pode dizer que exista alguma organização estrutural. Um padrão pode ser referenciado como uma quantidade ou descrição estrutural de um objeto ou de qualquer outro item de interesse, podendo ser complexo ou tão básico quanto um conjunto de medidas ou observações, como peso ou idade, por exemplo, geralmente sendo representado na forma de um vetor ou matriz.
- **Características:** as características podem ser entendidas como qualquer medição útil extraída de um processo, podendo ser, por exemplo, o resultado da aplicação de um algoritmo ou de um operador de extração de características nos dados de entrada. Intensidade de sinais e descrições geométricas de uma região são exemplos de *características*. Um grande esforço computacional é requerido neste processo, podendo os dados, muitas vezes, conter erros ou ruídos.
- **Vetor de características e espaço de características:** vetor de características é um vetor de dimensão  $d$ , denotado por  $x$ , que contém as características de forma organizada. Já um espaço de características é um espaço multidimensional que contém os vetores de características. Se, por exemplo, todas as características são números reais, pode-se dizer que o espaço de características é  $R^d$ .
- **Classificação:** a classificação consiste na atribuição de dados de entrada a uma ou mais classes  $c$  pré-especificadas. Esta atribuição é baseada na extração de características significativas ou atributos e no processamento ou análise destes atributos.

- **Reconhecimento:** o reconhecimento consiste no processo de classificação. Quando são analisados problemas de reconhecimento de padrões, são utilizadas  $(c + 1)$  classes, visto que uma classe deverá conter as saídas: “não classificado”, “não conhecido” ou “não pode decidir”.
- **Classe de padrões:** é um conjunto de padrões que possuem características similares. Uma das questões mais importantes em muitas aplicações de reconhecimento de padrões é identificar os atributos apropriados e formar uma boa medida de similaridade.
- **Pré-processamento:** é o processo de filtragem ou transformação dos dados de entrada para auxiliar na praticidade computacional, na extração das características e na minimização do ruído.

A partir desses conceitos, pode-se dizer que, dado um conjunto de características que definem um determinado objeto, a tarefa de reconhecer padrões consiste em fazer com que tais características sejam identificadas. De modo formal, o reconhecimento de padrões pode ser entendido como um mapeamento “opaco” entre o espaço de padrão e o espaço da classe associada. O mapeamento é dito “opaco” porque a função que leva ao espaço da classe associada não é conhecida. Segundo outra definição, dada por Fu (1982), o reconhecimento de padrões é uma associação de um padrão com uma determinada classe.

### **Reconhecimento de Padrões Auditivos**

Os sons não são reconhecidos apenas em forma de tom, intensidade e timbre, mas também em termos de origem da fonte sonora, cada fonte com seus próprios timbres, tons e intensidades. Quando os sons são percebidos como “fontes”, estruturas de um nível mais alto podem ser usadas para identificá-los. Deutsch (1981) e Bregman (1990), expuseram alguns fatores pelos quais os sons são agrupados como pertencendo a origens distintas.

- **Similaridade e diferença.** Para o agrupamento de sons deve ser considerada sua origem: similar ou distinta. Componentes que demonstram ter os mesmos atributos serão percebidos como relacionados e vice-versa. Os sons serão agrupados como sendo de uma origem comum se forem similares, i.e., se tiverem tons, timbres, potência, posição e intensidades parecidos.
- **Proximidade.** Componentes que estejam próximos com relação ao tempo ou à frequência serão percebidos como estando relacionados e vice-versa.
- **Continuidade.** Mudanças na intensidade ou frequência devem ser suaves e contínuas, para que determinados sons continuem a fazer parte de uma mesma fonte. Mudanças repentinas implicam novas fontes ativas.
- **Coerência.** Ligados por uma bem estruturada "continuidade", os componentes de uma fonte tendem a variar de maneira coerente, por exemplo, mudar de tom e intensidade juntos.
- **Frequência fundamental.** Se dois sons com frequências fundamentais diferentes são ouvidos juntos, os seus harmônicos não se confundem e serão percebidas duas fontes sonoras distintas. Se a fundamental for a mesma, a fusão dos harmônicos ocorrerá e os sons serão percebidos como pertencentes a uma única fonte.
- **Localização do som.** Sons originados de um mesmo local no espaço são tidos como de fonte comum. Sons cujas fontes estão separadas no espaço ajudam o sistema auditivo a se concentrar em uma única fonte, apesar do mascaramento provocado por outros sons.
- **Ritmo.** Padrões rítmicos tendem a ser percebidos como fontes. De acordo com Deutsch (1981), o ritmo é um dos mais poderosos fatores físicos de reconhecimento de padrões.

## Correlação

Ao obter dois diferentes sinais de um experimento, como, por exemplo, a temperatura e a pressão de um gás, uma questão que surge freqüentemente é se esses dois sinais estão relacionados, e se a ação em um sinal aciona uma resposta no outro sinal. Dependendo da simplicidade dos sinais, podemos encontrar correlações entre eles por mera inspeção visual. Entretanto, para sinais mais complexos (ou ruidosos), é quase impossível a detecção de uma correlação entre eles por inspeção visual. Neste caso, torna-se necessário utilizar recursos matemáticos.

A correlação matemática entre dois sinais  $g(t)$  e  $h(t)$  é definida como:

$$g(t) \circ h(t) = \int_{-\infty}^{+\infty} g^*(\tau) h(t + \tau) dt \quad [14]$$

A expressão acima multiplica duas séries temporais, sendo a segunda  $h(t)$  deslocada no tempo por uma quantidade  $\tau$  (conhecida como fator de atraso) e integra o sinal resultante. A propriedade de produto interno sugere uma interpretação da correlação matemática como uma medida de *similaridade* entre duas funções em relação a uma série de deslocamentos relativos entre elas (COSTA e CESAR 2001). A figura 11 ilustra este fato.

Uma maneira muito conveniente e eficiente de computar a correlação de dois sinais é fazer a FFT dos dois sinais tomar o complexo conjugado de um deles, calcular o produto entre eles e em seguida computar a IFFT do resultado para obter a correlação como uma função do tempo de atraso  $\tau$ , que nada mais é que a propriedade da FT que diz:

*Seja  $g(t)$  e  $h(t)$  funções reais de pares de transformadas:*

$$g(t) \leftrightarrow G(f) \quad e \quad h(t) \leftrightarrow H(f).$$

*Então:*

$$g(t) \circ h(t) \leftrightarrow G^*(f) H(f).$$

A seqüência de gráficos abaixo (fig.11) ilustra a facilidade de computar a correlação entre dois sinais quaisquer utilizando uma propriedade de correlação da FT e o *software* Matlab<sup>®</sup> 6.1:

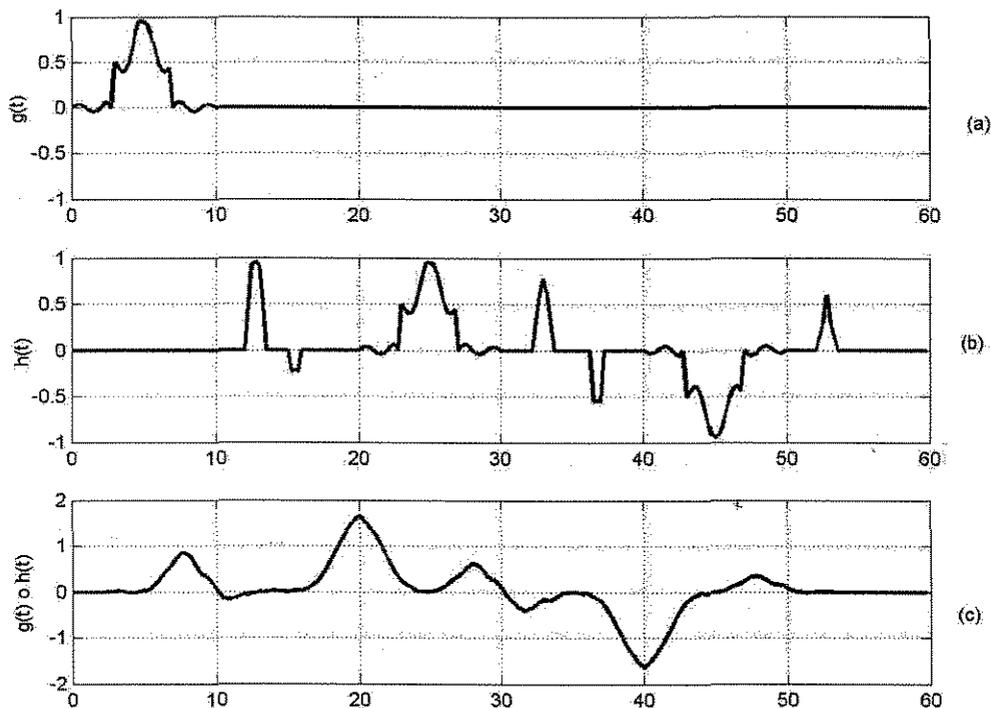


Figura 11 – (a) Uma função qualquer  $g_{(t)}$ , (b) Outra função qualquer  $h_{(t)}$ , (c) A correlação entre as duas funções,  $g_{(t)}$  e  $h_{(t)}$ , limitadas no tempo.

De acordo com a ilustração anterior, nesse tipo de análise os picos são identificados claramente e interpretados da seguinte maneira: a correlação  $f_{(t)} = (g \circ h)_{(t)}$  faz com que a função  $h_{(t)}$  deslize ao longo da função  $g_{(t)}$ ; calculando-se o produto interno para cada uma dessas situações, cada intensidade de correlação dá uma indicação da similaridade entre as funções. Na figura 11, por exemplo, a intensidade máxima é verificada para  $t = 20$  e  $t = 40$ , as quais correspondem às posições de referências contínuas na função  $g_{(t)}$ . Em outras palavras, a correlação permite buscar as posições em que duas funções são mais similares (no sentido amplo de produto interno).

Entretanto, podemos considerar que produtos internos implementados pela correlação têm sentido somente quando as funções têm suas amplitudes propriamente normalizadas, ou pelo menos próximas.

## Template Matching

Uma aplicação muito interessante da correlação em PDI é a técnica conhecida como *Template Matching*, a qual permite localizar numa imagem grande a ocorrência de objetos semelhantes a uma imagem pequena, o molde (*template*). Ao calcular a correlação entre a imagem grande e o molde, as posições nas quais a imagem grande se assemelha ao molde terão alto valor de correlação. Desta forma, uma busca de extremos na imagem da correlação fornecerá as posições e o valor desses extremos fornecerá o grau de semelhança.

