

# IDENTIFICAÇÃO DE TECIDO COM ATEROSCLEROSE PELO MÉTODO DE ANÁLISE DO COMPONENTE PRINCIPAL

*Thiago Siqueira Pinto, Landulfo Silveira Junior*

Laboratório de Espectroscopia Biomolecular/IP&D, UNIVAP, Av: Shishima Hifumi, 2911, Urbanova, 12244-000, São José dos Campos – SP, thiagosp10@hotmail.com, landulfo@univap.br

**Resumo** - A espectroscopia Raman pode ser empregada para discriminar tecidos normais de tecidos com patologias nas artérias humanas, sendo o espectro Raman a impressão digital da molécula, demonstrando assim suas informações bioquímicas específicas. A Análise dos Componentes Principais (PCA) vem sendo empregada com sucesso nas análises espectroscópicas que tem permitido a identificação tanto de alterações patológicas como na determinação de parâmetros ópticos de tecido normais e patológicos. Sendo a PCA uma ferramenta que permite separar a informação relevante de um ruído, um modelo de análise baseado nessas informações leva à classificação das amostras em análise com rapidez. Neste trabalho, pretende-se trabalhar com a técnica de Espectroscopia Raman como método de obtenção do sinal do tecido biológico e também utilizar o método da PCA para a identificação das características espectrais mais importantes dos espectros de tecidos.

**Palavras-chave:** Espectroscopia Raman, Análise dos Componentes Principais (PCA).

**Área do Conhecimento:** III - Engenharias

## Introdução

A área de Espectroscopia Biomolecular visando o diagnóstico de doenças é uma área em pleno desenvolvimento, sendo que o reconhecimento de padrões insere-se dentro desta área principalmente na busca de formas de identificação de características espectrais que diferenciem tecidos sadios dos tecidos com patologias.

Neste trabalho será desenvolvido o uso da Espectroscopia Raman para analisar o Espectro da amostra será utilizado o método PCA.

A Espectroscopia Raman é uma poderosa ferramenta analítica para o estudo de mudanças físicas e químicas que ocorrem em tecidos biológicos.

A Análise dos Componentes Principais (PCA) pode ser utilizada para simplificar a informação redundante, ou seja, reduzir o número de variáveis, facilitando assim a análise por causa da minimização no número de variáveis necessárias à caracterização do espectro. Esta técnica é freqüentemente denominada de técnica de compressão de dados; como cada variável não apresenta informação redundante, pode-se afirmar ainda que a PCA é uma técnica de reconhecimentos de padrões.

O método da PCA utiliza um algoritmo que basicamente: toma o conjunto de dados, subtrai a média; calcula a covariância; calcula os autovetores e autovalores da matriz de covariância; escolhe componentes de um vetor característico; e deriva o novo conjunto de dados.

Essa técnica utiliza conjuntos de espectros, onde estão todos os dados obtidos na leitura

espectral, e os decompõem em Componentes Principais (PC), que são as variações de maior ocorrência em todos os espectros das diferentes amostras. São realizados diversos PCs, e estes vem sendo cruzados no intuito de se obter uma melhor classificação das amostras estudadas.

Os Scores (SC) que são coeficientes de escala são calculados para cada Componente Principal.

Portanto, conhecendo-se os vetores dos Componentes Principais, pode-se estabelecer a importância de cada um deles na formação do espectro de uma amostra desconhecida que se queira classificar. Pode-se então utilizar essa informação para a classificação dos espectros dentro de categorias pré-determinadas.

Pretende-se aplicar a técnica PCA em um conjunto de dados de Espectros Raman obtidos de artérias coronárias humanas, classificados como normal, aterosclerótico e calcificado. A estes espectros foram adicionados ruído para que se possa entender e avaliar a técnica PCA como parte da metodologia de classificação dos espectros.

## Materiais e Métodos

Para esse experimento foram utilizados espectros Raman de 3 amostras de tecidos coronarianos com 3 tipos de diagnósticos: tecido normal, com aterosclerose e com calcificação. A estes espectros foram aplicados ruído gaussiano com duas intensidades diferentes, formando dois conjuntos de espectros: Espectros(1) e Espectros (2).

Para fazer o PCA foi utilizado o programa Matlab 6.1 e a Plotagem dos gráficos foi feita no programa Excel 7.0.

