

ESTUDO PRELIMINAR DA CLASSIFICAÇÃO DE ARRITMIAS CARDÍACAS UTILIZANDO ANÁLISE DE COMPONENTES PRINCIPAIS

por

J. NADAL¹ e R. B. PANERAI²

RESUMO -- Uma nova técnica de classificação de complexos do ECG é proposta, objetivando a monitoração automática de arritmias cardíacas. A partir dos sinais de 32 pacientes do "MIT-BIH Arrhythmia Database", 200 segmentos de ECG, cada um contendo as ondas P-QRS, são tomados como padrões representativos de diferentes tipos de batimentos cardíacos. A análise de componentes principais, aplicada a estes padrões, indica que apenas 10 componentes representam 99,2% da variância total do conjunto de sinais. A reconstrução de complexos, usando estas componentes, preserva a morfologia dos sinais. A classificação dos batimentos cardíacos de um dos pacientes do "Database", aplicando a regressão logística para determinar funções de separação, resulta em precisões de 99,1% (batimentos normais), 98,3% (extrasístoles ventriculares) e 91,4% (batimentos de fusão).

INTRODUÇÃO

A medicina conseguiu progressos em relação à morte súbita coronariana, uma das principais causas de óbito no Brasil, através da introdução de inovações tecnológicas. Ainda nos anos 50 a técnica de massagem cardíaca e os desfibriladores constituíram o primeiro marco neste quadro mórbido. O aparecimento das unidades coronarianas foi importante para a aplicação eficiente da cardioversão e para a monitoração visual contínua do ECG, uma vez que a fibrilação, principal causa da morte súbita, é geralmente precedida por arritmias cardíacas (Lown e Wolf, 1971).

A monitoração visual requer vigilância contínua e cuidadosa. Falhas de detecção são comuns devido à fadiga e desatenção do observador. Novos monitores eletrônicos surgiram, incorporando freqüencímetros cardíacos, capazes de detectar a ocorrência de taqui- e bradicardias, a partir de limites pré-fixados de freqüência. Tais sistemas apresentaram altos

¹-Professor Assistente,

²-Professor Titular,

Programa de Engenharia Biomédica, COPPE/UFRJ, Cx. Postal 68510, 21945-Rio de Janeiro-RJ

índices de falsos alarmes e foram substituídos por monitores computadorizados. A despeito do uso de computadores, dificuldades devidas à monitoração, em tempo real, levaram à severas restrições nos novos sistemas, sendo vários deles limitados à detecção da atividade ectópica ventricular, sem se ocupar em diagnosticar outras arritmias (ECRI, 1985).

Com a rápida evolução da informática, nos anos 70, reduziram-se as limitações tecnológicas. Como os novos minicomputadores processavam os algoritmos existentes com maior velocidade, e eram muito onerosos para atender um único paciente por vez, passou-se ao desenvolvimento de estações centrais. Em pouco tempo chegou-se a sistemas capazes de monitorar simultaneamente mais de dez leitos, adequados a grandes hospitais e clínicas especializadas, mas proibitivos para unidades de menor porte e de países em desenvolvimento, devido ao alto custo de atendimento por paciente.

A inovação seguinte deu-se com a introdução de microcomputadores, capazes de processar a monitoração de um único paciente. A adoção desses monitores de beira de leito, de baixo custo, é aparentemente mais adequada a países em desenvolvimento. Atualmente, com a disponibilidade de microcomputadores extremamente mais poderosos em termos de velocidade, capacidade de memória e arquitetura interna, e de co-processadores aritméticos de grande precisão e desempenho, pode-se pensar em pequenas estações centrais, baseadas em microcomputadores, ou em monitores de beira de leito que utilizem técnicas mais sofisticadas e precisas.