Como exemplo ilustrativo da técnica *Template Matching*, suponha-se que se queira localizar as ocorrências da letra “M” na imagem abaixo (fig.12), a qual contém um texto. Utilizando o *software Matlab® 6.1*, é possível extrair uma das letras “M” da imagem e efetuar a correlação entre o *template* “M” e a imagem. Os picos claros correspondem às ocorrências da letra “M” no texto.

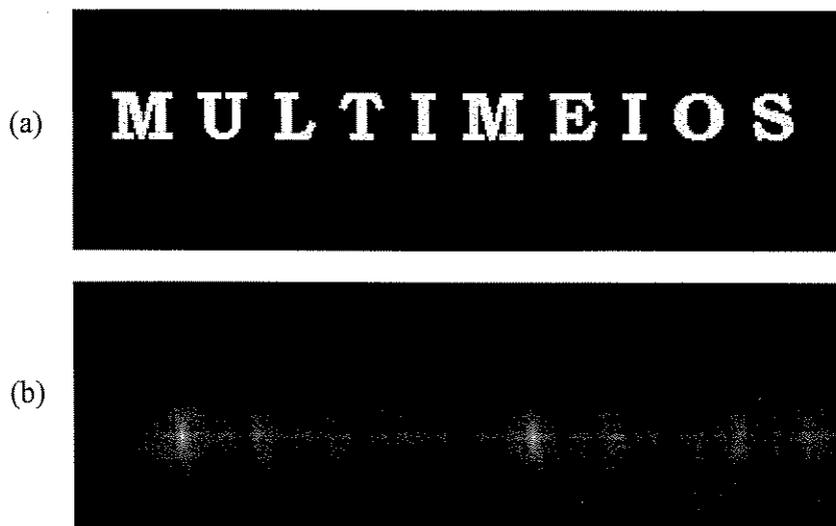


Figura 12 – (a) Uma imagem em que o molde (letra “M”) foi correlatado. (b) Localização e ocorrência do molde.

Este procedimento pode ser realizado no domínio das frequências ou do espaço. Quando o *template* é grande, o cálculo no domínio das frequências é mais veloz. Embora a correlação possa ser formulada em forma vetorial, trabalhar diretamente com um formato de imagem ou sub-imagem é mais intuitivo (e tradicional). Sendo assim, coincide uma sub-imagem  $h_{(x,y)}$  de tamanho  $J \times K$  dentro de uma imagem  $f_{(x,y)}$  de tamanho  $M \times N$ , e assume-se que  $J \leq M$  e  $K \leq N$ . Em sua forma mais simples, a correlação entre  $f_{(x,y)}$  e  $h_{(x,y)}$  é:

$$c(s, t) = \sum_x \sum_y f(x, y)h(x - s, y - t) \quad [15]$$

sendo  $s = 0, 1, 2, \dots, M - 1$ ;  $t = 0, 1, 2, \dots, N - 1$ , e a somatória é tomada sobre a região da imagem onde  $h$  sobrepõe a  $f$ . A figura 13 ilustra o procedimento: assume-se que a origem de  $f$  fica à esquerda no topo da imagem e a origem de  $h$  no centro da imagem. Para qualquer valor  $(s, t)$  dentro de  $f_{(x,y)}$ , aplicando-se a equação [15], obtém-se um valor de  $c$ . Como  $s$  e  $t$  variam,  $h_{(x,y)}$  move-se ao redor da área da imagem, dando a função  $c_{(s, t)}$ . O máximo valor de  $c_{(s, t)}$  indica a posição de melhor comparação de  $h_{(x,y)}$  sobre  $f_{(x,y)}$ . Note-se que a precisão é perdida para valores de  $s$  e  $t$  próximos às extremidades de  $f_{(x,y)}$ , com um erro proporcional ao tamanho de  $h_{(x,y)}$ .

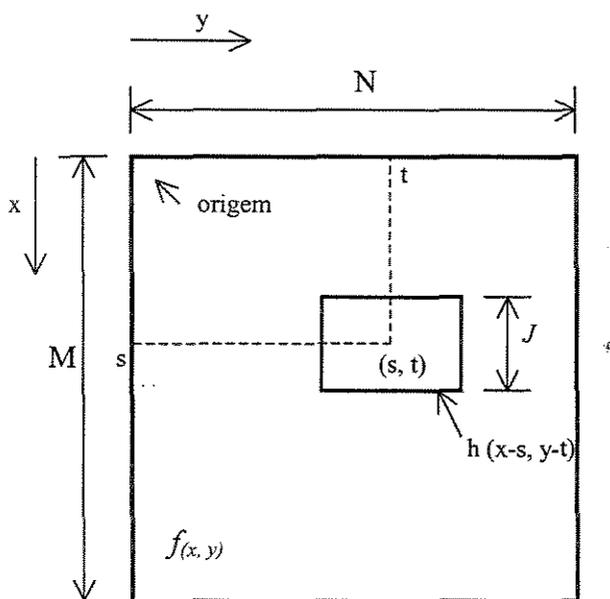


Figura 13 – Arranjo para obter a correlação de  $f_{(x,y)}$  e  $h_{(x,y)}$  no ponto  $(s, t)$ .

A função de correlação dada na equação [15] tem a desvantagem de ser sensível a mudanças de amplitude de  $f(x, y)$  e  $h(x, y)$ . Por exemplo, duplicando todos os valores de  $f(x, y)$  e  $h(x, y)$ , o valor de  $c_{(s, t)}$  também é duplicado, o que torna difícil interpretar uma comparação com outros resultados de similaridade. Um artifício frequentemente usado para contornar esse problema é executar a comparação com base no coeficiente de correlação normalizado ( $NCC$ ), o qual é definido como:

$$NCC(s, t) = \frac{\sum_x \sum_y [f(x, y) - \bar{f}(x, y)][h(x-s, y-t) - \bar{h}]}{\left\{ \sum_x \sum_y [f(x, y) - \bar{f}(x, y)]^2 \sum_x \sum_y [h(x-s, y-t) - \bar{h}]^2 \right\}^{1/2}} \quad [16]$$

onde  $\bar{h}$  é o valor médio de *pixels* em  $h(x, y)$  (computado só uma vez),  $\bar{f}(x, y)$  é o valor médio de  $f(x, y)$  na região coincidente com o local atual de  $h$  e a somatória é tomada sobre as coordenadas comum a ambos  $f$  e  $h$ . O  $NCC_{(s, t)}$  atinge valores de -1 a 1, independentemente das mudanças de escala na amplitude de  $f(x, y)$  e  $h(x, y)$  (GONZALEZ e WOODS, 2002). Se o  $NCC_{(s, t)}$  for igual a -1, isso indica que o molde encontrou o seu “negativo” (o termo “negativo” utilizado aqui é análogo ao negativo de um filme fotográfico).

O *Template Matching* tem sido muito aplicado em diversas áreas. Suas aplicações vão desde o reconhecimento de impressões digitais, alinhamento de superfícies, acompanhamento de objetos numa seqüência de imagens; reconhecimento de faces; detecção de alvos a partir de imagens de radar; diagnóstico de doenças a partir de imagens até identificação de estrelas. Isto demonstra que se trata de uma ferramenta muito poderosa no reconhecimento de padrões.

## Teoria da comunicação

As origens da moderna teoria da comunicação remontam a inúmeros trabalhos na área da comunicação elétrica. Significativo avanço foi conseguido com os trabalhos de Huntley sobre a quantidade de informação, nos quais ele define informação de uma mensagem ( $H$ ) em função do número de possíveis seqüências de símbolos que podem ser selecionados ( $s$ ) e do número de símbolos selecionados ( $n$ ), através da equação  $H = n \log(s)$ . Esta equação contém a origem do relacionamento da comunicação com a entropia. Gabor, em 1946, relacionou as diferentes variáveis da teoria da comunicação, fazendo analogia com o princípio de incerteza de Heisenberg e com os trabalhos de Wiener, um matemático preocupado com problemas de comunicação e controle (DIB, 1974).

Coube, porém, a Claude Shannon, em 1948, apresentar em *A Mathematical Theory of Communication* um estudo sistemático a partir de um modelo relativamente simples, com o qual demonstrou uma série de teoremas relativos ao processo de comunicação. Atribui-se a Warren Weaver a difusão das idéias de Shannon em outras áreas que não a engenharia, através do artigo *Recent contributions to the Mathematical Theory of Communication*, publicado em 1949.

A teoria da comunicação está relacionada com as leis matemáticas que regem a transmissão e o processamento da informação. Mais concretamente, ocupa-se da medição quantitativa da informação, da representação (codificação/decodificação) e da capacidade que os sistemas têm para transmitir e processar as informações, caracterizando um sistema de comunicação.

A transmissão de informação depende de sinais (por exemplo, letras e algarismos), em meio aos quais a fonte de informações pode realizar a sua escolha. Em geral, estes sinais são codificados na estação emissora, isto é, são transformados em outros sinais que se prestam para a transmissão.

Esta transformação, ou seja, esta relação exata entre dois tipos de sinais é conhecida como *codificação*; a transformação das letras do alfabeto em sinais do código Morse constitui um exemplo de codificação.

A informação transmitida é decodificada pelo receptor, sendo transmitida ao destinatário. A codificação pode referir-se tanto à transformação de voz ou de imagem em sinais elétricos ou eletromagnéticos como à cifração de mensagens para garantir-lhes privacidade.

A teoria da informação ocupa-se também das perturbações (ruído) na transmissão das informações. Estas perturbações podem estar localizadas em qualquer etapa do processo de comunicação. Uma das tarefas da informática consiste em determinar os métodos de codificação e decodificação e proteger as informações contra os fatores capazes de perturbá-las.