O método utilizado para a análise dos espectros foi o da Análise dos Componentes Principais (PCA). A Análise dos Componentes Principais consiste essencialmente em rescrever as coordenadas das amostras em outro sistema de eixo mais conveniente para a análise dos dados. Em outras palavras, as n-variáveis originais geram, através de suas combinações lineares, n-componentes principais, cuja principal característica, além da ortogonalidade, é que são obtidos em ordem decrescente de máxima variância, ou seja, a componente principal 1 detém mais informação estatística que a componente principal 2, que por sua vez tem mais informação estatística que a componente principal 3 e assim pôr diante.

Este método permite a redução da dimensionalidade dos pontos representativos das amostras, pois embora a informação estatística presente nas n-variáveis originais seja a mesma dos n-componentes principais, é comum obter em apenas 2 ou 3 das primeiras componentes principais mais que 90% desta informação. O gráfico da componente principal 1 versus a componente principal 2 fornece uma janela privilegiada (estatisticamente) para observação dos pontos no espaço n-dimensional.

A Análise de Componentes Principais também pode ser usada para verificar a importância das próprias variáveis originais escolhidas, ou seja, as variáveis originais com maior peso (*loadings*) na combinação linear dos primeiros componentes principais são as mais importantes do ponto de vista estatístico.

Portanto, a tarefa de quem trabalha com estatística multivariada consiste em interpretar a distribuição dos pontos no gráfico de componentes principais e identificar as variáveis originais com maior peso na combinação linear das componentes principais mais importantes.

O conjunto de espectros foi transportado do Excel para o Matlab, para que se aplicasse a Análise dos Componentes Principais.

Para executar os dados do Excel no Matlab, primeiro salva-se o arquivo como extensão .TXT. Logo após, é acessado o arquivo pelo Matlab através do comando "load nome\_do\_arquivo.TXT. Com isso o arquivo é carregado em forma de matriz, onde cada linha (x) representa o comprimento de onda e cada coluna (y) o número da amostra.

Com a matriz já carregada foram inseridos os comandos para se calcular o PCA, [COEFF, SCORE] = princomp (nome\_do\_arquivo). Assim, vão se formar mais duas matrizes que são:

- COEFF que tem dimensão yXy.
- SCORE que tem dimensão xXy.

Depois de obter-se COEFF e SCORE, estas variáveis são transportadas para o Excel em formato .TXT para que se faça a plotagem dos pontos, através do seguinte comando: save score.txt SCORE -ASCII e save coeff.txt COEFF -ASCII.

## Resultados

Após esses procedimentos foram adquiridas as matrizes COEFF(1), SCORE(1); e COEFF(2), SCORE(2). O COEFF é responsável pela classificação dos espectros.

Foram feitas as classificações para espectros(1), conjunto de dados que apresenta pequena intensidade do ruído gaussiano e espectros(2), conjunto de dados que apresenta alta intensidade do ruído gaussiano, e os resultados foram localizados nas três primeiras colunas. A partir da coluna 4 até a 27 se tem apenas ruído. Os resultados foram separados por cores nas seguintes figuras:

a) COEFF do conjunto de dados dos Espectros(1).

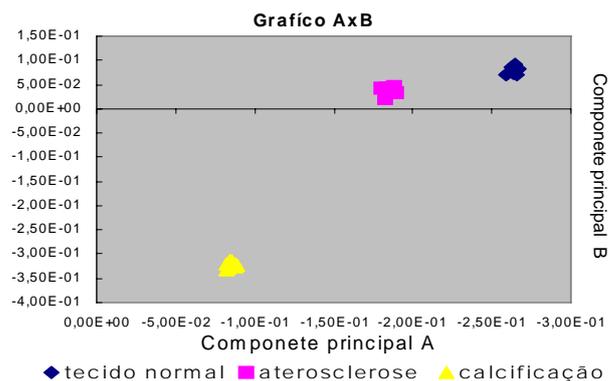


Figura 1 - Plotagem do Componente Principal A versus Componente Principal B do conjunto Espectro(1).