Classificação de Complexos QRS para Monitores de Arritmias

Todos os monitores automáticos de arritmias baseiam-se na detecção prévia dos complexos QRS. Pereira et alii (1988) apresentam um detector baseado em compressão do ECG. Lima (1986) trabalha com a primeira derivada do ECG. Após a localização da onda R, segue-se a extração e análise dos complexos QRS para a detecção da ocorrência de arritmias. A classificação dos complexos normalmente é efetuada por dois métodos, apresentados a seguir.

Comparação de Padrões— Um exemplo desta técnica, proposto por Feldman et alii (1971), baseia-se no cálculo da correlação cruzada entre complexos para formar diferentes famílias de batimentos cardíacos. Os primeiros complexos analisados, com formas similares, constituem um grupo normal, e a média desses sinais define um "padrão normal". A cada novo complexo calcula-se a correlação com o padrão normal. Se o coeficiente de correlação for menor que um limiar, o complexo é considerado anormal, sendo então comparado com os padrões de outros grupos já constituídos. Caso não exista nenhum similar, forma-se um novo grupo. Esta técnica consome um tempo razoável para efetuar todas as correlações necessárias. Por este motivo, a maioria dos sistemas é sujeita a restrições, tais como o uso de poucas amostras para o cálculo das correlações e a limitação do número máximo de famílias admitidas.

Técnica paramétrica— Este método utiliza características do complexo para determinar anormalidades. Usualmente toma-se como parâmetros alguns detalhes reconhecidos como discriminantes pelos próprios especialistas em ECG, tornando o método empírico. O

sistema descrito por Nolle (1972), por exemplo, baseia-se em quatro parâmetros do complexo QRS para a classificação: altura, duração, área absoluta e "off-set". A classificação baseia-se no cálculo de distâncias Euclidianas entre agrupamentos.

Técnicas alternativas— Vários sistemas mais recentes são baseados em outras técnicas, como os exemplos a seguir. Flosi (1988) combina as duas técnicas supra-citadas. Florenz et alii (1975) baseiam-se em máquinas de estados finitos. Henkind et alii (1985) utilizam sistemas especialistas. Jarisch e Mark (1985) usam modelos estocásticos de análise de agrupamentos. Mead et alii (1982) estudam a classificação no domínio da frequência.

O presente trabalho propõe a aplicação da análise de componentes principais (ACP) na classificação de arritmias. A ACP constitui um método clássico de análise para grandes massas de dados, pela sua capacidade de reduzir o volume de dados. O método foi escolhido por combinar aspectos das duas técnicas clássicas: tal qual a técnica de correlação cruzada, considera todo o complexo, e assim como a técnica paramétrica, extrai características do sinal. Tem ainda a vantagem de que as componentes são ortogonais entre si e sua extração é menos empírica.

MATERIAIS E MÉTODOS

O "MIT-BIH Arrhythmia Database"

O desenvolvimento do projeto baseia-se em um banco de dados, com eletrocardiogramas previamente classificados por cardiologistas (MIT, 1982). O "MIT-BIH Arrhythmia Database" contém trechos de 30 minutos de sinal, em duas derivações, de 48 pacientes. Uma das derivações de cada sinal, originalmente na forma analógica, é digitalizada por meio de conversão A/D, com frequência de amostragem de 240 Hz, e armazenada em disco, em um microcomputador. São gerados dois arquivos para cada paciente: um contendo as amostras do sinal, em 10 "bits" e outro com a classificação de cada complexo, e anotações de ritmo, referenciadas aos sinais por um contador do número de amostras. Dois terços dos sinais do banco de dados (32 pacientes) são usados para o desenvolvimento do projeto, sendo os demais deixados de lado, para permitir avaliações não tendenciosas dos algoritmos. Com base nas anotações disponíveis, cada batimento cardíaco é isolado, formando um segmento de 120 amostras (500 ms), com o máximo da onda R na 60^a amostra. Cada segmento é gravado em disco, juntamente com o intervalo RR (distância entre a onda R atual e a anterior) e a respectiva classificação. Ao todo dispõe-se de 75.904 registros, incluindo batimentos normais e diferentes tipos de arritmias.