O conceito de informação utilizado neste trabalho refere-se apenas ao aspecto mensurável, o qual pode ser expresso mediante equações matemáticas. No jogo de dados, por exemplo, sempre se obtém apenas um dos seis números possíveis. Antes da jogada, todos os seis números possuem a mesma probabilidade ( $p$ ) de caírem, isto é,  $p = 1/6$ . Em cada jogada é eliminada uma incerteza do jogador e pode-se afirmar que cada jogada possui idêntico valor informativo. Esta observação pode ser generalizada:

*Informação é a eliminação mensurável da incerteza relativa à determinada situação.*

A eliminação da incerteza relativa a um fato é tanto maior quanto menor for a probabilidade ( $p$ ) deste fato ocorrer. Portanto, a medida do conteúdo informativo ( $I$ ) deve ser o inverso de  $p$ . Outra condição adicional para a medida da informação consiste em equiparar  $I$  a zero quando se tem certeza absoluta da ocorrência do fato, isto é, nos casos em que  $p = 1$ . Esta condição é preenchida pela aplicação da função logarítmica, uma vez que  $\log 1 = 0$ .

Vemos, portanto, que o conteúdo informativo mensurável de um evento pode ser definido por:

$$I = \log_2 (1/p) \quad [17]$$

Para esta definição, foi escolhido o logaritmo da base 2. Esta escolha corresponde à escolha da unidade da medida de informação em bits, o dígito binário. Desde que o número de estados possíveis seja  $2^N$  ou  $\log_2 2^N = N$ , é possível, com esta equação, calcular matematicamente o conteúdo informativo de cada lance do jogo de dados, que é 2,58 *bits*.

Esta definição de medida da informação pode ocorrer com idêntica probabilidade, de  $1/p$ , como é o caso no dado regular. Trata-se, porém, de uma simplificação; geralmente os diversos sinais e estados de uma fonte de informação não apresentam a mesma probabilidade. As letras do alfabeto constituem um exemplo de probabilidade desigual, visto que elas ocorrem com frequência (probabilidade) desigual num determinado texto.

No caso especial de probabilidades idênticas, a equação [17] pode ser transformada: quando  $n$  representa o número total de sinais possíveis, símbolos ou possíveis estados da fonte de informação, teremos que  $p = 1/n$ . No caso do dado,  $n = 6$ . Ou seja, o conteúdo informativo de uma fonte de informações corresponde ao logaritmo dual do número  $n$  de todos os estados possíveis da fonte de informações.

Este conceito quantitativo da informação não leva em consideração a importância da informação para o jogador. No jogo de dados, por exemplo, a valorização subjetiva dos números varia de um jogador para outro, de acordo com numerosos fatores (tipo de jogo, experiências anteriores, participantes do jogo).

## Modelo de comunicação de Shannon

Um conceito fundamental na teoria da informação é o de que a quantidade de informação contida numa mensagem é um valor matemático bem definido e mensurável. O termo quantidade não se refere ao volume de resultados, mas à probabilidade de que uma mensagem, dentro de um conjunto de mensagens possíveis, seja recebida. O significado de comunicação pode ser analisado através do modelo de Shannon e Weaver (1949).

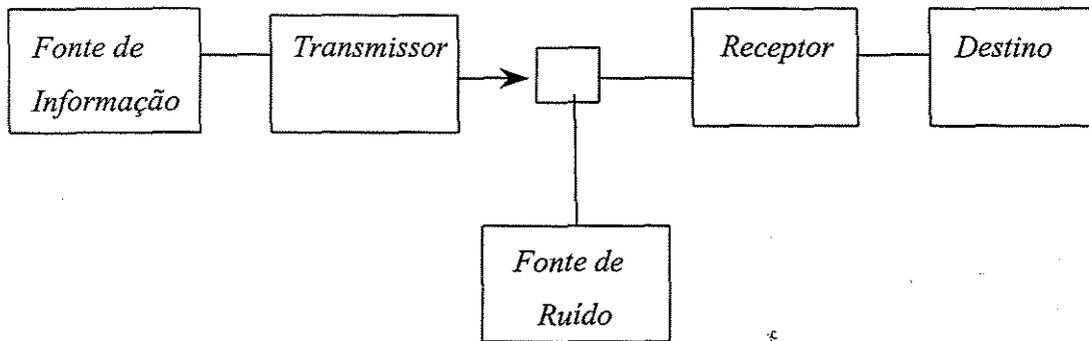


Figura 14 – Modelo de Comunicação de Shannon e Weaver.

Segundo Shannon e Weaver (1949), a fonte de informação produz a mensagem ou conjunto de mensagens a serem transmitidas. A mensagem pode corresponder: (a) a uma seqüência de letras, figuras e sons; (b) a uma função do tempo  $f(t)$  (telefonia ou rádio); (c) a uma função do tempo e de outras variáveis (televisão em branco e preto); (d) a três funções de diversas variáveis,  $f(x,y,t)$ ,  $g(x,y,t)$ ,  $h(x,y,t)$  (caso de televisão a cores), etc.

O transmissor transforma a mensagem em sinal, o qual é enviado ao receptor através de um canal adequado para a transmissão do sinal. No caso de telefonia, o canal correspondente é um fio por onde circula uma corrente elétrica variável. No caso do rádio, o sinal é uma onda eletromagnética. No caso dos pássaros é uma onda mecânica tridimensional longitudinal.

Ao receber a mensagem, o receptor, em geral, opera de modo inverso ao do transmissor, reconstruindo as mensagens originais, que são passadas ao destinatário (pessoa ou objeto para o qual a mensagem foi preparada). Durante a transmissão do sinal, diversos acontecimentos podem alterar o sinal original, constituindo o ruído.

Entre as principais questões ligadas ao processo de comunicação, Weaver (1949) assinala: (a) a medida da quantidade de informação transmitida e da capacidade do canal de comunicação; (b) as características de um processo de codificação eficiente; (c) as características do ruído e como este afeta a mensagem recebida pelo destinatário.

Ainda segundo o mesmo autor, o processo de comunicação parece envolver problemas em três níveis: técnico (quanto à exatidão com que os símbolos da comunicação podem ser transmitidos), semântico (quanto à precisão com que os símbolos transmitidos transportam os significados desejados) e eficiência do processo (quanto à eficiência com que os símbolos transmitidos podem afetar a conduta do modo desejado).

Para Shannon uma mensagem é muito informativa se a chance de sua ocorrência for pequena. Caso contrário, se uma mensagem é muito previsível, então a quantidade de informação é pequena. Para completar sua análise quantitativa do canal de comunicação, Shannon introduziu a *taxa de entropia*  $S$ , uma quantidade que mede a taxa de produção de informação de uma fonte e também a capacidade de informação que esta transporta, chamada de capacidade do canal de comunicação.

$$S = \sum_{i=1}^n p_i I_i \quad [18]$$

Shannon mostrou que, se a taxa de entropia (eq.18), a quantidade de informação que você deseja transmitir, excede a capacidade de canal, é inevitável a ocorrência de erros incorrigíveis na transmissão. Ele também mostrou que, se a taxa de entropia do remetente for menor que a capacidade do canal, então há uma maneira de codificar a informação de forma que ela seja recebida sem erros. Isso ocorre até mesmo se o canal distorcer a mensagem durante transmissão.

Shannon adaptou sua teoria para analisar o ser humano no idioma escrito e mostrou que ele é bastante redundante, pois usa mais símbolos e palavras que o necessário para a transmissão de mensagens. Presumivelmente, o ser humano usa esta redundância para melhorar sua habilidade de reconhecer mensagens confiantemente e comunicar tipos diferentes de informações. A transmissão de informações depende do conjunto de sinais onde a fonte de informações realiza a sua escolha.

As pesquisas realizadas com textos sistematicamente mutilados revelaram, por exemplo, que a língua alemã escrita só aproveita em média 1,3 bit por letra. Teoricamente, o conteúdo informativo médio de cada uma das 26 letras do alfabeto é de  $\log_2 26 = 4,7$  bits; portanto, a redundância média é de  $(4,7 - 1,3)$  bits = 3,4 bits por letra.

Entretanto, este desperdício aparentemente supérfluo de sinais não constitui senão um aspecto parcial da redundância. As vantagens da redundância manifestam-se diante das *perturbações no canal de transmissão*, como, por exemplo, nas ligações telefônicas deficientes, na radiorecepção perturbada pela estática, na escrita pouco legível, na comunicação sonora animal em meio aos obstáculos da mata. Em tais casos, é graças à redundância da linguagem que um determinado texto pode ser decifrado com base em uma fração de sinais identificáveis. De um modo geral, a teoria da comunicação mostra que a transmissão de uma informação é tanto mais protegida contra distúrbios quanto maior for a redundância contida na codificação.

Maria Luisa da Silva (2001) utilizou a teoria da comunicação de Shannon para caracterizar a estrutura e organização do canto da espécie *Turdus rufiventris*. Considerando que também há informação na frequência das notas e na ritmicidade do canto (o que implica afirmar que há informação quando um indivíduo emite somente uma nota), mediu-se a variação e avaliou-se a funcionalidade dos sinais de comunicação, criando um índice que integra os valores de entropia e ritmicidade do canto individual para amenizar a heterogeneidade da amostra e elucidar a magnitude da variação.

## Bioacústica e comunicação sonora

A Bioacústica é um campo científico moderno que se beneficia da tecnologia de gravações e análise dos sons por meio da qual se pode facilmente conservar e definir, em termos de parâmetros físicos, o próprio sinal de comunicação, contribui para a ornitologia de campo, esclarecendo vários aspectos da comunicação sonora em aves (VIELLIARD 1987).

Embora recentes, as contribuições da bioacústica para a ornitologia têm sido fundamentais, principalmente quanto ao estabelecimento dos conceitos básicos da evolução, da biogeografia, da taxonomia, da etologia e da biologia da conservação.

A divergência no estabelecimento de critérios na definição da comunicação alimenta a polêmica sobre como avaliar uma determinada situação e verificar se os indivíduos envolvidos estão ou não se comunicando. Os exemplos práticos mostram como é difícil se chegar a uma definição de comunicação que não seja ambígua e que possa ser facilmente aplicada a todos os casos (HADDAD 1995).

Deve haver no mínimo três elementos no processo de comunicação: fonte, mensagem, destinatário. A fonte pode corresponder a um indivíduo (falando, escrevendo, desenhando, gesticulando), a mensagem pode ser na forma escrita, sonora ou conter outros sinais que podem ser interpretados de forma significativa, e o destinatário pode ser um indivíduo escutando, prestando atenção ou lendo (SCHRAMM 1965).

