

CLASSIFICAÇÃO DE PADRÕES AUDITIVOS USANDO MÁQUINAS DE VETOR DE SUPORTE E REDES NEURAS ARTIFICIAIS

M. J. S. Pedrosa*, D. M. Dourado* e L. B. Felix*

*Núcleo Interdisciplinar de Análise de Sinais - NIAS/Departamento de Engenharia Elétrica/UFV, Viçosa, Brasil

matheus.pedrosa@ufv.br

Resumo: A Atenção Seletiva Auditiva refere-se à capacidade mental de resistir a estímulos distratores e focar em informações relevantes. A quantificação dos estados de atenção seletiva por meio do Eletroencefalograma cria um paradigma para o uso desta habilidade, inerente ao ser humano, como meio para uma Interface Cérebro Computador. Como classificador dos estados da atenção, foram utilizadas Redes Neurais Artificiais (RNAs) e Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machines*). Os resultados mostraram que para estimulação nas frequências moduladoras 32/38 Hz e portadoras 500/2000 Hz, as RNAs expressam taxa de acerto médio de 87,5% contra 85,71% do SVM. Por outro lado, os estudos mostram que o melhor classificador a ser usado quando se trata de todas as combinações de frequências é o SVM, que expressa uma taxa de acerto médio de 53,53% contra 51,5% da RNA.

Palavras-chave: EEG, SVM, Atenção Seletiva Auditiva, Resposta Evocada em Regime Permanente, RNA.

Abstract: *The auditory selective attention refers to the mental capacity to resist distracting stimuli and to focus on relevant information. The quantification of the states of selective attention, by means of Electroencephalogram, created a paradigm for the use of this ability that is inherent to humans, as a means for Brain Computer Interface. To classify the states of attention, Artificial Neural Network (ANN) and Support Vector Machines (SVM) were used. The results showed that stimulation using the modulate frequency 32/38 Hz and carrier frequency 500/2000 Hz, the ANN expressing average hit rate of 87.5% to 85.71% of the SVM. Furthermore, the studies show that the best classifier to be used when it comes to all frequency combinations is Support Vector Machine (SVM), which expresses a medium hit rate of 53.53% to 51.5% of the ANN.*

Keywords: EEG, SVM, Auditory Selective Attention, Auditory Steady-State Response, RNA.

Introdução

Ao registro eletrográfico de potenciais bioelétricos no escalpo dá-se o nome de Eletroencefalograma (EEG). São de especial interesse os sinais de EEG coletados concomitantemente à apresentação de estímulos sensoriais, pois tais sinais carregam

informações acerca da via neural e dos núcleos cerebrais ativadas no processamento da informação sensorial – este potencial elétrico relacionado a eventos é conhecido como Resposta Evocada (RE). As REs são usadas de várias formas na clínica médica, por exemplo: audiometria usando respostas auditivas em regime permanente (ASSRs, do inglês *Auditory Steady-State Response*). Uma ASSR ocorre quando um som é apresentado a uma taxa suficientemente elevada, de modo que a resposta ao estímulo atual sobreponha-se à resposta ao estímulo anterior [1].

A audição seletiva, isto é, focando atenção em um estímulo dentre dois, causa uma alteração quantificável na resposta medida [2]. Desta forma, a implementação de uma Interface Cérebro-Computador (ICC) auditiva poderia ser alcançada através da apresentação de sons diferentes ao indivíduo e coleta de sinais EEG [2]. O problema de classificação de padrões pode ser resolvido utilizando Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machines*) - a qual é fundamentada na Teoria da Aprendizagem Estatística [3] - ou por Rede Neural Artificial (RNA) - que é um processador constituído de unidades de processamento simples, que têm propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso [4].

O objetivo é desenvolver um sistema de classificação de padrões a partir de uma ICC, explorando os efeitos de atenção nas ASSRs no qual o usuário faça escolhas binárias. Este sistema faz a classificação, a partir dos classificadores SVM e RNA, dos sinais EEG registrados em resposta a dois estímulos auditivos simultâneos. Para comunicar a decisão, o indivíduo deve focar a atenção em um dos dois estímulos auditivos apresentados.