No presente trabalho utilizam-se trechos de apenas 80 amostras de cada batimento, com a onda R na 50^a posição, de modo que cada trecho contém o complexo QRS e a onda P, quando esta ocorre. De cada complexo P-QRS é removida a tendência (flutuação da linha de base), subtraindo-se uma reta ajustada às primeiras e últimas amostras. A opção por arquivos com pontos em excesso permite, no futuro, estender o número de amostras para o

processamento.

Técnica de Componentes Principais

Do total de complexos, 200 são selecionados, sendo 144 normais e 56 de outros tipos, mantendo-se a proporção existente no banco de dados. Os 200 padrões constituem uma matriz X, e a matriz de covariância C é calculada por (Panerai et alii, 1988):

$$c_{ij} = (1/N-1) \sum_{k=1}^N (x_{ik} - \bar{x}_i) (x_{jk} - \bar{x}_j) \quad \{ i, j = 1, 2, \dots, M \} \quad (1)$$

onde \bar{x} é a média dos padrões, $N=200$, e $M=80$ (número de amostras).

As componentes principais (CPs) são obtidas pela solução de:

$$C \cdot w_p = k_p w_p \quad \{ p = 1, 2, \dots, P \} \quad (2)$$

onde $P \leq M$, w_p é a p-ésima CP (autovetor) e k_p é o correspondente autovalor (Morrison, 1976). Cada autovalor tem valor proporcional à fração da variância total representada pelo respectivo w_p . Os autovetores concentram o máximo possível da variância dos sinais nas primeiras CPs e constituem uma base ortogonal.

Os coeficientes de um complexo P-QRS são dados por:

$$a_{qp} = \sum_{i=1}^m x_{qi} w_{ip} \quad \{ q = 1, 2, \dots, Q; p = 1, 2, \dots, P \} \quad (3)$$

onde $Q=75.904$ e P é o número julgado satisfatório de CPs.

Após o cálculo dos coeficientes, cada complexo pode ser reconstruído fazendo-se a transformação inversa (Morrison, 1976):

$$\hat{x}_{qi} = x_i + \sum_{p=1}^P a_{qp} w_{pi} \quad \{ q = 1, 2, \dots, Q; i = 1, 2, \dots, M \} \quad (4)$$

onde \hat{x}_{qi} constitui uma aproximação de x_{qi} .

O poder de separação das CPs é estudado, inicialmente, através de diagramas de espalhamento, usando-se pares de componentes. Uma alternativa para a classificação é o emprego de funções de separação, uma para cada par de padrões distintos (Tou e Gonzales, 1974). Esta técnica é utilizada como exemplo, para um dos pacientes do banco de dados. As funções são obtidas calculando-se regressões logísticas, através do pacote estatístico "BMDP Statistical Software" (Dixon, 1985). Os resultados obtidos são comparados aos valores esperados

(classificação do banco de dados), para avaliar-se o desempenho do classificador.

RESULTADOS

A média dos 200 complexos P-QRS, assumidos como padrões, e as primeiras 10 CPs obtidas são exibidas na figura 1. Os demais autovetores não têm maior importância posto que os 10 primeiros concentram 99,2% da variância total dos dados.

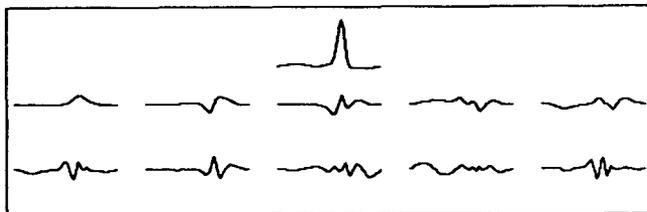


Figura 1: Parte superior: média dos 200 padrões usados no cálculo das componentes principais (escala arbitrária). Parte inferior: as 10 primeiras componentes principais, obtidas com os mesmos padrões.