Weaver (1949) assinala que a palavra comunicação inclui todos os procedimentos pelos quais uma mente pode afetar outra. A comunicação visa, muitas vezes, alcançar resultados predeterminados. Quando isso não ocorre, diz-se que a comunicação falhou. Mas para avaliar se a comunicação ocorreu ou não, é necessário definir em termos objetivos e mensuráveis qual o resultado que se pretende alcançar com a comunicação. Devem-se estabelecer, em termos comportamentais, os critérios (respostas desejadas) que irão indicar

se ocorreu a mudança de comportamento desejada e se, portanto, a comunicação foi eficiente.

Como em qualquer sistema de comunicação, os sinais acústicos da comunicação animal precisam manter a informação que carregam ao longo das três etapas da comunicação: emissão, transmissão e recepção. O sinal emitido deve, portanto, chegar ao receptor de maneira que sua função seja captada e identificada.

A emissão e a recepção de um sinal são realizadas através de capacidades fisiológicas e anatômicas especializadas que são adequadas à modalidade do sinal e ao meio onde vive o animal em questão.

A base da comunicação animal corresponde ao conceito de espécie biológica. Assim, utilizamos a palavra canto para designar o sinal de comunicação vocal que tem como função biológica primordial o reconhecimento específico (VIELLIARD 1987). Dessa forma, a noção de especificidade do canto das aves já foi observada por Florence em 1830, no diário de bordo da expedição Langsdorff pelo Brasil (VIELLIARD 1993) e por Altum (1868), que reconhecem o caráter específico do canto, embora o definam como um chamado de acasalamento, independente do tipo de som.

O canto é único para cada espécie sendo que algumas espécies possuem dialetos regionais. É temporariamente delimitado pelas estações do ano e geralmente executado apenas pelos machos. É usado, principalmente, para anunciar e delimitar o território e atrair a fêmea. O canto pode ser uma repetição sucessiva de piados ou padrões complexos de unidades contínuas de som, as notas, formando longos e repetidos temas ou frases.

## O Sabiá-laranjeira *Turdus rufiventris*

O Sabiá-laranjeira *Turdus rufiventris* é uma espécie de pássaro conspícua e muito comum da fauna Neotropical. Vive na mata (floresta mesófila, floresta secundária), bordas de floresta ombrófila, parques, quintais e até mesmo no centro das cidades quando há algum ajardinamento (SILVA 1997).

Sua aparência física, ventre cor laranja, o bico e a região orbicular amarelo-vivo contrastam com a garganta clara estriada de preto, não é tão extraordinária quanto seu canto, melodioso, variado e principal responsável pela sua popularidade no Brasil. Seu canto é composto de apitos e trinados de altura média (alcance de frequência entre 1 a 4 kHz, ver fig.15). As frases são articuladas em longas seqüências de notas regularmente espaçadas e emitidas sucessivamente. Alguns indivíduos costumam emitir notas de curta duração amplamente moduladas no final da frase (SILVA, 1997).

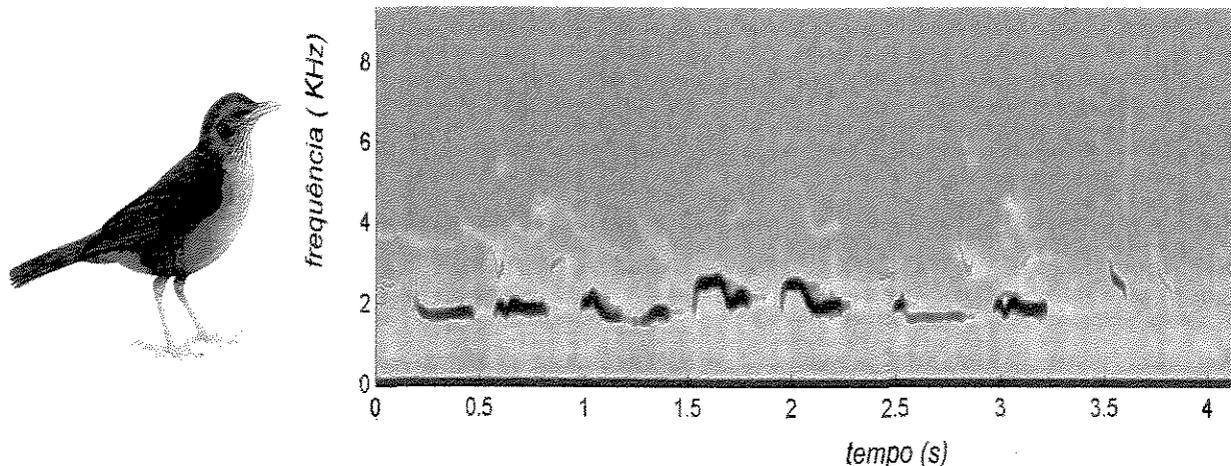


Figura 15 – O Sabiá-laranjeira e o espectrograma de um canto típico da espécie.

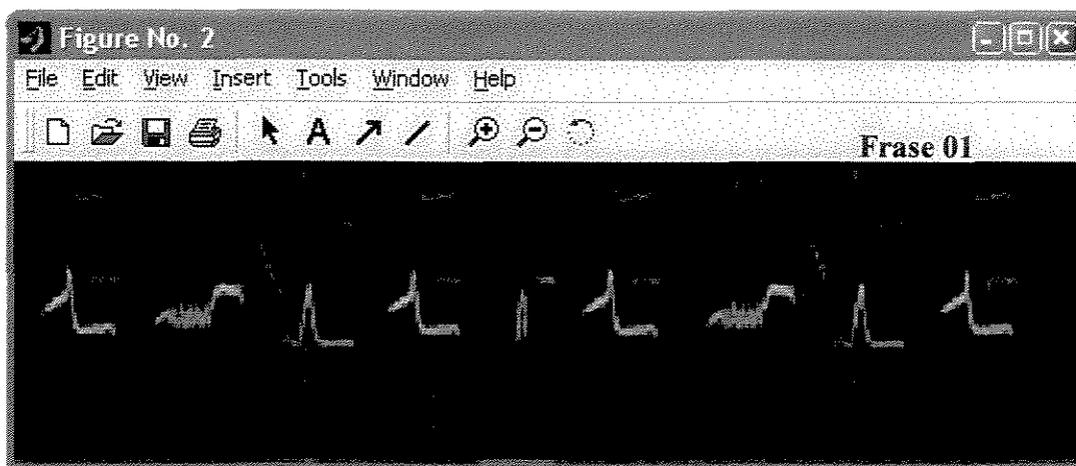
O Sabiá-laranjeira é abundante e difundido, permitindo o estudo de muitos indivíduos de diferentes localidades (SILVA *et al.* 2000). Ocorre no Brasil Oriental e Central, do Maranhão até a Paraíba, Rio Grande do Sul e Mato Grosso, além de Uruguai, Paraguai e regiões limítrofes da Bolívia e Argentina (SCHAUENSEE 1970).

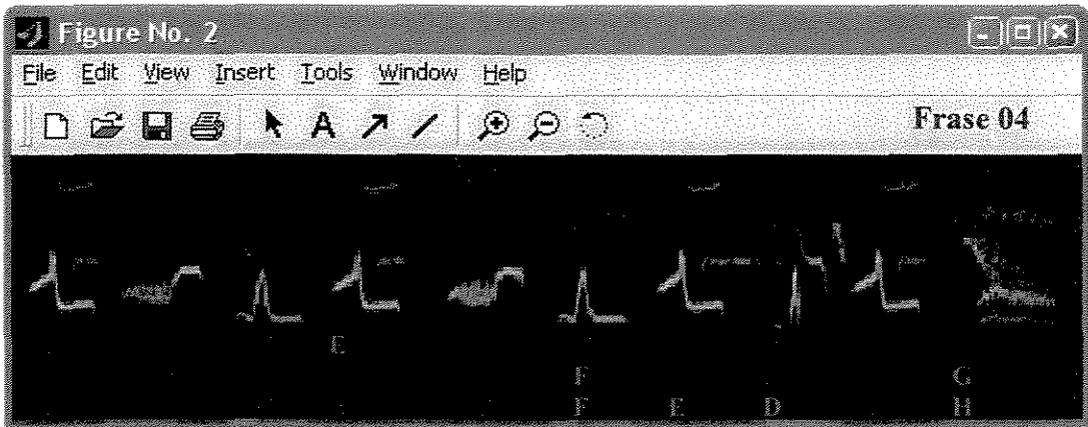
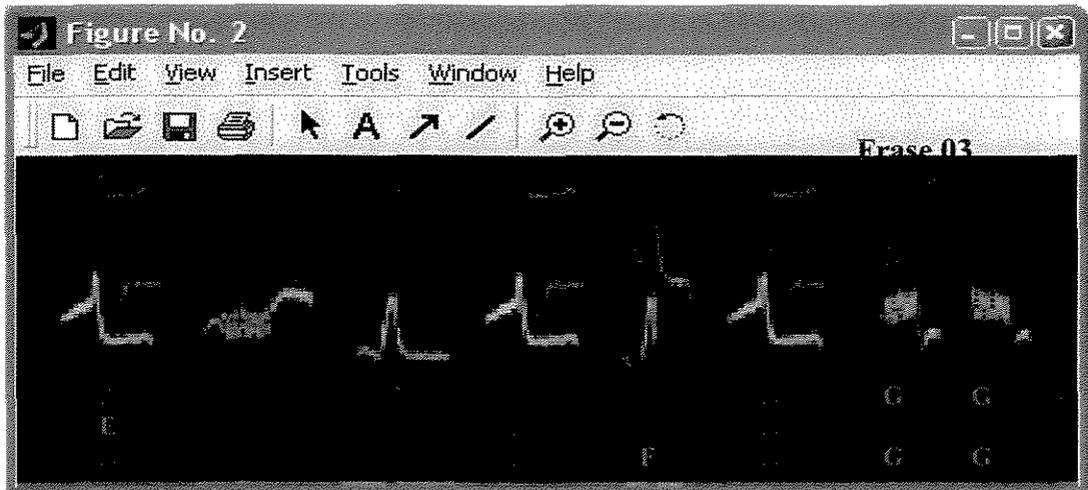
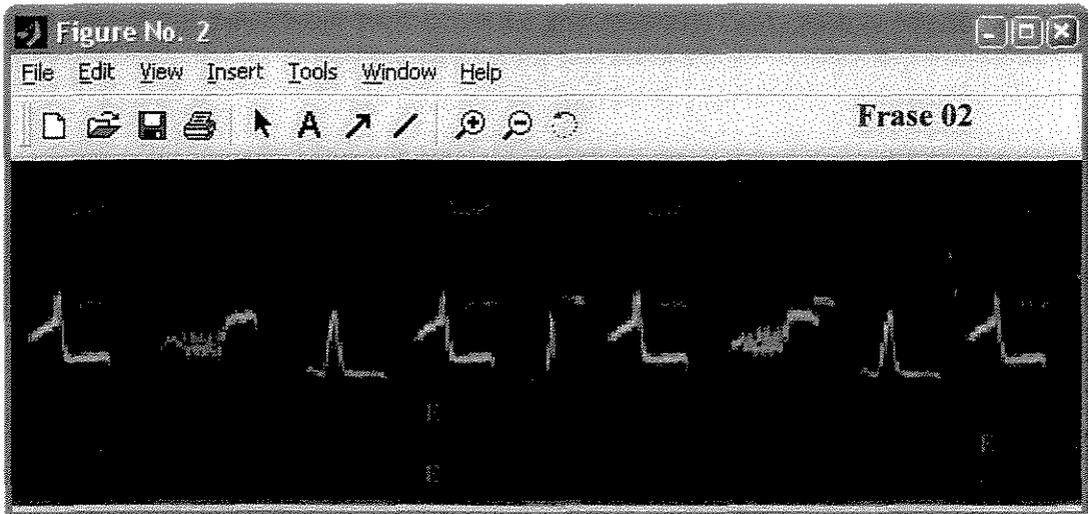
## RESULTADOS

