



Estudo piloto sobre identificação de indivíduos a partir de dados de EEG por meio de conectividade funcional utilizando sincronização por motifs

Palavras-Chave: biometria, eletroencefalografia, motifs

Manuela Von Ah Davanço [IFGW - UNICAMP]

Marina Cabral de Paulo [IFGW - UNICAMP]

Marcela Stracieri Janchevis Preiss [INSTITUTO PRINCIPIA]

Prof.^a Dr.^a Gabriela Castellano (orientadora) [IFGW - UNICAMP]

INTRODUÇÃO:

A eletroencefalografia (EEG) é uma das técnicas mais antigas para a investigação da atividade cerebral [1]. Ela permite medir a atividade elétrica dos neurônios por meio da colocação de eletrodos diretamente sobre o escalpo de um indivíduo. Além disso, esta técnica possui as vantagens de ter relativamente baixo custo, ser portátil e ter uma ótima resolução temporal (da ordem de milissegundos). Recentemente, a possibilidade de usar a EEG para produzir uma assinatura do indivíduo, isto é, para fins biométricos, tem sido estudada em vários trabalhos [2]–[4]. A maioria desses trabalhos utilizou medidas extraídas diretamente das séries temporais de eletrodos específicos, como a densidade espectral de potência [2], [5], [6].

No entanto, cada vez mais vem sendo explorado o fato de que regiões cerebrais atuam em conjunto para exercer tarefas cognitivas, sensoriais ou motoras. Isso levou ao estudo de redes cerebrais funcionais, ou da conectividade funcional [7]. A conectividade funcional é calculada comparando a atividade de pares de regiões cerebrais, utilizando alguma medida de similaridade [8]. Várias medidas de similaridade em sido exploradas no contexto de uso da EEG para biometria, como a coerência espectral [4], a correlação de Spearman aplicada à transformada de Hilbert das séries temporais [9], o valor de bloqueio de fase ou *phase locking value* [10], a parte imaginária do *phase locking value* [11], o índice de atraso de fase ou *phase lag index* [12], e informação mútua aplicada a padrões ordinais [13], entre outras.

O objetivo do presente trabalho foi utilizar da medida de sincronização por motifs [14], também baseada em padrões ordinais [15], para o cálculo de redes cerebrais funcionais a partir de dados de EEG, e explorar a viabilidade de usar estas redes para fins biométricos.

METODOLOGIA:

Foi utilizado um banco de dados online chamado “EEG Motor Movement/ Imagery Dataset” (<https://physionet.org/content/eegmmidb/1.0.0/>) para obter-se os dados de EEG usados [16], [17].

Neste banco, encontram-se dados de 109 indivíduos, cada um com 14 aquisições de EEG, sendo estas feitas com diferentes tarefas. Para este trabalho piloto, foram utilizadas apenas as aquisições de repouso, sendo uma com olhos abertos e outra com olhos fechados, de 20 indivíduos.

Inicialmente, foi realizada uma limpeza dos dados a fim de reduzir artefatos e ruídos. Para isso, foi feito o uso da plataforma MATLAB em conjunto com o software EEGLab, o qual permite a análise de dados de EEG. No software, primeiramente, foi possível realizar uma limpeza visual dos dados através da plotagem desses sinais em gráficos, removendo artefatos identificados, por exemplo, pela grande diferença de amplitude do sinal em certos trechos. Em seguida, ainda no software, foi utilizada a ferramenta de *independent component analysis* (ICA), onde é possível analisar componentes do sinal de EEG separadamente, identificando artefatos como piscadas de olhos ou movimentos musculares, removendo-as posteriormente. O ICA também foi utilizado para retirar a componente da banda α de frequência dos sinais (8-12 Hz), nas aquisições de olho fechado, pois é uma banda característica de quando se fecham os olhos [18], e esta poderia trazer diferenças ao comparar os sinais de olhos abertos com os de olhos fechados. Também foi utilizado o método de *common average referencing* (CAR), implementado na plataforma MATLAB, para remover artefatos comuns a todos os eletrodos. O método consiste em calcular uma média entre os sinais de todos os eletrodos, e então subtrair esse valor do sinal de cada eletrodo.

Após este pré-processamento dos dados, foi utilizado o método de sincronização por motifs [14] para a análise dos sinais. O método foi implementado no MATLAB. Uma série de motifs consiste em uma série de padrões de comportamento que podem ser extraídos de um sinal de EEG, como pode-se ver na Figura 1.