Materiais e Métodos

Estímulos – A estimulação foi aplicada por meio de sons modulados em amplitude (AM), gerados de acordo com a Equação 3 [5].

$$c(t) = A * \text{sen}(2\pi f_c t) \quad (1)$$

$$m(t) = 1 - a * 0.5 * (\cos(2\pi f_m t) + 1) \quad (2)$$

$$s(t) = c(t) * m(t) \quad (3)$$

Onde f_c é a frequência portadora, a é a profundidade de modulação, A é a amplitude máxima do sinal e f_m é a frequência moduladora.

Os estímulos foram gerados digitalmente, utilizando *Matlab*, com uma frequência de amostragem de 44100 Hz. Neste trabalho, utilizou-se uma profundidade de modulação de 100%, pois segundo [6] esse valor evoca a ASSR de maior amplitude. A amplitude máxima (A) foi ajustada para que a intensidade sonora de cada tom AM fosse de 75 dB SPL.

Fones de ouvido – Os níveis de tensão elétrica para calibração do fone, de acordo com o nível e intensidade sonora desejada, foram fornecidos pelo Núcleo de Estudos e Pesquisas em Engenharia Biomédica da Universidade Federal de Minas Gerais (NEPEB), onde foi feita a calibração do equipamento utilizando um ouvido artificial acoplado a um medidor de níveis sonoros.

Protocolo de estimulação – No experimento foi necessário à realização de 114 exames, distribuídos em 14 combinações distintas de frequências. As combinações de frequências estimuladas foram definidas de maneira que se pudessem diferenciar os ruídos estimulados, logo facilitaria que o paciente conseguisse focar a atenção. Todos os sinais foram coletados de jovens entre 18 e 24 anos e sem problemas auditivos aparentes, sendo todos eles postos sentados em uma cabine audiométrica. Os estímulos gerados pela ferramenta GUI (*Graphical User Interface*) do *Matlab* foram levados para dentro da cabine através de um cabo blindado. O protocolo experimental foi aprovado pelo comitê de ética em pesquisa com seres humanos da Universidade Federal de Viçosa (UFV). A Tabela 1 mostra as frequências de modulação utilizadas no trabalho.

Tabela 1: Parâmetros de estimulação AM.

Frequência moduladora (Hz)	Frequência Moduladora Corrigida (Hz)
32	31.132
34	34.656
36	35.831
38	39.356
41	41.705

Os parâmetros foram definidos a fim de provocar a redução do espalhamento espectral, para isso foi utilizado o método de correção da frequência de estimulação aplicado *a priori*. Tal correção é feita a partir do ajuste da frequência de modulação do estímulo, de tal forma que cada janela de dados contenha número primo de ciclos do sinal, ou seja, a frequência do estímulo é ajustada para garantir um número primo de oscilações em um trecho de comprimento fixo, como mostra a Tabela 1. Durante a estimulação variou-se diversas combinações de frequências moduladoras e frequências portadoras de 500, 1000 e 2000 Hz.

Os estímulos foram executados na faixa de 30 a 40Hz, pois nestas frequências as ASSR apresentam as maiores respostas, além de serem mais influenciadas pelo estado de atenção [5, 7, 8]. Durante as etapas de estimulação os voluntários foram orientados a manter os olhos fechados, a fim de facilitar a concentração e evitar artefatos oculares [9]. Cada etapa de estimulação teve dois minutos e intervalos entre estimulações de trinta segundos. Antes de cada etapa na qual o indivíduo tivesse que focar a atenção em um determinado estímulo, era estimulada apenas a orelha (monoaural) no lado do foco de atenção durante dez segundos, a fim de instruir e facilitar o indivíduo na tarefa cognitiva. A evolução das etapas de estimulação é ilustrada na Figura 1.