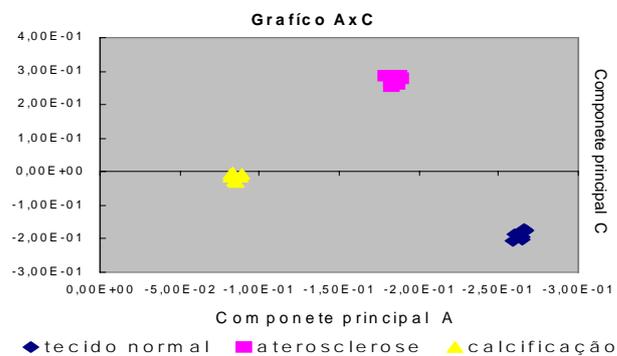


Figura 2 - Plotagem da Componente Principal A versus Componente Principal C do conjunto Espectro(1).

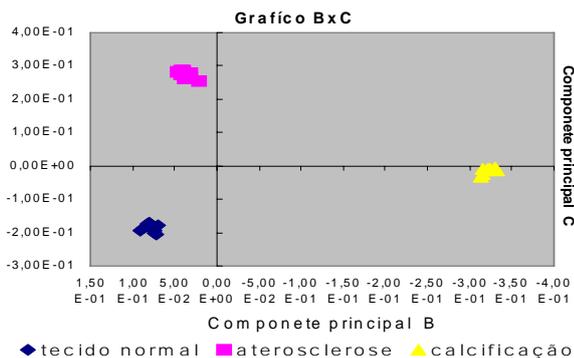


Figura 3 - Plotagem da Componente Principal B versus Componente Principal C do conjunto Espectro(1).

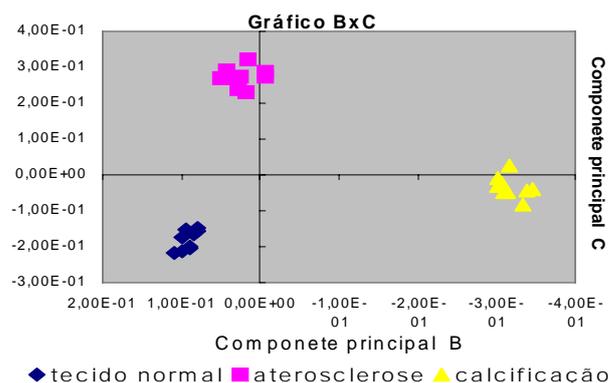


Figura 6 - Plotagem da Componente Principal A versus Componente Principal C do conjunto Espectro(2).

b) COEFF do conjunto de dados dos Espectros(2)

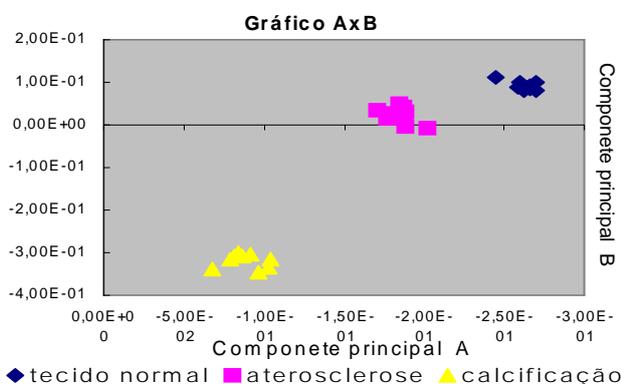


Figura 4 - Plotagem da Componente Principal A versus Componente Principal B do conjunto Espectro(2).

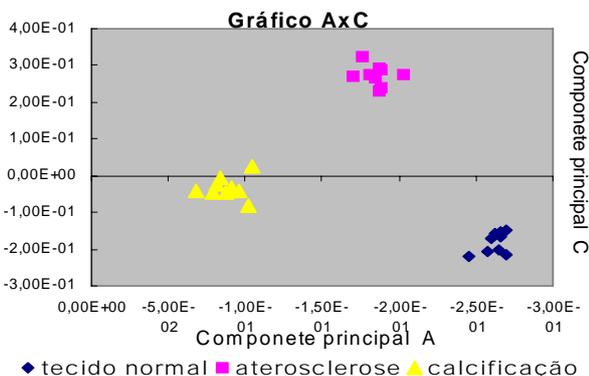


Figura 5 - Plotagem da Componente Principal A versus Componente Principal C do conjunto Espectro(2).