Levando-se em consideração a análise do reconhecimento de três indivíduos, foram realizados diversos testes sobre a viabilidade do SITM para classificação de notas. A rotina desenvolvida também determina a entropia de Shannon da seqüência de notas, mas não da ritmicidade do canto. Os resultados da entropia não são apresentados por serem conseqüências diretas da classificação. Se a classificação estiver correta os cálculos da Entropia também estarão.

Para facilitar a interpretação, os resultados seguem a seguinte legenda: as letras azuis expressam a classificação das notas obtida por um observador humano treinado usando os métodos tradicionais de comparação; as vermelhas representam os resultados obtidos pelo SITM e na cor verde estão as notas em que os resultados do SITM diferem do método usual.

No ASN estão armazenados cinco frases co sete tipos de nota do indivíduo 08. A primeira e a segunda frases têm a mesma seqüência de nove notas. A terceira frase possui oito notas, a quarta dez notas e a quinta frase dezesseis notas. Foram realizados três testes, conforme o esquema que se segue:





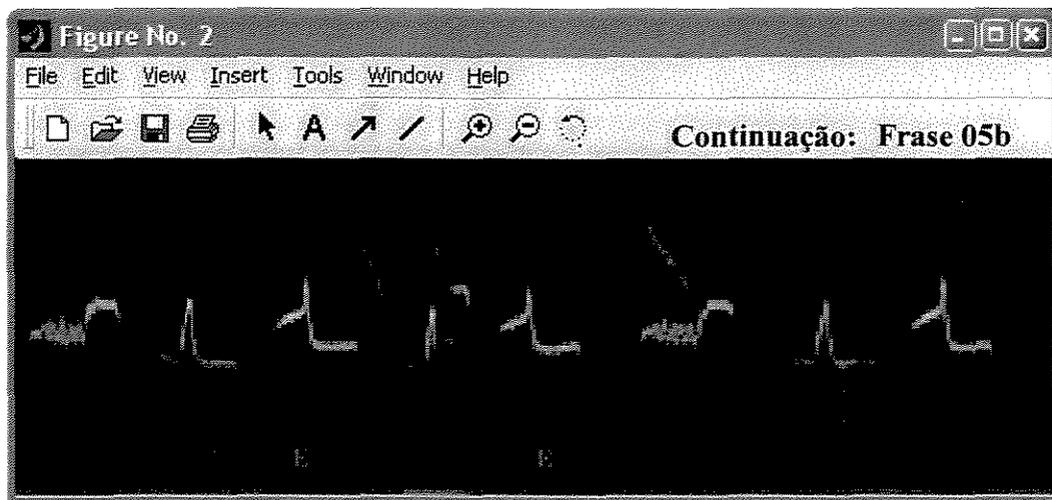
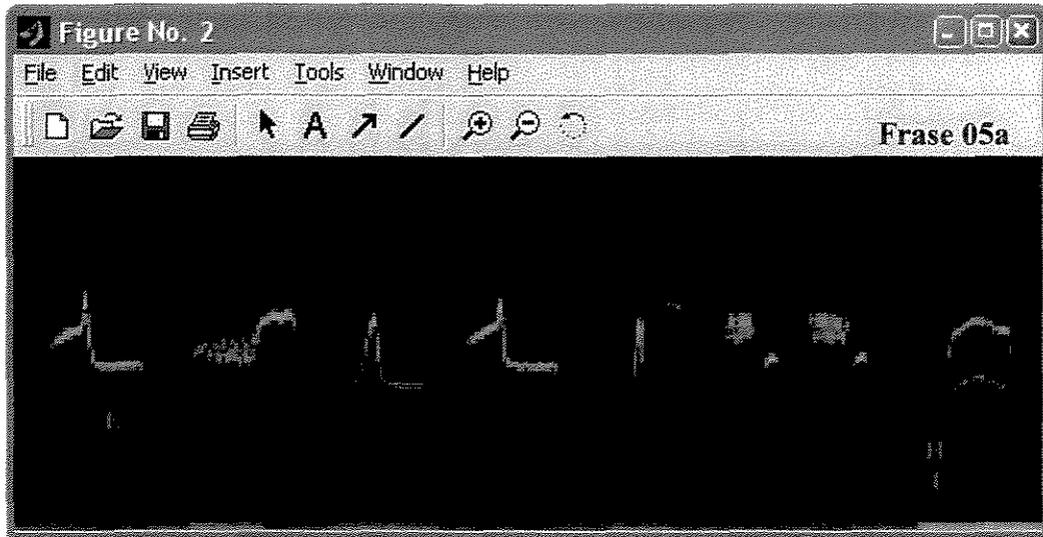


Figura 16 – Frases do indivíduo 08.

Em cada teste foi obtida uma certa quantidade de acertos, a quantidade total de acertos determina uma média de sucesso  $x$  dentro de um certo número de amostras  $n$ . Segundo Inferência Estatística o IC para proporções é:

$$P(p_o - z_\alpha \sigma_p \leq p \leq p_o + z_\alpha \sigma_p) = 1 - \alpha \quad [19]$$

$$\text{Sendo } p_o = x/n; \quad [20]$$

$$\sigma_p = (p_o q_o / n)^{1/2} \quad [21]$$

e  $Z_\alpha$  um valor tabelado.

Assim, adotando um nível de incerteza de  $\alpha = 4\%$ , os resultados demonstraram que o SITM, com uma confiança de 96%, reconhece corretamente de 77% a 97% das notas do indivíduo 08.

Do indivíduo 15, existem apenas duas frases gravadas. A primeira com 40 notas e a segunda com 53. Ao analisar esse canto, o SITM reconheceu três tipos de notas diferentes enquanto a análise visual distingue 7 tipos. Foram feitos cinco testes como indica o esquema abaixo:

<b>Frase 1a</b>	<b>Frase 1b</b>	<b>Frase 1c</b>	<b>Frase 1d</b>
A B C D A	B E F E F	A B E F G	A B E F A
A B C D A	B E F E F	A B E F G	A B E F A
A B C D A	B E F E F	A B E F G	A B E F A
A B C D A	B E F E F	A B E F G	A B E F A
A B C D A	B E F E G	A B E F H	A B E F A
A B C D A	B E F E F	A B E F G	A B E F A
<b>Frase 1e</b>	<b>Frase 1f</b>	<b>Frase 1g</b>	<b>Frase 1h</b>
C D A B E	F E F A B	C D A B E	F A B E F
C D A B G	B E B A B	C D A B E	F A B G B
H D A B G	F E F A B	H D A B E	F A B E F
C D A B G	B E B A B	C D A B E	F A B G B
I D A B E	G E G A B	I D A B J	G A B J K
C D A B G	F E F A B	H D A B E	F A B E F
<b>Frase 2a</b>	<b>Frase 2b</b>	<b>Frase 2c</b>	<b>Frase 2d</b>
F E F A B	E F A B E	F A B C D	A B E F E
F H B A B	E F A B H	F A B I D	A B J F E
F E B A F	I F A F H	F A B C D	A F J F K
F H B A B	E F A F H	F A B I D	A B E F E
L H G A B	E F A G E	F A G I D	A G J F E
F H B A B	E F A B H	F A B I D	A F J F B
<b>Frase 2e</b>	<b>Frase 2f</b>	<b>Frase 2g</b>	<b>Frase 2h</b>
F A B C D	A B E F E	F D A B E	F A B E F A
F A B K D	A B L M N	M D A B L	O A B P O A
B A B L D	A B L B E	F D A B E	M A B E N A
F A B K D	A B L F E	F D A B E	F A B E F A
F A B M D	A B N F E	F D A G O	P A B P P A
B A B L D	A B L F E	F D A B J	B A B L F A

Frase 2i

B C D A B  
 Q R D A Q  
 O K D A B  
 M K D A B  
 B K D A B  
 M K D A B

Frase 2j

E F E F A B C  
 L B L M S M K  
 L F L M A M R  
 E B L M A M K  
 L P O P A M K  
 L B L M A M K

Utilizando o mesmo nível de incerteza, os resultados demonstraram que o SITM, com uma confiança de 96%, reconhece corretamente de 64% a 82% das notas do indivíduo 15.

O indivíduo 26 possui 4 frases todas elas com 10 notas; no total são 11 tipos de notas.

Frase 01

A B C D E F G H H I  
 A B C D E F G H H I  
 A B C D E F G H H I  
 A B C D E F G H H I  
 A B C D E F G H H I  
 A B C D E F G H H I

Frase 02

A B C D E F G H H I  
 A B C D E F G H H I  
 A B C D E F G H H I  
 A B C D E F G H H I  
 J B C D E F G H H I  
 A B C D E F G H H I

Frase 03

A B C J D E F G H H  
 A B C J D E F G H H  
 A B K J D E F G H H  
 J B C K D E F G H H  
 J B C K D E F G H H  
 A B C J D E F G H H

Frase 04

I K C E F G J D E F  
 I L C E F G K D E F  
 I K H E F G K D E F  
 I L M E F G K D E F  
 I L M E F G J D E F  
 I L C E F G K D E F

Com o indivíduo 26, o SITM apresentou uma confiança de 96% e reconhece corretamente de 74% a 97% suas 11 notas.