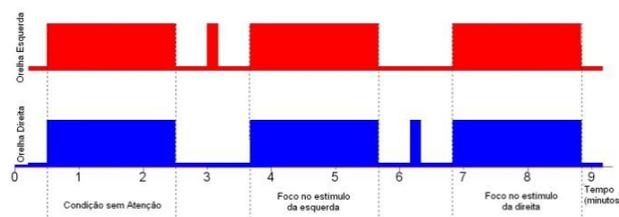


Figura 1: Linha do tempo para os estímulos auditivos. Estímulo da orelha esquerda (superior) e orelha direita (inferior), com frequência das portadoras de 500Hz e 2000Hz, respectivamente.

Aquisição de sinais – Para coletar os sinais de EEG foi utilizado o Sistema Internacional 10-20. Os sinais foram coletados usando o amplificador de sinais biológicos (*BrainNet* BNT-36), com filtro passa-alta e passa-baixa de 0,1 e 100 Hz, respectivamente, e frequência de amostragem de 601,5 Hz. Foram utilizados os canais T3, F3, P3, O1, T4, P4, O2, Fz e Pz do Sistema Internacional 10-20, os quais obtiveram uma melhor resposta segundo [10].

Processamento do sinal – Os sinais são condicionados a processos de filtragem e remoção de artefatos. A cada trecho adquirido será aplicado um filtro passa-faixa, observando a frequência da moduladora, bem como um filtro rejeita-faixa de 60 Hz, para remover a componente espectral da rede elétrica, que geralmente apresenta elevado nível de potência [10].

Subsequentemente a rejeição de artefatos, fez-se uma etapa de subtração dos sinais, onde se tomou como referência a situação em que o indivíduo não estivesse focando a atenção a qualquer estímulo.

Redução de dimensionalidade – O método utilizado foi a análise das componentes principais (PCA, do inglês *Principal Component Analysis*), que tem por finalidade básica a análise dos dados visando sua redução, eliminação de sobreposições e a escolha das formas mais representativas dos mesmos a partir de combinações lineares das variáveis originais. Trabalhou-se com até quatro componentes principais, uma vez que analisando a Figura 2, percebe-se que essas componentes compreendem cerca de 95 % da variabilidade dos dados.

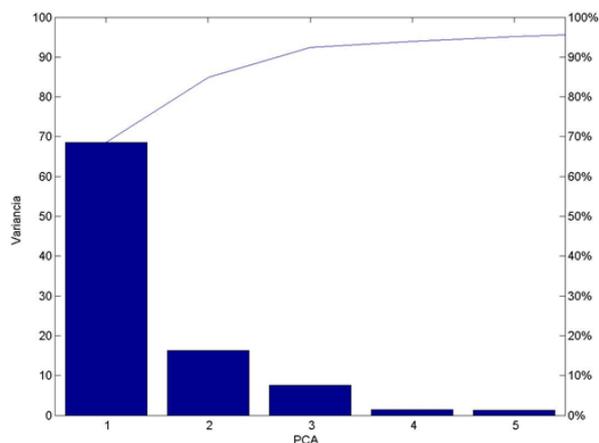


Figura 2: Gráfico da variância por componente principal e da variância acumulada.

Extração de características – O modelo autoregressivo (AR, do inglês *Autoregressive*) é um modelo estocástico e útil na representação de um grande número de séries temporais. Tal técnica foi aplicada na representação temporal a partir dos valores das componentes principais, e desenvolveram-se modelos desde a quarta até a décima ordem.

Classificação de padrões – Foram abordados dois classificadores distintos, SVM e RNA, em ambos tentase diferenciar a atenção ao estímulo esquerdo da atenção ao estímulo direito.

A fim de possibilitar meios para comparação com o SVM, adotou-se uma composição de RNA simples, *feedforward network*, cuja função de formação foi baseada na otimização de Levenberg-Marquardt. No algoritmo da RNA, fez sucessivas simulações, adotando sempre uma camada neural escondida e variando-se o número de neurônios que a compunham para obtenção do melhor resultado.

Determinou-se que as entradas de ambos os classificadores seriam os coeficientes obtidos a partir da utilização do modelo AR, sendo 70% do conjunto de dados para treinamento e 30% para validação. Durante o treinamento foi realizado o processo de validação cruzada. O método escolhido foi o método das k-partições (do inglês *K-fold Cross-Validation*), a qual é uma ferramenta útil no processo de mineração dos dados.