De acordo com os gráficos apresentados, o conjunto de dados Espectros(1) estão melhor agrupados do que os dos Espectros(2), pois a intensidade do ruído em Espectros(2) é maior fazendo com que os pontos estejam mais distantes (espalhados) uns dos outros, isso já não ocorre em Espectros(1) pois seus pontos estão mais próximos.

c) SCORE do conjunto de dados Espectros(1)

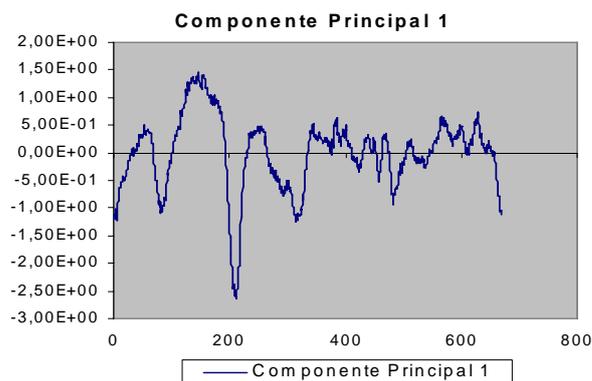


Figura 7 - Espectro da Componente Principal 1, do conjunto Espectros(1).

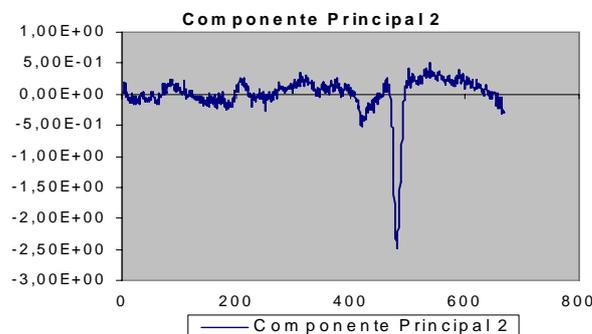


Figura 8 - Espectro da Componente Principal 2, do conjunto Espectros(1).

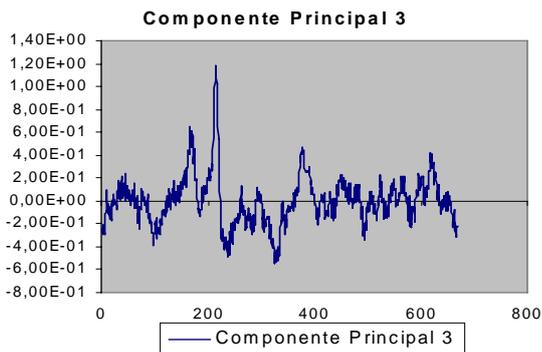


Figura 9 - Espectro da Componente Principal 3, conjunto Espectros(1).

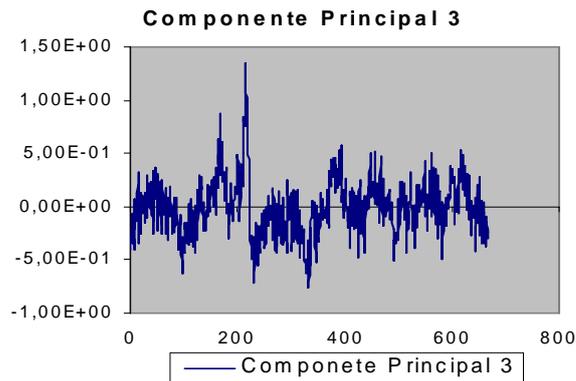


Figura 13 - Espectro da Componente Principal 3, conjunto Espectros(2).

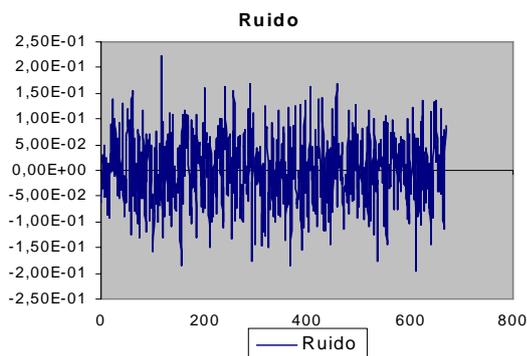


Figura 10 – Espectro da componente Principal 4 (ruído), conjunto Espectro(1).