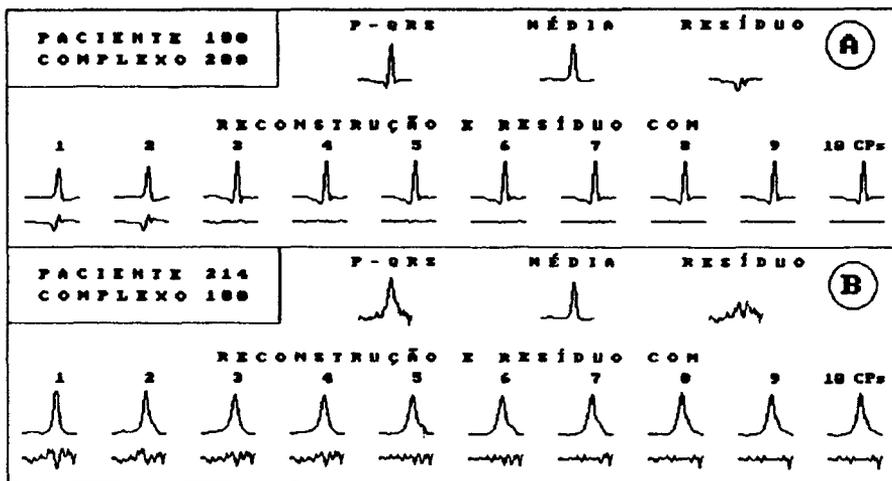


Figura 2. Reconstrução de complexo P-QRS utilizando componentes principais. Escala de amplitudes é arbitrária e igual para todos os sinais. [A] Exemplo típico de complexo sem ruído (parte superior esquerda), padrão médio (centro), e diferença entre os dois sinais (parte superior direita). Parte inferior: sinais reconstruídos com 1 a 10 componentes principais e seus correspondentes resíduos. (B) Exemplo típico de complexo P-QRS com ruído.

A figura 2 mostra dois exemplos de reconstrução de batimentos, utilizando a média e as 10 primeiras CPs. Observa-se, em 3-A, que a partir da 4ª CP o erro da reconstrução torna-se pequeno. No caso de sinais contaminados por ruído, como no exemplo em 3-B, a reconstrução atua como um filtro, pois a variabilidade do ruído somente aparecerá nas componentes de maior ordem.

Após a extração das CPs, são calculados os coeficientes dos complexos P-QRS e obtidos diagramas de espalhamento, para avaliação do poder de separação das componentes. Um exemplo, para o paciente 208 do banco de dados, é mostrado na figura 3. Sobre a figura são exibidas as funções de separação, calculadas por meio de regressão logística. Os resultados de classificação, para os batimentos do respectivo paciente, são apresentados na tabela 1. Os valores obtidos correspondem a 97,8% de classificações corretas, com parciais de 99,1% para os batimentos normais, 98,3% para extra-sístoles ventriculares (PVCs) e 91,4% para batimentos de fusão. Os erros de classificação correspondem a 2,2%, sendo assim distribuídos: 23,4% (batimentos normais), 26,6% (extra-sístoles) e 50,0% (batimentos de fusão).

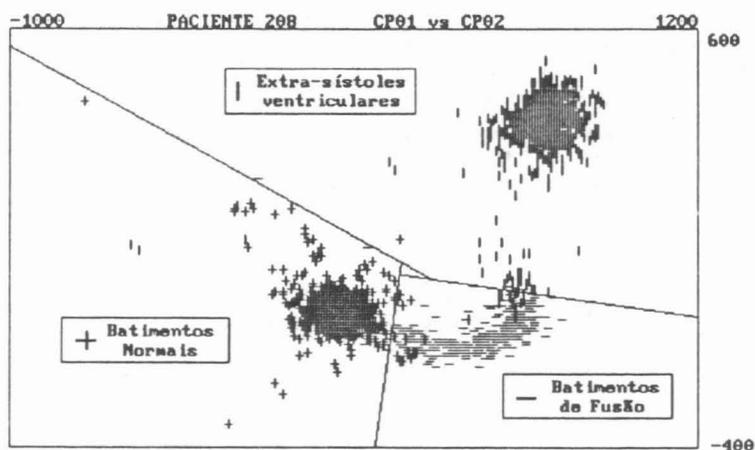


Figura 3. Diagrama de espalhamento dos batimentos do paciente 208 do "MIT-BIH Arrhythmia Database", utilizando as 2 primeiras componentes principais. As linhas que delimitam as três classificações correspondem a funções de separação, calculadas por meio de regressões logísticas.

