

PROCESSAMENTO DE SINAIS UTILIZANDO PCA EM ESPECTROS DE FLUORESCÊNCIA DE BACTÉRIAS

Daniel Vieira de Paula¹, Douglas Braga¹, Landulfo Silveira Jr.²

UNIVAP/FEAU, Av. Shishima Hifumi, 2911, Urbanova – São José dos Campos – SP
daniell_sjc@yahoo.com.br¹; landulfo@univap.br²

Resumo – O processamento de sinais através da técnica de análise de componentes principais (PCA) é uma importante ferramenta utilizada para identificar as variáveis mais importantes no espaço das componentes principais. A demonstração de diferenças no espectro de autofluorescência em três gêneros bacterianos e sua posterior separação, através da técnica PCA em grupos com semelhança espectral, permite o desenvolvimento de uma rotina que, pelo padrão de fluorescência obtido, identifique automaticamente o gênero bacteriano, de maneira rápida. Foram obtidos espectros de fluorescência de três bactérias (*E. coli*, *E. faecalis*, *S. aureus*) em diferentes comprimentos de onda de excitação. Foi desenvolvida uma rotina baseada nas diferenças dos espectros dos componentes principais, e os resultados indicaram que o comprimento de onda de excitação de 410nm foi o melhor separador espectral dos tipos de bactérias.

Palavras-chave: Identificação bacteriana, espectroscopia de fluorescência, reconhecimento de padrões, análise dos componentes principais (PCA), processamento de sinais.

Área do Conhecimento: III - Engenharias

Introdução

A bacteriologia moderna identifica gêneros e espécies bacterianas por métodos bioquímicos, manuais ou automatizados, com tempo de diagnóstico que duram em torno de 24 h, além de serem caros devido aos reagentes utilizados. O desenvolvimento de um método óptico baseado na espectroscopia de fluorescência, que não necessite de reagentes e com curto tempo de processamento, poderá diminuir tanto o tempo de obtenção do resultado quanto o custo de cada exame diagnóstico (GIANNA, 2001).

A espectroscopia de fluorescência é um método convencionalmente empregado no estudo de materiais para a determinação da composição química em função de bandas de emissão fluorescentes características. O mesmo princípio pode ser utilizado para a determinação da presença de agentes fluorescentes em tecidos biológicos, que podem ser úteis no diagnóstico (MACIEL e BAGNATO, 2008).

Novas ferramentas e técnicas matemáticas têm sido adaptadas para processar informações obtidas de uma medição. Entre elas, está o reconhecimento de padrões através do processamento de sinais utilizando o PCA (Análise de Componentes Principais), uma técnica estatística poderosa que pode ser utilizada para redução do número de variáveis e para fornecer uma visão estatisticamente privilegiada do conjunto de dados. A análise de componentes principais fornece as ferramentas adequadas para

identificar as variáveis mais importantes (JOLLIFFE, 2002).

Em um dado experimento que necessite classificar os dados em grupos bem definidos, a partir das informações espectrais das amostras, pode-se verificar quais variáveis, ou seja, os componentes principais, são mais importantes para a correta identificação destes grupos (SILVEIRA JR., 2001).

O objetivo deste estudo é utilizar a técnica estatística multivariada PCA para a identificação de grupos de bactérias – *E. coli*, *E. faecalis* e *S. aureus*, através da verificação das diferenças dos espectros de autofluorescência, obtidos em trabalhos anteriores, em períodos distintos, pela análise das diferenças dos componentes principais do conjunto de dados. Desta forma, será possível comprovar a possível eficiência do uso dos padrões de fluorescência típicos para cada grupo estudado, em um conjunto de comprimentos de onda de excitação, visando uma rápida identificação dos agentes de infecções bacterianas.

Metodologia

Foi realizada a análise PCA utilizando os espectros de fluorescência obtidos em estudos anteriores (GIANNA, 2001; SANTOS, 2004), primeiramente para aprendizado e familiarização da técnica PCA aplicada a processamento de sinais em seguida detectar se o tempo decorrido entre os experimentos causa diferenças significativas na análise espectral, e se estas

diferenças ainda permitem identificar os gêneros bacterianos em questão.

Para o processamento dos dados, foi utilizada a função "princomp", do programa Matlab 4.2 cedido pela UNIVAP. A técnica PCA permite concentrar ao máximo a variância do conjunto de dados (A) em um número reduzido de variáveis, chamadas de componentes principais (PCs). Um conjunto de coeficientes de escala, os escores (ES), são calculados para cada componente principal. Os escores representam a contribuição de cada PC na variância do espectro. Quando as componentes principais são multiplicadas pelos escores e então somadas, temos a reconstrução do espectro original:

$$A = PC \times E$$

O espectro de fluorescência de cada amostra bacteriana foi obtido no Espectrofluorímetro FluoroMax-2 marca SPEX, nos comprimentos de onda de excitação entre 330 e 510 nm em intervalos de 20 nm e emissão coletada 20 nm acima da excitação até 700 nm.