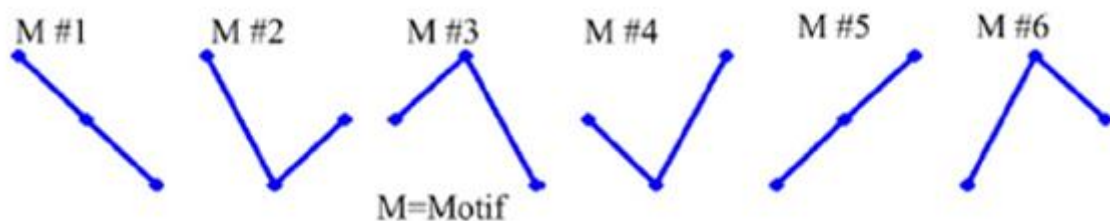


Figura 1: Conjunto de possíveis padrões presentes no sinal de EEG para motifs de três pontos (grau 3), utilizados neste trabalho. Figura extraída de [19].

Desta forma, a série temporal de cada eletrodo no sinal de EEG pode ser “traduzida” para uma série de motifs (neste trabalho foram utilizados motifs de 3 pontos, como na Figura 1). Com isso, pode-se utilizar as séries de motifs de dois eletrodos diferentes para analisar quão parecidos são os comportamentos destas séries, para diferentes tempos de atraso (t). Neste trabalho, não foram usados atrasos (ou seja, foi usado $t = 0$ para todas as análises).

Matematicamente, a comparação sem atraso de duas sequências A e B de motifs é dada por:

$$c_{AB} = \sum_{i=1}^{L_M} J_i$$

em que $J_i = 1$ se $M_{A_i} = M_{B_i}$ e $J_i = 0$ caso contrário, e L_M é o comprimento (número de amostras) da série de motifs. Desta forma, cada vez que o mesmo motif é encontrado em ambas as séries, atribui-se o valor 1 ao coeficiente J correspondente àquela interação. Para um motif de grau n , o comprimento L_M será dado, em termos do comprimento L_0 da série de EEG original, por:

$$L_M = L_0 - (n - 1)$$

Finalmente, o grau de sincronização Q_{AB} entre as séries, que varia entre 0 e 1, é definido como:

$$Q_{AB} = \frac{c_{AB}}{L_M}$$

Assim, a sincronização por motifs permite a construção de uma matriz de conectividade $N \times N$ (em que N é o número de eletrodos utilizados para a aquisição), na qual cada elemento é o grau de sincronização Q_{AB} entre duas séries temporais de EEG. Neste trabalho, os dados de EEG utilizados foram adquiridos através de 64 eletrodos, portanto obtiveram-se matrizes de conectividade de 64×64 .

A fim de verificar a possibilidade de utilizar esses sinais de EEG em um sistema de biometria, calcularam-se as matrizes de conectividade para trechos de 1 s dos sinais, sendo que, na aquisição de olhos abertos utilizaram-se o décimo e penúltimo segundo da série, e na aquisição de olhos fechados apenas o penúltimo segundo. Primeiramente, compararam-se as matrizes dentro do mesmo sinal (décimo e penúltimo segundo da aquisição de olhos abertos) e então, matrizes de aquisições diferentes (penúltimo segundo tanto de olhos abertos quanto fechados).

Para estas comparações, foi calculada a distância euclidiana entre as respectivas matrizes, dada por:

$$d_{XY} = \sqrt{\sum_{i=1}^N \sum_{j=1}^N (X_{i,j} - Y_{i,j})^2}$$

onde $X_{i,j}$ e $Y_{i,j}$ são, respectivamente, o elemento i, j da matriz de conectividade dos sujeitos X e Y .

RESULTADOS E DISCUSSÃO:

Na Tabela 1, temos as distâncias para a comparação entre aquisições diferentes (olhos abertos e fechados), e na Tabela 2 temos as porcentagens de acertos na identificação dos indivíduos. Note que S_n são os indivíduos a serem comparados. A tabela para as distâncias da comparação entre trechos da mesma aquisição (olhos abertos) é similar à Tabela 1.

Note que, na comparação de trechos diferentes de uma mesma aquisição, obtivemos uma porcentagem de acertos de 100%, significativamente maior que a porcentagem da comparação de aquisições diferentes, 35%. Essa diferença entre as porcentagens era esperada, pois na primeira comparação estamos analisando trechos diferentes, mas de um mesmo sinal, portanto estes

possuem maior similaridade. Já na segunda comparação, analisamos trechos de sinais diferentes, portanto é esperado que estes possuam mais diferenças entre si.