Tabela 1 Resultados da classificação dos batimentos do paciente 208, usando as duas primeiras Componentes Principais. (PVC = extra-sístoles ventriculares)

Classificação do "MIT - BIH Database"				
R e s u l t a d o		Normal	Fusão	PVC
	Normal	1569	29	4
	Fusão	14	340	13
	PVC	1	3	974

DISCUSSÃO E CONCLUSÃO

Foram apresentados alguns resultados da utilização da análise de componentes principais na classificação de complexos P-QRS do eletrocardiograma. Apenas dez componentes foram consideradas para a representação dos sinais, por explicarem 99,2% da variância total dos complexos. A importância deste fato pode ser melhor avaliada observando-se o poder de reconstrução dos sinais destas primeiras CPs. A partir dos exemplos apresentados verifica-se que as características morfológicas de cada complexo estão fortemente concentradas nas primeiras CPs. O resíduo após a extração da 4ª CP já é pouco significativo, alterando-se muito pouco com a contribuição de maior número de componentes.

Foi apresentado o resultado de classificação somente para os sinais de um paciente do "MIT-BIH Arrhythmia Database". Devido ao grande volume de dados - intervalo RR e mais 6 componentes de 75.094 complexos P-QRS - não é possível determinar, subjetivamente, um método de classificação que seja aplicável indistintamente a todo o conjunto de pacientes. Uma alternativa para investigação futura consiste em ampliar a aplicação da regressão logística. A partir desta análise, utilizando-se uma técnica de regressão passo-a-passo ("Stepwise"), é possível selecionar as componentes que fornecem melhor discriminação, e a partir desta seleção estabelecer os métodos de classificação mais adequados.

Os resultados obtidos, não conclusivos, são bastante promissores. Cabe destacar que os índices observados para os batimentos de fusão, inferiores, são justificáveis, pois estes complexos apresentam formas intermediárias entre o batimento normal e a extra-sístole ventricular. Em tais batimentos, os ventrículos são ativados por duas frentes de onda simultâneas: uma advinda do nóculo sinusal, que alcançou o ventrículo, e outra de foco

ectópico, disparada antes da chegada da primeira (Carneiro, 1983).

Do presente estudo prospectivo, alguns pontos positivos podem ser destacados: 1) a técnica é efetiva em termos de redução de dimensionalidade; 2) os diagramas de espalhamento, mesmo com apenas duas CPs, mostram separação entre complexos distintos; e 3) a utilização da técnica para a extração automática de parâmetros de complexos P-QRS é promissora, particularmente porque estes parâmetros levam em conta todo o complexo, são ortogonais entre si, e sua extração não depende de técnicas subjetivas.

AGRADECIMENTOS

Os autores agradecem à FINEP pelo suporte ao projeto, e a Carlos Alberto C. Pires pelo apoio na digitalização dos sinais.