## DISCUSSÃO

Esquemas de classificação podem diferir entre si, segundo a variável contextual, como, por exemplo: (1) *habitat* e efeitos associados à propagação do som; (2) a identidade do cantor e o mecanismo da produção do som; (3) a maneira como um receptor da espécie categoriza o som; ou (4) o comportamento funcional desses sons. Não há nenhuma razão para procurar encontrar um esquema único de classificação para um conjunto de sons que seja válido para todos esses diferentes contextos. Portanto, o que está sendo avaliado aqui é somente a capacidade de reconhecimento de notas de uma espécie comparado com o método tradicional.

A vantagem do SITM em relação a um observador humano treinado é o processamento simultâneo de um grande número de dados e o armazenamento das informações que proporciona facilidades para novas análises. Os resultados apresentaram intervalos de confiança significativos, entretanto, devido ao caráter inédito da pesquisa, não é suficiente para que o SITM possa ser utilizado em laboratório.

No caso da análise do indivíduo 15, por exemplo, o equívoco ocorrido foi ocasionado pelo pequeno dimensionamento das imagens quando selecionadas com o *mouse*. Para contornar esse problema, o canto foi dividido em várias partes, cada uma delas com uma média de cinco notas, a exemplo do que foi feito com a quinta frase do indivíduo 08. Esse artifício futuramente pode ser substituído por uma barra de rolagem (ou *slider*, como denominam os usuários familiarizados com a interface gráfica do Matlab).

Foi observado também que, conforme a base de dados aumenta, o programa se “confunde”, ocasionando erros incorrigíveis até o momento. Na maioria das vezes, o erro não demonstra ser tão grave, entretanto, como é o caso, ele se propaga por toda a amostra. Esse problema pode ser contornado com ajustes no algoritmo e com uma melhor definição no grau de NCC. Ainda assim, é provável que haja uma propagação de erros considerável, devido a erros causais gerados pelo observador.

## CONCLUSÃO

O processamento de imagem tem uma vantagem em relação ao processamento de som no que diz respeito ao reconhecimento de padrões. Além de apresentar um formato mais “leve”, o que implica um processamento mais rápido, a representação visual dos sons complementa a análise auditiva, tornando viável o reconhecimento da imagem que representa um padrão sonoro.

A classificação semi-automática de sons através de imagens aplicadas à bioacústica, especificamente ao canto da espécie *T. rufiventris*, tem uma limitação óbvia ao levar muito em conta a intensidade de brilho, além de erros causais de manuseio do *mouse*. Os resultados de classificação das notas apresentaram intervalos de confiança significativos, entretanto não o suficiente para que o SITM seja utilizado em laboratório. Mesmo assim, ele ainda pode se tornar um método muito eficaz, já que o que prejudicou seriamente os resultados foram propagações de erros iniciais. Estes erros podem ser corrigidos com a otimização do algoritmo de classificação e com a utilização de novos recursos. A eficácia do método pode ser melhorada com o acréscimo de mais uma análise na medida de similaridade, como detalhes na forma de onda da nota através da sua representação gráfica tridimensional. Outra alternativa, seria explorar mais a fundo ferramentas usuais em PDI como a morfologia matemática, para análises de forma e contorno. Acredita-se que a implementação de uma rede neural de Hopfield elimine completamente a propagação de erros.

Juntamente com ferramentas atuais como *Wavellets* e *redes neurais artificiais* o SITM pode ser uma ferramenta essencial para compor um *software* científico moderno na área de Bioacústica. Espera-se que a integração desses e/ou de novos conceitos tenda a se transformar numa teoria sólida, de modo que se estabeleçam critérios para avaliar tais sinais com o mínimo possível de distorções conceituais e que se aproxime ao máximo o conhecimento humano a natureza da comunicação sonora animal.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALTUM, B. (1868). – *Der Vögel und sein Leben*. W. Niemann, Münster, 196 pp.
- BEECHER, M. D. (1988). - Spectrographic analysis of animal vocalizations: Implications of the “uncertainty principle. *Bioacoustics* 1(1): 187-207.
- BERLIN, B. & O’NEILL, J. P. (1981). - The pervasiveness of onomatopoeia in Aquaruna and Uambisa bird names. *Journal of Ethnobiology* 1(2): 238-261.
- BREGMAN, A S. (1990). - *Auditory Scene Analysis: The Perceptual Organization of sound*. The MIT Press, Cambridge, Massachusetts.
- BRIGHAM, E. O. (1988). – *The fast Fourier transform and its applications*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- CARDOSO, A J. & VIELLIARD, J. (1990). - Vocalizações de anfíbios anuros de um ambiente aberto, em Cruzeiro do Sul, Estado do Acre. *Revta. Bras. Biol.* 50(1): 229-242.
- CLARK, C. W., MARLER, P. & BEAMAN, K. (1987). - Quantitative analysis of animal vocal phonology; an application to Swamp Sparrow song. *Ethology* 76: 101-115.
- COHEN, L. (1995). - *Time-frequency analysis: theory and applications*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- COSTA, L. F., CESAR JR, R M. (2001). - *Shape analysis and classification: theory and practice*. CRC Press, Inc. Boca Raton, FL, USA.
- DEUTSCH, D. (1981). – Response to 'Auditory Perception in Music Teaching and Learning' by J.J. Heller and W.C. Campbell. In *Proceedings of the National Symposium on the Applications of Psychology to the Teaching and Learning of Music*, Reston, Virginia.
- DIB, C. Z. (1974). - *Tecnologia da Educação e sua aplicação à aprendizagem de física*. Ed. Pioneira, São Paulo.
- HADDAD, C.F.B. (1995). – Comunicação em Anuros (Amphibia). *Anais de Etologia* 13: 116-132.
- HANSELMAN, D. e LITTLEFIELD, B. (1999). – Matlab® 5 Versão do Estudante – *Guia do Usuário*. MACKRON Books, São Paulo.
- FU, K. S. (1982). - *Application of Pattern Recognition*. CRC Press, Cleveland.

- GONZALEZ, R. C. & WOODS, R. E. (2002). - *Digital Image Processing*, 2nd edition. Prentice Hall, Upper Saddle River, NJ
- KETTLE, R. & VIELLIARD, J. (1991). - Documentation standards for wildlife sound recordings. *Bioacoustics* 3: 235-238.
- MORETTIN, L.G. (1995). - *Estatística básica, vol.2: Inferência*. MACKRON Books, São Paulo.
- OPPENHEIM, A. V. & SCHAFER, R. V. (1975). - *Digital Signal Processing*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- SCHRAMM, W. (1965). - How Communication Works. In W. Schramm (org.), *The Process and Effects of Mass Communication*. Urbana, Illinois Press: Univ.Ill. Press, 1954, 1965.
- SCHALKOFF, R. (1992). - *Pattern Recognition*. Statistical, Structural and Neural Approaches. John Wiley, New York.
- SCHAUENSEE, M. R. de (1970). - A Guide to the Birds of South America. The Academy of Natural Science of Philadelphia, Wynnewood, Pennsylvania.
- SHANNON, C. E. (1948). - A Mathematical Theory of Communication. *The Bell System Technical Journal* 27: 379-423, 623-656, July, October.
- SHANNON, C. E. & WEAVER, W. (1949). - *The Mathematical theory of communication*. The Univ. Illinois Press, Urbana, Illinois Press: Univ.Ill. Press.
- SILVA, M. L. (1997). - Descrição do repertório vocal do Sabiá-laranjeira *Turdus rufiventris* (Aves, Passeriformes, Turdinae). Dissertação de Mestrado em Neurociências e Comportamento, Universidade de São Paulo.
- SILVA, M. L. (2001). - Estrutura e organização de sinais de comunicação complexos: o caso do Sabiá-laranjeira *Turdus rufiventris* (Aves, Passeriformes, Turdinae). Tese de Doutorado, USP, São Paulo.
- SILVA, M. L., PIQUEIRA, J.R.C. & VIELLIARD, J. (2000). - Using Shannon entropy on measuring the individual variability in the Roufus-bellied Thrush *Turdus rufiventris* vocal communication. *J. theor. Biol.* 207: 57-64.
- TCHERNICHOVSKI, O., NOTTEBOHM, F., HO, C. E., PESARAN, B. & MITRA, P. P. (2000). - A Procedure For An Automated Measurement of Song Similarity. *Animal Behaviour* 59: 1167-1176.

VIELLIARD, J. M. E. (1987). - Uso da bioacústica na observação das aves. In: *II Encontro Nac. Anilhad. Aves*, Rio de Janeiro, pp.98-121..

VIELLIARD, J. M. E. (1993). - *A Zoophonia de Hercule Florence*. UFMT, Cuiabá.

VIELLIARD, J. M. E. (2000). - Estado atual das pesquisas em bioacústica e sua contribuição para o estudo e a proteção das aves do Brasil. *A Ornitologia no Brasil*, UERJ, pp. 287-301.

WEAVER, W. (1949). - Recent contributions to the mathematical theory of communication. Pp. 1-29 in C.E. Shannon & W. Weaver. *The Matematical Theory of communication*. University of Illinois Press.