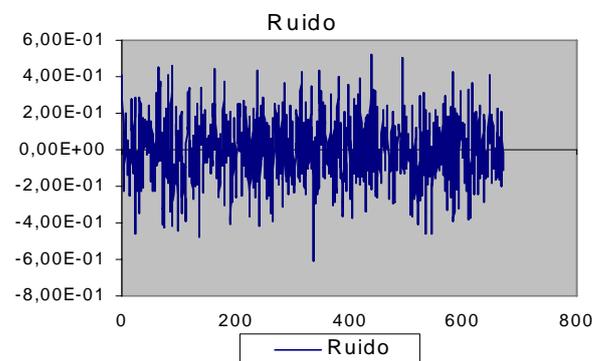


Figura 14 – Espectro da componente Principal 4 (ruído), conjunto Espectros(2).

d) SCORES do conjunto de dados Espectros(2)

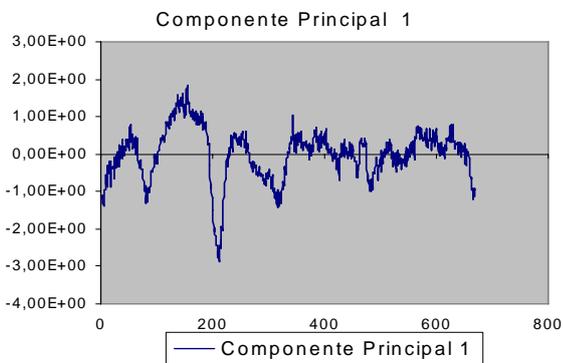


Figura 11 - Espectro da Componente Principal 1, conjunto Espectros(2).

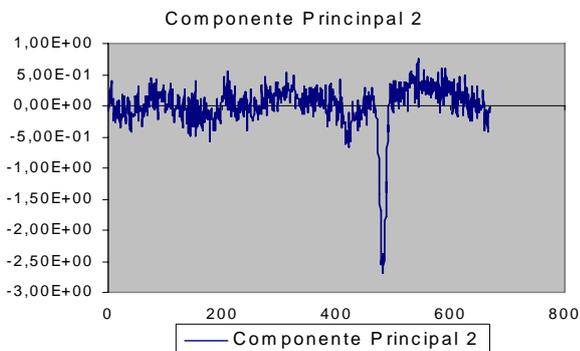


Figura 12 - Espectro da Componente Principal 2, conjunto Espectros(2).

**Conclusão**

A obtenção dos resultados de identificação e classificação das amostras se deu pela técnica estatística de Análise dos Componentes Principais, que foi aplicada no software Matlab, e os resultados estão sendo apresentados nas figuras de 1 a 14.

Através desse gráficos é possível classificar e identificar os tecidos e observar as diferenças de intensidade de ruídos entre Espectros(1) e Espectros(2).

Observou-se ainda que em virtude da diferença de intensidade de ruído, os resultados de agrupamento dos dados no conjunto Espectros(1) ficaram melhor apresentados que nos gráficos do conjunto Espectros(2).

A Análise dos Componentes Principais realizadas nesse trabalho foi efetivamente adequada para a classificação dos tecidos, onde cada Componente Principal apresentou características únicas, embora o nível de ruído fosse diferente nos dois grupos.

**Agradecimentos**

Projeto com suporte financeiro da FVE/Univap – IP&D, através de Bolsa PIBIC-Univap.

### Referências

[1] Lindsay I Smith. A tutorial on Principal Components Analysis. February 26, 2002.

Disponível em: [http://www.cs.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf). Acesso em 15/03/05.

[2] Luiz Fernando C.de Oliveira. Espectroscopia Molecular. Maio 2001. Disponível em:

<http://sbqensino.foco.fae.ufmg.br/uploads/601/espect.pdf>. Acesso em 20/03/2005.

[3] Landulfo Silveira Jr. Sistema de espectroscopia Raman no infravermelho.

Disponível em: <http://www1.univap.br/landulfo/pesq3.htm> Acesso em: 02/04/2005.

[4] Internet site. <http://www.dq.fct.unl.pt/qof/chem10.html>. Acesso em 10/04/2005.