REFERÊNCIAS

- CARNEIRO, E.F. (1983), O Eletrocardiograma, Livraria Ateneu, Rio de Janeiro, 3ª edição.
- DIXON, W.J., BROWN, M.B., ENGELMAN, L., FRANE, J.W., HILL, M.A., JENRICH, R.I., TOPOREK, J.D. (1985), BMDP Statistical Software, University of California Press, Berkeley.
- ECRI (1985), "Arrhythmia Monitoring Systems", Health Devices, Emergency Care Research Institute, August, pages 287-327.
- FELDMAN, C.L., AMAZEEN, P.G., KLEIN, M.D. et alii (1971), "Computer Detection of Ventricular Ectopic Beats", Computers and Biomedical Research, Number 3, pages 666-674.
- FLORENZ, M.K., ROLNITSKY, L.M., BIGGER, Jr., J.T. (1975), "A Rapid ECG Processing Computer Program Using the Finite State Machine Approach", Computers in Cardiology 1984, IEEE Computer Society, pages 145-150.
- FLOSI, J.O.C. (1988), Sistema de Classificação Automática dos Complexos QRS do Eletrocardiograma, Tese de Mestrado, Programa de Engenharia Biomédica, COPPE/UFRJ, Rio de Janeiro .
- HENKIND, S., TEICHOLZ, L., HARRISON, M. (1985), "Intensive Care Unit Monitoring Using a Real-Time Expert System", Computers in Cardiology 1984, IEEE Computer Society, pages 7-12.
- JARISCH, W.R., MARK, R.G. (1985), "A Stochastic Modeling Approach to Cluster Identification with Application to Arrhythmia Detection", Computers in Cardiology 1984, IEEE Computer Society, pages 185-188.
- LIMA, C.E.G. (1986), Aperfeiçoamento e Avaliação de um Algoritmo para Detecção e Delineação, em Tempo Real, de Complexos QRS, Tese de Mestrado, Programa de Engenharia Biomédica, COPPE/UFRJ, Rio de Janeiro.
- LOWN, B., WOLF, M. (1971), "Approaches to Sudden Death from Coronary Heart Disease",

- Circulation, Volume XLIV, pages 130-142.
- MEAD, C.N., PULL, H.R., CHENG, J.-S., CLARK, K.W., THOMAS, L.J. (1982), "A Frequency-Domain-Based QRS Classification Algorithm", Computers in Cardiology 1981, IEEE Computer Society, pages 351-354.
- MIT-BIH (1982), Arrhythmia Data Base: Tape Directory and Format Specification, Harvard University, Massachusetts Institute of Technology, Biomedical Engineering Center for Clinical Instrumentation, Massachusetts.
- MORRISON, D.F. (1976), Multivariate Statistical Methods, McGraw- Hill Book Company, New York.
- NOLLE, F.M. (1972), ARGUS. A Clinical Computer System for Monitoring Electrocardiographic Rhythms, D.Sc. Dissertation, Washington University, Saint Louis.
- PANERAI, R.B., FERREIRA, A.L.A.S., BRUM, O.F. (1988), "Principal Component Analysis of Multiple Noninvasive Blood Flow Derived Signals", IEEE Transaction on Biomedical Engineering, Volume BME- 35, Number 7, pages 533-538.
- PEREIRA, W.C.A., SCHLINDWEIN, F.S., LIMA, C.E.G. (1988), "Algoritmo para Detecção e Delimitação de Complexos QRS em Tempo-Real, Utilizando duas Derivações de ECG", Revista Brasileira de Engenharia - Caderno de Engenharia Biomédica, Volume 4, Número 2, páginas 5-24.
- TOU, J.T. e GONZALES, R.C. (1974), Pattern Recognition Principles, Addison-Wesley, London.

PRELIMINARY STUDY OF THE CLASSIFICATION OF CARDIAC ARRHYTHMIAS BY MEANS OF PRINCIPAL COMPONENT ANALYSIS

ABSTRACT -- A new technique for the classification of P-QRS complexes of the electrocardiogram is proposed, foreseeing application in the automatic monitoring of cardiac arrhythmias. Using data from 32 patients of the "MIT-BIH Arrhythmia Database", 200 electrocardiographic segments, containing the P-QRS waves, are selected as representative patterns of different cardiac contractions. Principal component analysis of this data set indicates that 99.2% of the total variance can be represented by only 10 components. Thus, signals reconstructed using only these 10 components are observed to preserve shape. Logistic regression is applied to construct discriminant functions for the classification of cardiac contractions. The results achieved in one patient taken from the database give correct classification of 99.1% (normal beats), 98.3% (ventricular premature contractions) and 91.4% (fusion beats).