## APÊNDICE

### Pseudocódigo Matlab do SITM

```
001 - % Teste
002 - clear
003 - erro = 0;
004 -
005 - %-----%
006 -
007 - % O arquivo deve ser Mono, 44,100 Hz, 16 bit PCM wave file.
008 -
009 - canto = wavread ('08002.wav');
010 - specgram (canto, 1024, 2, 1024); % specgram(a,nfft,fs>window)
011 - colormap (gray);
012 - fig = imcrop;
013 -
014 - % Tratamento da imagem
015 -
016 - se = strel ('ball',1, -8);
017 - E = imerode (flipud (fig), se);
018 - [m,n] = size(E);
019 - figure ('Units', 'pixels', 'Position', [100 100 m n]);
020 - canto = image (E);
021 - colormap (gray);
022 - set (gca, 'Position', [0 0 1 1]);
023 - axis off;
024 - axis image;
025 -
026 - %-----%
027 -
028 - eval('load -mat teste.mat', 'erro = 1;');
029 -
030 - if erro == 1
031 -     'nao existe base de dados' % executa sem base
032 -     semBase;
033 -
034 -     tamanhoDaFrase = length(frase)/3;
035 -     contP = 1;
036 -     contF = 0;
037 -     indiceDaFrase = 1;
038 -     contFrase1 = 1;8
039 -     while contFrase1 < tamanhoDaFrase+1
040 -         indiceDaFrase = (contFrase1-1)*3+1;
041 -         if frase(indiceDaFrase) ~= char(0)
042 -             elementoDaFrase(1) = frase(indiceDaFrase);
043 -             elementoDaFrase(2) = frase(indiceDaFrase+1);
044 -             elementoDaFrase(3) = frase(indiceDaFrase+2);
045 -             frase(indiceDaFrase) = char(0);
046 -
047 -             contF = contF + 1;
048 -             p(contP).nota(1) = elementoDaFrase(1);
049 -             p(contP).nota(2) = elementoDaFrase(2);
050 -             p(contP).nota(3) = elementoDaFrase(3);
050 -             for contFrase2 = contFrase1+1:tamanhoDaFrase;
```

```

051 -             indiceDaFrase = (contFrase2-1)*3+1;
052 -             if elementoDaFrase(1) == frase(indiceDaFrase)
053 -             if elementoDaFrase(2) == frase(indiceDaFrase+1)
054 -             if elementoDaFrase(3) == frase(indiceDaFrase+2)
055 -                 frase(indiceDaFrase) = char(0);
056 -                 contF = contF + 1;
057 -                 end
058 -             end
059 -         end
060 -     end
061 -     p(contP).f = contF;
062 -     contP = contP + 1;
063 -     contF = 0;
064 - end
065 -     contFrase1 = contFrase1 + 1;
066 - end
067 -
068 - else
069 - 'carregando base de dados'
070 - comBase;                % executa comBase
071 -
072 - tamanhoDaFrase = length(frase_atual)/3;
073 - contP = length(p)+1;
074 - contF = 0;
075 - indiceDaFrase = 1;
076 - contFrase1 = 1;
077 - while contFrase1 < tamanhoDaFrase+1
078 -     indiceDaFrase = (contFrase1-1)*3+1;
079 -     if frase_atual(indiceDaFrase) ~= char(0)
080 -
081 -         naBase = 0;
080 -         for contFrase2 = 1:length(p)
081 -             if p(contFrase2).nota(1) == frase_atual(indiceDaFrase)
082 -                 if p(contFrase2).nota(2) == frase_atual(indiceDaFrase+1)
083 -                     if p(contFrase2).nota(3) == frase_atual(indiceDaFrase+2)
084 -                         frase_atual(indiceDaFrase) = char(0);
085 -                         p(contFrase2).f = p(contFrase2).f + 1;
086 -                         naBase = 1;
087 -                     end
088 -                 end
089 -             end
090 -         end
091 -         if naBase == 0
092 -
093 -             elementoDaFrase(1) = frase_atual(indiceDaFrase);
094 -             elementoDaFrase(2) = frase_atual(indiceDaFrase+1);
095 -             elementoDaFrase(3) = frase_atual(indiceDaFrase+2);
096 -             frase_atual(indiceDaFrase) = char(0);
097 -             contF = contF + 1;
098 -             p(contP).nota(1) = elementoDaFrase(1);
099 -             p(contP).nota(2) = elementoDaFrase(2);
100 -             p(contP).nota(3) = elementoDaFrase(3);

```

```

101 -   for contFrase2 = contFrase1+1:tamanhoDaFrase
102 -       indiceDaFrase = (contFrase2-1)*3+1;
103 -       if elementoDaFrase(1) == frase_atual(indiceDaFrase)
104 -           if elementoDaFrase(2) == frase_atual(indiceDaFrase+1)
105 -               if elementoDaFrase(3) == frase_atual(indiceDaFrase+2)
106 -                   frase_atual(indiceDaFrase) = char(0);
107 -                   contF = contF + 1;
108 -               end
109 -           end
110 -       end
111 -   end
112 -   p(contP).f = contF;
113 -   contP = contP + 1;
114 -   contF = 0;
115 - end
116 - end
117 - contFrase1 = contFrase1 + 1;
118 - end
119 - end
120 -
121 - %-----%
122 -
123 - % Entropia de Shannon
124 -
125 - totalDeNotas = 0;
126 - for contadorDoNumeroDeNotas = 1:length(p)
127 -     totalDeNotas = totalDeNotas + p(contadorDoNumeroDeNotas).f;
128 - end
129 - for contadorDoNumeroDeDi = 1:length(p)
130 -     di(contadorDoNumeroDeDi) = totalDeNotas/p(contadorDoNumeroDeDi).f;
131 -     i2(contadorDoNumeroDeDi) = p(contadorDoNumeroDeDi).f/totalDeNotas;
132 - end
133 -
134 - for contadorDoNumeroDeIi = 1:length(p)
135 -     ii(contadorDoNumeroDeIi) = log2(di(contadorDoNumeroDeIi));
136 - end
137 -
138 - s = 0;
139 - for contadorDoNumeroDesi = 1:length(p)
140 -     s = s + di2(contadorDoNumeroDesi)*ii(contadorDoNumeroDesi);
141 - end
142 - 'Entropia de Shannon:'
143 - s
144 - save teste notas p;
145 -

```

Sub-rotina semBase

```
001 - % SEM BASE
002 - % canto = imread('Figural1.jpg');
003 - % canto = rgb2gray(canto); % Transforma uma imagem RGB em PB
004 -
005 - n=input('numero de notas:');
006 -
007 - % para ocorrer a comparação deve existir pelo menos duas notas
008 -
009 - nota1 = imcrop;
010 - nota2 = imcrop; % Seleciona a nota desejada com o mouse
011 -
012 - matriz = padronizaMatrizes(nota1,nota2);
013 - i = 3;
014 - while (i < n + 1)
015 - nota = imcrop;
016 - matriz = padronizaMatrizes(matriz, nota);
017 - i = i + 1;
018 - end
019 -
020 - grau(1,1)=0;
021 - for i = 1:(n - 1)
022 -     for j = (i + 1):n
023 -         % escolha do template
024 -         ncc = normxcorr2(matriz(:,:,j), matriz(:,:,i));
025 -         grau(i,j) = max(ncc(:))*100;
026 -     end
027 - end
028 - grau(j,j) = 0;
029 -
030 - % grau
031 -
032 - padrao = 1;
033 - for i = 1: length(grau)
034 -     sequencia(i) = 0;
035 - end
036 - padraoUsado = 0;
037 -
038 - for i = 1:length(grau)
039 -     for j = 1:length(grau)
040 -         if ((grau(i,j) > 76)|(grau(i,j)==0))&(sequencia(j) == 0)
041 -             seqüência (j) = padrao;
042 -             padraoUsado = 1;
043 -         end
044 -     end
045 -     if padraoUsado == 1
046 -         padrao = padrao + 1;
047 -         padraoUsado = 0;
048 -     end
049 - end
050 -
```

```

051 - j = 1;
052 - i = 1;
053 - while j < length(sequencia) + 1
054 -     if sequencia(j) < 26
055 -         frase(i) = char(sequencia(j)+64);
056 -         frase(i+1)= ' ';
057 -         frase(i+2)= ' ';
058 -     else
059 -         frase(i) = char(fix(sequencia(j)/26));
060 -         frase(i+1)= char(mod(sequencia(j)/26));
061 -         frase(i+2)= ' ';
062 -     end
063 -     i = i + 3;
064 -     j = j + 1;
065 - end
066 - frase
067 - % seqüência baseado no vetor de entrada p/ determinar elementos
068 - % repetidos existem
069 -
070 - elementos = 0; % não existem elementos marcados
071 - tamanhoElementos = 1;
072 - pertence = 0;
073 - outputMatrizCont = 1;
074 -
075 - for j = 1:length(sequencia)
076 -     for i = 1:length(elementos)
077 -         if sequencia(j) == elementos(i)
078 -             pertence = 1;
079 -         end
080 -     end
081 -     if pertence == 0
082 -         outputMatriz(:, :, outputMatrizCont) = matriz(:, :, j);
083 -         outputMatrizCont = outputMatrizCont + 1;
084 -         elementos(tamanhoElementos) = sequencia(j);
085 -         tamanhoElementos = tamanhoElementos + 1;
086 -
087 -         % marca sequencia(j) em elementos(length(elementos)+1);
088 -     end
089 -
090 -     pertence = 0;
091 - end
092 -
093 - for i = 1: length (elementos)
094 -     if elementos(i) < 26
095 -         notas(i).nome(1) = char (elementos(i) + 64);
096 -         notas(i).nome(2) = ' ';
097 -         notas(i).nome(3) = ' ';
098 -
099 -         notas(i).numero = elementos(i);
100 -         notas(i).outputMatriz = outputMatriz(:, :, i);

```

```

100 -     else
101 -         notas(i).nome(1) = char(fix(elementos(i)/26));
102 -         notas(i).nome(2) = char(mod(elementos(i)/26));
103 -         notas(i).nome(3) = ' ';
105 -         notas(i).numero = elementos(i);
106 -         notas(i).outputMatriz = outputMatriz(:,:,i);
107 -     end
108 - end
109 - i = 1;
110 - j = 1;
111 - while j < length(notas) + 1
112 -     frase_atual(i) = notas(j).nome(1);
113 -     frase_atual(i+1) = notas(j).nome(2);
114 -     base_do_canto(i+2) = notas(j).nome(3);
115 -     j = j + 1;
116 -     i = i + 3;
117 - end
118 -
119 - base_do_canto
120 - % elementos

```

#### Sub-rotina comBase

```

001 - % COM BASE
002 -
003 - n = input ('numero de notas: '); % lendo a entrada pelo prompt do
matlab
004 - % para ocorrer a comparacao deve existir pelo menos duas notas
005 -
006 - notal = imcrop;
007 - nota2 = imcrop; % Seleciona a nota desejada com o auxilio do mouse
008 -
009 - matriz = padronizaMatrizes(notal,nota2);
010 -
011 - i = 3;
012 -
013 - while (i < n + 1)
014 -     nota = imcrop; % Seleciona a nota desejada com o auxilio do
mouse
015 -     matriz = padronizaMatrizes(matriz, nota);
016 -     i = i + 1;
017 - end
018 -
019 - % até aqui foram lidas as matrizes de entrada
020 - % fatalmente elas estarão fora de padrão
021 - % a idéia é padronizar os dois vetores de matrizes pegando um 023 -
% elemento de cada- % vez
024 - % outputMatriz = notas(1).outputMatriz;
025 - % end
026 - % if length(notas) > 1
027 - % tamanho = size(matriz)
028 -
029 - for i = 1:length(notas)
030 -     outputMatriz(:,:,i) = notas(i).outputMatriz;
031 - end