Os dados coletados de cada amostra bacteriana foram listados no programa Microsoft Excel. Os gráficos espectrais obtidos foram posteriormente normalizados pelo pico mais intenso, e a média de cada grupo de bactérias foi plotada conforme mostrado nas Figuras 1, 2 e 3.

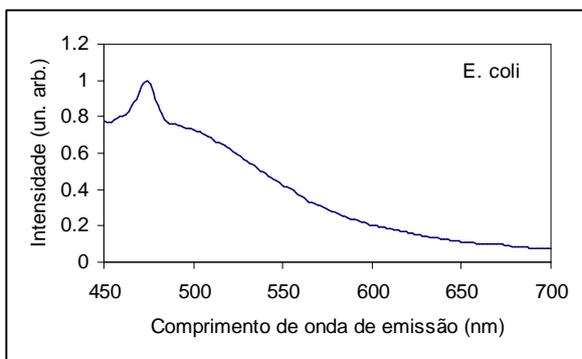


Figura 1: Média dos espectros de autofluorescência das amostras de *E. coli* com excitação em 410 nm.

Foi aplicada a técnica PCA, sendo que os primeiros componentes principais são os responsáveis pela maior diferenciação da fluorescência. Foram utilizados os quatro primeiros PCs: 1, 2, 3 e 4, a fim de verificar a melhor separação entre os espectros bacterianos para permitir sua identificação.

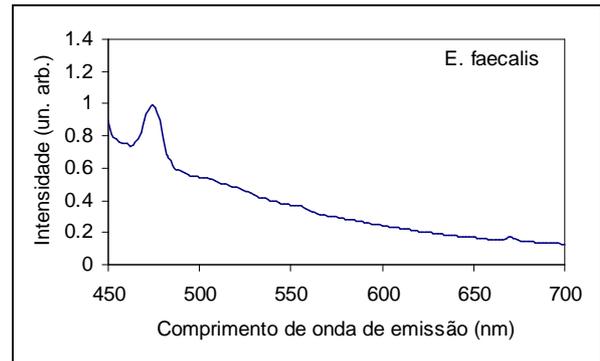


Figura 2: Média dos espectros de autofluorescência das amostras de *E. faecalis* com excitação em 410 nm.

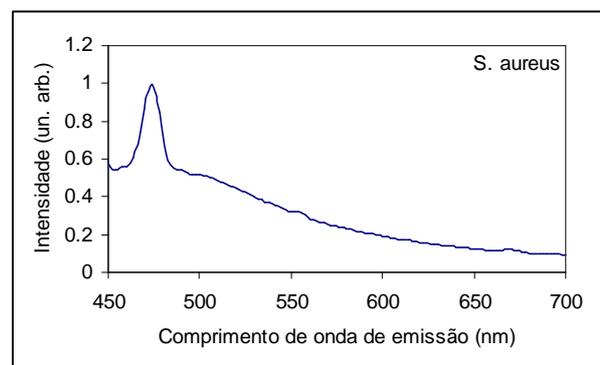


Figura 3: Média dos espectros de autofluorescência das amostras de *S. aureus* com excitação em 410 nm.

Resultados

A Figura 4 demonstra a plotagem da primeira análise dos 3 grupos distintos de bactérias, com 30 amostras, sendo 10 de cada gênero as quais tiveram os espectros submetidos ao processamento de sinais PCA, identificadas previamente por um método automatizado, que servirão como calibração para as demais análises.

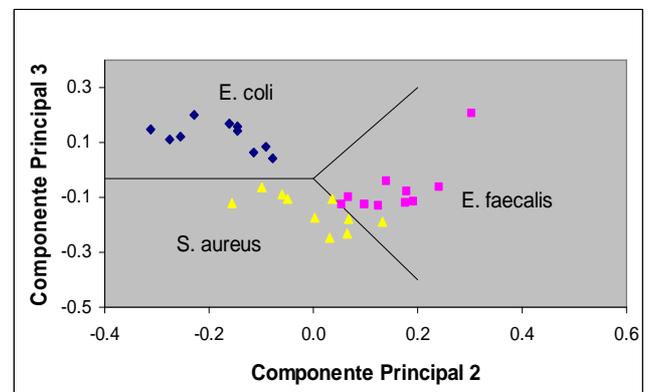


Figura 4: Plotagem dos escores das PC 2 e PC 3 das 30 amostras bacterianas testadas na primeira experiência.

O – *E. coli*, Δ – *S. aureus*, □ – *E. faecalis*

A Figura 5 demonstra a plotagem da segunda análise, também com 30 amostras sendo 10 de cada gênero. A mesma realizada em data próxima à calibração.