Resultados

A Tabela 2 ilustra os resultados obtidos na classificação da atenção esquerda/direita usando SVM e utilizando todas as combinações de frequências existentes no conjunto de dados.

Tabela 2: Taxa de acerto percentual no conjunto de validação na classificação do SVM, com suas respectivas variações do modelo AR e PCAs, utilizando-se 30% dos dados para validação com todas as combinações de frequências.

Componente Principal / Ordem	Taxa de Acerto (%)						
	4	5	6	7	8	9	10
1 ^a	50	60,61	48,48	56,06	53,03	48,48	48,48
1 ^a e 2 ^a	50	50	63,64	56,06	53,03	57,58	51,52
1 ^a , 2 ^a e 3 ^a	50	50	72,73	51,52	51,52	50	50
1 ^a , 2 ^a , 3 ^a e 4 ^a	53,03	53,03	62,12	50	50	58	50

A RNA com o melhor desempenho é composta por apenas uma camada escondida, sendo constituída de quatro neurônios artificiais, cuja função transferência é a hiperbólica sigmoide, seguida de uma camada de saída com um neurônio, que foi ativada por uma função de transferência linear.

Na Tabela 3 observam-se os resultados obtidos na classificação do foco de atenção usando RNA e utilizando todas as combinações de frequências existentes no conjunto de dados.

Tabela 3: Taxa de acerto percentual no conjunto de validação na classificação da RNA, com suas respectivas variações do modelo AR e PCAs, utilizando-se 30% dos dados para validação com todas as combinações de frequências.

Componente Principal / ordem	Taxa de Acerto (%)						
	4	5	6	7	8	9	10
1 ^a	56.1	53	59.1	51.5	50	50	50
1 ^a e 2 ^a	54.5	53	62.1	69.7	54.5	59.1	57.6
1 ^a , 2 ^a e 3 ^a	40.9	56.1	63.6	60.6	51.5	57.6	57.6
1 ^a , 2 ^a , 3 ^a e 4 ^a	54.5	62.1	66.7	48.5	54.5	57.6	53

A Tabela 4 ilustra os resultados obtidos na classificação da atenção esquerda/direita usando SVM. Esses resultados ilustram a combinação das frequências moduladoras iguais a 32/38 Hz e portadoras iguais a 500/2000 Hz.

Tabela 4: Taxa de acerto percentual no conjunto de validação na classificação do SVM, com suas respectivas variações do modelo AR e PCAs, utilizando-se 30% dos dados para validação.

Componente Principal / Ordem	Taxa de Acerto (%)						
	4	5	6	7	8	9	10
1 ^a	50	100	100	75	75	75	50
1 ^a e 2 ^a	75	100	100	100	100	75	75
1 ^a , 2 ^a e 3 ^a	100	100	100	75	100	75	100
1 ^a , 2 ^a , 3 ^a e 4 ^a	75	100	75	100	75	75	100

Na Tabela 5 observam-se os resultados obtidos na classificação do foco de atenção usando RNA. Esses resultados ilustram a combinação das frequências moduladoras iguais a 32/38 Hz e portadoras iguais a 500/2000 Hz

Tabela 5: Taxa de acerto percentual no conjunto de validação na classificação da RNA, com suas respectivas variações do modelo AR e PCAs, utilizando-se 30% dos dados para validação.

Taxa de Acerto (%)							
Componente Principal / Ordem	4	5	6	7	8	9	10
1 ^a	75	50	75	50	100	50	100
1 ^a e 2 ^a	100	100	75	50	100	50	100
1 ^a , 2 ^a e 3 ^a	100	100	100	100	100	75	100
1 ^a , 2 ^a , 3 ^a e 4 ^a	100	100	100	100	100	100	100

Discussão

Nos resultados expressos na Tabela 4 e Tabela 5, notam-se expressivos valores percentuais de acerto quanto ao foco de atenção, denotando assim que em ambas as situações propostas, os classificadores foram capazes de contornar a dificuldade de generalizar. Observaram-se níveis de aprendizados consistentes, acarretando assim em significativas taxas percentuais de acerto para o grupo de amostras usado na validação.