```

```

032 - % end
033 - % problema
034 - % ncc = normxcorr2 (outputMatriz(:, :, 2), outputMatriz(:, :, 1));
035 - % max(ncc(:))*100
036 - novaMatriz = padronizaMatrizes(outputMatriz, matriz(:, :, 1));
037 - % carrega as matrizes anteriores
038 - if size(matriz, 3) > 1
039 -     for i = 2: size(matriz, 3)
040 -         novaMatriz = padronizaMatrizes(novaMatriz, matriz(:, :, i));
041 -     end
042 - end
043 - % na nova matriz
044 - % grau
045 - % matriz = novaMatriz;
046 - % a nova matriz e a somatoria entre a base de dados e a nova ent
047 -
048 - outTam = size(outputMatriz, 3);
049 - clear outputMatriz;
050 - clear matriz;
051 - for i = 1: outTam
052 -     outputMatriz(:, :, i) = novaMatriz(:, :, i);
053 - % carrega a matriz padronizada em output
054 - end
055 - j = 1;
056 - for i = outTam+1: size(novaMatriz, 3)
057 -     matriz(:, :, j) = novaMatriz(:, :, i);
058 -     j = j + 1;
059 - end
060 -
061 - % atualiza a base de dados inicial
062 - for i = 1:length(notas)
063 -     notas(i).outputMatriz = outputMatriz(:, :, i);
064 - end
065 -
066 - n = size(novaMatriz, 3);
067 - grau(1, 1) = 0; % usa a matriz anterior no calculo da recorrência
068 - for i = 1:(n - 1)
069 -     for j = (i + 1):n
070 - % escolha do template
071 - ncc = normxcorr2(novaMatriz(:, :, j), novaMatriz(:, :, i));
072 - grau(i, j) = max(ncc(:))*100;
073 -     end
074 - end
075 - grau(j, j) = 0;
076 - % grau
077 - % baseado no vetor de entrada pode-se determinar quantos
078 - % elementos repetidos existem
079 - padrao = 1;
080 - for i = 1: length(grau)
081 -     seqüência (i) = 0;
082 - end
083 - padraoUsado = 0;
084 - for i = 1:length(grau)
085 -     for j = 1:length(grau)
086 -         if ((grau(i, j) > 76) | (grau(i, j) == 0)) & (sequencia(j) == 0)

```

```

087 -             sequencia(j) = padrao;
088 -             padraoUsado = 1;
089 -         end
090 -     end
091 - if padraoUsado == 1
092 -     padrao = padrao + 1;
093 -     padraoUsado = 0;
094 - end
095 - end
096 - i = 1;
097 - for j = length(notas)+1:length(sequencia)
098 -     sequenciaOficial(i) = sequencia(j);
099 -     i = i + 1;
100 - end
101 -
102 - j = 1;
103 - i = 1;
104 - while j < length(sequencia) + 1
105 -     if sequencia(j) < 26
106 -         frase(i) = char(sequencia(j)+64);
107 -         frase(i+1) = ' ';
108 -         frase(i+2) = ' ';
109 -     else
110 -         frase(i) = char(fix(sequencia(j)/26));
111 -         frase(i+1) = char(mod(sequencia(j)/26));
112 -         frase(i+2) = ' ';
113 -     end
114 -     i = i + 3;
115 -     j = j + 1;
116 - end
117 - % frase
118 - % sequencia
119 - j = 1;
120 - i = 1;
121 - while j < length(sequenciaOficial) + 1
122 -     if sequenciaOficial(j) < 26
123 -         frase_atual(i) = char(sequenciaOficial(j)+64);
124 -         frase_atual(i+1) = ' ';
125 -         frase_atual(i+2) = ' ';
126 -     else
127 -         frase_atual(i) = char(fix(sequenciaOficial(j)/26));
128 -         frase_atual(i+1) = char(mod(sequenciaOficial(j)/26));
129 -         frase_atual(i+2) = ' ';
130 -     end
131 -     i = i + 3;
132 -     j = j + 1;
133 - end
134 - frase_atual
135 - % sequenciaOficial
136 - elementos = 0;
137 - tamanhoElementos = 1;
138 - pertence = 0;
139 - outputMatrizCont = 1;
140 -
141 - clear outputMatriz;

```

```

142 - for j = 1:length(sequencia)
143 -     for i = 1:length(elementos)
144 -         if sequencia(j) == elementos(i)
145 -             pertence = 1;
146 -         end
147 -     end
148 -     if pertence == 0
149 -         outputMatriz(:, :, outputMatrizCont) = novaMatriz(:, :, j);
150 -         outputMatrizCont = outputMatrizCont + 1;
151 -         elementos(tamanhoElementos) = sequencia(j);
152 -         tamanhoElementos = tamanhoElementos + 1;
153 -
154 -         % marca sequencia(j) em elementos(length(elementos)+1);
155 -         end
156 -         pertence = 0;
157 -     end
158 -
159 - for i = 1: length(elementos)
160 -     if elementos(i) < 26
161 -         notas(i).nome(1) = char(elementos(i)+64);
162 -         notas(i).nome(2) = ' ';
163 -         notas(i).nome(3) = ' ';
164 -         notas(i).numero = elementos(i);
165 -         notas(i).outputMatriz = outputMatriz(:, :, i);
166 -     else
167 -         notas(i).nome(1) = char(fix(elementos(i)/26));
168 -         notas(i).nome(2) = char(mod(elementos(i)/26));
169 -         notas(i).nome(3) = ' ';
170 -         notas(i).numero = elementos(i);
171 -         notas(i).outputMatriz = outputMatriz(:, :, i);
172 -     end
173 - end
174 -
175 - i = 1;
176 - j = 1;
177 - while j < length(notas) + 1
178 -     base_do_canto(i) = notas(j).nome(1);
179 -     base_do_canto(i+1) = notas(j).nome(2);
180 -     base_do_canto(i+2) = notas(j).nome(3);
181 -     j = j + 1;
182 -     i = i + 3;
183 - end
184 - base_do_canto
185 - % elementos

```

#### Sub-rotina padronizaLinha

```
001 - % Padronização da Linha da Matriz
002 - function matriz = padronizaLinha(matriz1,matriz2);
003 -
004 - sobra = size(matriz2,1) - size(matriz1,1);
005 - elemento = matriz2(1,1);
006 - tamLinha = size(matriz1,2);
007 -
008 - for i = 1: tamLinha
009 -     linha(i) = elemento;
010 - end
011 -
012 - for i = 1:sobra
013 -     matriz1 = [matriz1;linha];
014 - end
015 - matriz = matriz1;
```

#### Sub-rotina padronizaColuna

```
001 - % Padronização da Coluna da Matriz
002 - function matriz = padronizaColuna(matriz1,matriz2);
003 -
004 - sobra = size(matriz2,2) - size(matriz1,2);
005 - elemento = matriz2(1,1);
006 - tamColuna = size(matriz1,1); % o tamanho da linha eh o numero de
linhas
007 - for i = 1: tamColuna
008 -     coluna(i) = elemento;
009 - end
010 -
011 - coluna = coluna';
012 -
013 - for i = 1:sobra
014 -     matriz1 = [matriz1,coluna];
015 - end
016 -
017 - matriz = matriz1;
```

#### Sub-rotina padronizaMatriz

```
001 - % Padronização da Matrizes
002 - function matriz = padronizaMatrizes(matriz1, matriz2)
003 - % esta função busca e corrige colunas
004 - % matriz1 eh a matriz de 3 dimensoes que contem os cantos
005 - % matriz2 é a matriz com o canto medido
006 - % length(size(a)) numero de dimensões
007 - % matriz1(:, :, 1)
008 -
009 - numeroDeLinhasNota1 = size(matriz1(:, :, 1), 1);
010 - % a analise é feita sobre qq elemento da matriz
011 - numeroDeColunasNota1 = size(matriz1(:, :, 1), 2);
012 - numeroDeLinhasNota2 = size(matriz2, 1);
013 - numeroDeColunasNota2 = size(matriz2, 2);
014 -
```

```

015 - alteradoAux = 0;
016 - alteradoAux2 = 0;
017 - if numeroDeLinhasNota1 < numeroDeLinhasNota2
018 -     for i = 1:size(matriz1,3)
019 -         matrizAux(:, :, i) = padronizaLinha(matriz1(:, :, i), matriz2);
020 -         % a segunda entrada eh sempre maior
021 -     end
022 -     alteradoAux = 1;
023 - else
024 -     matriz2 = padronizaLinha(matriz2, matriz1(:, :, 1));
025 -     matrizAux = matriz1;
026 - end
027 -
028 - if numeroDeColunasNota1 < numeroDeColunasNota2
029 -     for i = 1:size(matriz1,3)
030 -         matrizAux2(:, :, i) = padronizaColuna(matrizAux(:, :, i), matriz2);
031 -     end
032 -     alteradoAux2 = 1;
033 - else
034 -     matriz2 = padronizaColuna(matriz2, matrizAux(:, :, 1));
035 - end
036 -
037 - if (alteradoAux == 1) & (alteradoAux2 == 1)
038 -     matrizAux2(:, :, size(matrizAux2,3)+1) = matriz2;
039 -     matriz = matrizAux2;
040 - elseif (alteradoAux2 == 1)
041 -     matrizAux2(:, :, size(matrizAux2,3)+1) = matriz2;
042 -     matriz = matrizAux2;
043 - elseif (alteradoAux == 1)
044 -     matrizAux(:, :, size(matrizAux,3)+1) = matriz2;
045 -     matriz = matrizAux;
046 - elseif (alteradoAux == 0) & (alteradoAux2 == 0)
047 -     matrizAux(:, :, size(matrizAux,3)+1) = matriz2;
048 -     matriz = matrizAux;
049 - end

```