	S01	S02	S03	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S22
S01	20,3	37,2	78,7	201,3	61,5	103,9	65,2	161,6	62,0	60,5	180,8	34,4	80,2	49,5	114,9	122,9	52,5	38,0	61,7	105,8
S02	32,3	29,2	79,6	175,9	64,2	98,6	57,3	155,8	54,2	53,9	195,4	36,8	71,2	45,9	111,3	121,5	51,7	37,7	57,0	102,6
S03	43,7	52,1	64,1	207,9	69,9	140,8	62,0	135,6	40,2	51,3	189,6	42,7	57,4	43,9	104,9	109,7	73,8	45,8	53,7	87,7
S05	200,2	193,2	243,4	38,1	225,3	152,4	226,4	334,4	207,3	209,3	376,4	195,6	225,8	201,1	286,8	256,0	191,4	193,6	208,2	271,7
S06	60,7	55,6	99,5	178,8	55,6	63,3	86,3	191,1	71,5	74,8	230,9	64,8	104,5	60,8	146,8	128,3	71,4	49,8	68,8	130,3
S07	49,1	60,1	115,2	198,9	74,9	89,2	87,9	190,4	87,4	79,4	225,5	55,3	103,4	77,9	143,7	146,4	63,9	55,5	82,9	138,5
S08	40,2	35,7	67,5	186,8	67,1	108,4	37,5	143,4	41,8	48,1	201,4	38,7	60,9	39,6	104,5	113,3	60,6	38,8	48,7	86,4
S09	196,8	202,4	213,2	330,0	213,9	281,8	195,5	263,6	182,1	202,1	352,5	189,6	189,3	182,7	228,6	244,3	232,4	195,0	200,1	226,1
S10	54,4	58,4	82,9	207,4	80,0	133,7	62,4	149,9	39,8	55,2	210,8	49,9	64,8	51,3	107,6	122,3	77,3	50,3	59,2	100,9
S11	67,3	53,1	88,1	212,4	98,8	114,1	89,3	167,2	66,0	50,4	212,0	68,3	98,9	67,3	143,8	119,0	82,4	57,5	61,4	109,3
S12	63,6	71,0	123,7	223,3	100,3	145,0	98,1	197,6	88,0	91,3	192,6	65,4	107,6	70,6	152,3	158,2	84,9	70,9	101,4	140,9
S13	45,5	41,9	72,4	190,3	73,8	113,0	62,9	142,6	42,2	55,0	217,6	41,3	66,9	38,3	111,8	112,3	72,2	36,4	50,6	92,7
S14	52,4	51,6	77,0	209,1	79,1	132,2	65,1	146,5	44,5	48,4	213,4	48,3	56,7	46,0	113,2	120,2	78,0	47,0	58,3	93,8
S15	39,5	39,1	76,8	196,1	62,8	108,4	58,4	158,4	50,0	53,9	191,6	37,9	74,3	40,4	114,4	124,7	66,0	29,7	56,1	99,3
S16	45,6	53,1	83,7	197,9	71,2	118,6	68,3	150,0	54,9	59,4	225,1	44,0	66,3	50,4	107,2	124,9	72,3	44,5	64,1	104,4
S17	96,5	86,3	89,0	234,3	116,0	178,7	89,5	164,0	64,0	90,8	264,0	87,6	88,9	72,7	146,1	133,0	127,5	76,2	84,5	103,3
S18	36,7	37,9	81,5	199,6	71,0	104,0	68,8	163,6	55,1	58,6	204,8	40,0	78,3	44,2	128,2	124,4	45,3	34,8	57,7	103,8
S19	36,6	48,3	73,3	202,9	62,3	131,4	61,7	152,7	56,0	66,9	205,9	39,9	73,2	46,4	118,9	124,8	78,9	33,0	68,9	100,8
S20	51,5	50,2	67,4	161,7	73,2	96,0	59,5	146,3	47,4	55,5	224,2	44,0	60,7	39,9	116,1	120,7	73,9	40,7	46,4	94,3
S22	35,1	38,9	56,9	182,3	62,9	111,4	51,3	131,7	35,6	48,2	202,2	37,8	54,1	34,6	102,7	104,1	65,4	32,6	46,2	74,9

Tabela 1: Distâncias euclidianas calculadas para a comparação entre aquisições diferentes. Em verde, os resultados que foram corretos (tiveram a menor distância euclidiana na comparação de trechos de sinal do mesmo indivíduo) e em vermelho, os resultados que apresentaram erro (menor distância não corresponde à comparação de sinal do mesmo indivíduo).

Trechos diferentes do mesmo sinal	Trechos de sinais diferentes
100%	35%

Tabela 2: Porcentagens de acertos na biometria de cada comparação.