Abrangendo a análise, tentou-se classificar os padrões auditivos utilizando todo o conjunto amostral. Foi notório o desempenho inferior ao obtido na situação de apenas uma combinação de frequências. Tais resultados observados na Tabela 2 e Tabela 3 permitem justificar a dificuldade quanto ao processo de aprendizado de ambos os classificadores, uma vez, trabalharam com um maior volume de dados.

Para fins comparativos calculou-se a taxa de acerto médio de cada classificador. Essa taxa foi calculada pelo somatório da taxa de acerto de cada configuração, ou seja, das variações do modelo AR e do número de componentes principais. Essa grandeza ajudou a mensurar a capacidade de classificação de cada ferramenta para as situações propostas. Inicialmente para a situação de estimulação com 32/38 Hz de moduladora e 500/2000 Hz de portadora, observa-se ligeira sobreposição da RNA em relação ao SVM, tendo elas 87,5% e 85,71% de acerto percentual médio respectivamente, mostrando assim uma facilidade maior da RNA nesse contexto de classificação. Já para a situação de maior número de dados, o SVM apresenta 53,53% de acerto percentual médio contra 51,5% da RNA, revelando assim que no contexto experimental abordado nesse trabalho, o SVM mostrou-se com maior capacidade de aprendizado quando submetido a um maior volume dados.

Conclusão

Note que para as distintas situações propostas a capacidade de reconhecimento de padrões auditivos de ambos os classificadores foram muito semelhantes, entretanto qualitativamente diferentes conforme a maior dificuldade de generalização. De um modo geral, tanto SVM quanto RNA podem ser usadas como ferramentas úteis no estudo de padrões auditivos. No entanto vale ressaltar a necessidade de um estudo coeso do problema para se decidir o melhor método a ser usado, e assim propiciar a obtenção do melhor resultado possível dentro do contexto prático.

Agradecimentos

Agradecemos a FAPEMIG, CNPQ e CAPES pelo apoio financeiro.

Referências

- [1] Dimitrijevic, A; John, M. S; Van Roon, P; Purcell, D. W; Adamonis, J; Ostroff, J; Nedzelski, J. M and Picton, T. W (2002). “Estimating the audiogram using multiple auditory steady-state response.” *J. Am. Acad. Audiol.*, v.13, p. 205- 224.
- [2] S. A. Hillyard, R. F. Hink, V. L. Schwent, and T. W. Picton (1973). Electrical signs of selective attention in the human brain. *Science*.
- [3] Vapnik, V., (1995). The nature of statistical learning theory. New York: Springer- Verlag.
- [4] Haykin, S., (2002). *Redes Neurais, Princípios e Prática*. Bookman.
- [5] John M. S., Lins O. G. e Boucher B. L. (1998) “Multiple Auditory Steady-State Responses (MASTER): Stimulus and Recording Parameters.” *Ear and Hearing*, v. 37 p. 59-82.
- [6] John M. e Picton T. W. (2000), “Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements” *Hearing Research*, v. 14,1 p. 57-79.
- [7] Linden R. D., Campbell K. B. e Hamel G.. (1985) “Human Auditory Steady-State Evoked Potentials during Sleep” *Ear and Hearing*, v. 6, p. 167-174.
- [8] Ross B., Draganova R. e Picton T. W. (2003) “Frequency specificity of 40-Hz auditory Steady-state Response”, *Hearing Research*, v. 186, p. 1-4.
- [9] Kim D. W., Hwang H. J. e Lim J. H. (2011) “Classification of selective attention to auditory stimuli: Toward vision-free brain computer interfacing” *Journal Neuroscience Methods*, v. 197. - p. 180-185.
- [10] Ranaudo F. S. (2012), *Atenção Seletiva Auditiva Usando Potenciais Evocados em Regime Permanente e Coerência Espacial*, Dissertação de Mestrado. UFRJ, Rio de Janeiro.