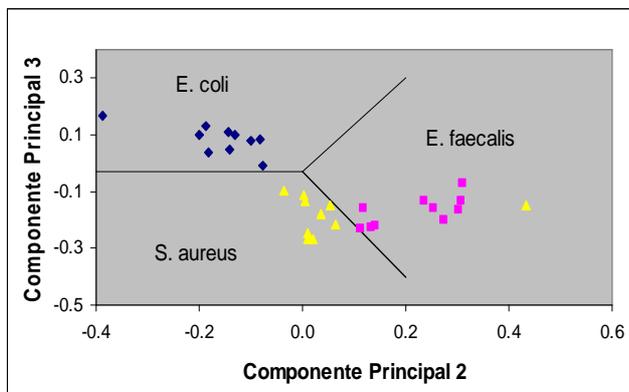


Figura 5: Plotagem dos escores das PC 2 e PC 3 das 30 amostras bacterianas testadas na segunda experiência.

O – E. coli, Δ - S. aureus, □ – E. faecalis

A Figura 6 demonstra a plotagem da terceira análise realizada com 18 amostras sendo 6 de cada gênero, e em data distante da primeira e da segunda análise.

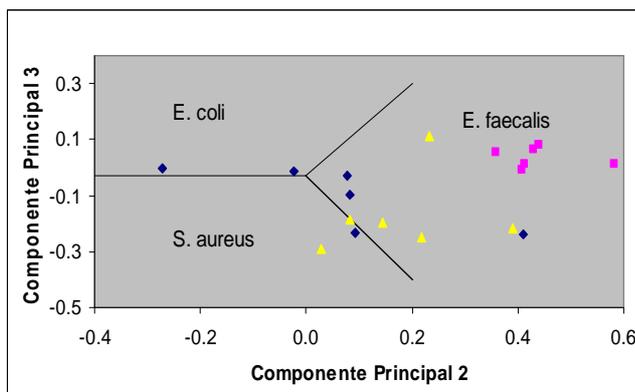


Figura 6: Plotagem dos escores das PC 2 e PC 3 das 18 amostras bacterianas testadas na última experiência.

O – E. coli, Δ - S. aureus, □ – E. faecalis

Discussão

No desenvolvimento do trabalho, verificou-se que na primeira e segunda análise, após a aplicação do método PCA, cada grupo de bactéria, se agrupa em uma mesma região de identificação, fornecendo consistência ao método empregado, conforme mostrado nas Figuras 4 e 5. Entretanto, na terceira análise, não foi verificado o mesmo padrão de agrupamento.

Esta diferença encontrada entre as duas primeiras e a terceira análise pode se dar ao fato de que as experiências foram executadas com

longo intervalo de tempo entre elas e, que na última, o método de PCA não havia sido originalmente aplicado.

Após aplicação do método PCA no terceiro experimento, não aconteceu o mesmo agrupamento das bactérias. Esta inconsistência pode ser resultado de uma diferença na calibração do equipamento responsável pelas leituras dos espectros, que não teve sua repetibilidade avaliada.

Conclusão

O processamento pela análise de componentes principais é uma ferramenta bastante útil e eficaz, que ajuda a revelar a informação mais relevante, que pode estar aparentemente oculta. A eficiência total relativa ao conjunto de teste pode indicar qual a quantidade de componentes que é necessária para ter-se uma boa representação do espaço original de dados de entrada, reduzindo drasticamente a dimensão deste espaço.

A utilização dos espectros de fluorescência para a obtenção do modelo de identificação necessita fundamentalmente de uma avaliação da repetibilidade da resposta do equipamento ao longo do tempo, a fim de que erros de diagnóstico não ocorram por falhas de calibração.

Referências

- GIANNA, H. E. Identificação de gêneros bacterianos por espectroscopia de fluorescência de multi-excitação, Dissertação de Mestrado, Universidade do Vale do Paraíba, São José dos Campos, 2001.
- MACIEL, V. H; BAGNATO, V. S. Espectroscopia de Fluorescência no diagnóstico de lesões teciduais. Disponível em: www.abfm.org.br/c2006/trabalhos/H_vitori_352.pdf . Acesso em 30 de março de 2008.
- SILVEIRA JR., L. Correlação entre a técnica de espectroscopia Raman e a análise histopatológica das placas ateromatosas em artérias coronárias humanas". Tese de Doutorado, Universidade de São Paulo, São Paulo, 2001.
- JOLLIFFE, I. T. Principal Component Analysis Ed. Springer-Verlag, 2ª Ed., New York, 2002.
- SANTOS, P. M. Metodologia de Identificação de Cepas Bacterianas por Espectroscopia de Fluorescência, Trabalho de graduação, Universidade do Vale do Paraíba, São José dos Campos, 2004.