Com isso, pode-se notar que, em se tratando da comparação de uma mesma aquisição de sinal, foi possível identificar os indivíduos através dos métodos utilizados. Porém, se tratando de uma comparação entre sinais de aquisições diferentes, obtivemos uma porcentagem mais baixa e, portanto, não foi possível identificar os indivíduos adequadamente neste caso.

Algumas possíveis razões para a segunda comparação não ter ocorrido como esperado, envolvem, por exemplo, o fato de ter-se retirado a componente da banda alfa dos dados obtidos com olhos fechados, mas não dos com olhos abertos, aumentando as diferenças entre estes sinais. O número de indivíduos comparados (20) pode ter sido insuficiente para se obter um resultado mais robusto, assim como o intervalo de tempo escolhido para cada sinal (1 s) pode ter sido muito pequeno, ou até muito grande.

CONCLUSÕES:

Após pré-processar os dados de EEG com ajuda do software EEGLab, foi possível aplicar o método da sincronização por motifs e compararam-se as matrizes de conectividade através do cálculo da distância euclidiana, obtendo-se resultados satisfatórios para a comparação entre uma mesma aquisição de sinal, mas não tão satisfatórios na comparação entre aquisições diferentes.

Um próximo passo neste trabalho seria filtrar os dados de diferentes aquisições de forma igual, evitando criar outras diferenças entre estes, e aumentar o espaço amostral, analisando os sinais de mais indivíduos, assim como mudar o intervalo de tempo utilizado dos sinais, de um segundo.

BIBLIOGRAFIA

- [1] L. F. Haas, "Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography," *J. Neurol. Neurosurg. Psychiatry*, vol. 74, no. 1, pp. 9–9, Jan. 2003.
- [2] S. Marcel and J. R. Millan, "Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 743–752, Apr. 2007.
- [3] P. Campisi, D. La Rocca, and G. Scarano, "EEG for Automatic Person Recognition," *Computer (Long Beach, Calif.)*, vol. 45, no. 7, pp. 87–89, Jul. 2012.
- [4] D. La Rocca *et al.*, "Human Brain Distinctiveness Based on EEG Spectral Coherence Connectivity," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 9, pp. 2406–2412, Sep. 2014.
- [5] J. Berkhout and D. O. Walter, "Temporal Stability and Individual Differences in the Human EEG: An Analysis of Variance of Spectral Values," *IEEE Trans. Biomed. Eng.*, vol. BME-15, no. 3, pp. 165–168, Jul. 1968.
- [6] F. Su, H. Zhou, Z. Feng, and J. Ma, "A biometric-based covert warning system using EEG," in *2012 5th IAPR International Conference on Biometrics (ICB)*, 2012, pp. 342–347.
- [7] K. J. Friston, "Functional and Effective Connectivity: A Review," *Brain Connect.*, vol. 1, no. 1, pp. 13–36, Jan. 2011.
- [8] A. M. Bastos and J.-M. Schoffelen, "A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls," *Front. Syst. Neurosci.*, vol. 9, Jan. 2016.
- [9] R. Cox, A. C. Schapiro, and R. Stickgold, "Variability and stability of large-scale cortical oscillation patterns," *Netw. Neurosci.*, vol. 2, no. 4, pp. 481–512, Oct. 2018.
- [10] E. Pereda, M. García-Torres, B. Melián-Batista, S. Mañas, L. Méndez, and J. J. González, "The blessing of Dimensionality: Feature Selection outperforms functional connectivity-based feature transformation to classify ADHD subjects from EEG patterns of phase synchronisation," *PLoS One*, vol. 13, no. 8, p. e0201660, Aug. 2018.
- [11] S. I. Dimitriadis, C. Salis, I. Tarnanas, and D. E. Linden, "Topological Filtering of Dynamic Functional Brain Networks Unfolds Informative Chronnectomics: A Novel Data-Driven Thresholding Scheme Based on Orthogonal Minimal Spanning Trees (OMSTs)," *Front. Neuroinform.*, vol. 11, Apr. 2017.
- [12] M. Garau, M. Fraschini, L. Didaci, and G. L. Marcialis, "Experimental results on multi-modal fusion of EEG-based personal verification algorithms," in *2016 International Conference on Biometrics (ICB)*, 2016, pp. 1–6.
- [13] I. Kottlarz *et al.*, "Extracting Robust Biomarkers From Multichannel EEG Time Series Using Nonlinear Dimensionality Reduction Applied to Ordinal Pattern Statistics and Spectral Quantities," *Front. Physiol.*, vol. 11, Feb. 2021.
- [14] R. S. Rosário, P. T. Cardoso, M. A. Muñoz, P. Montoya, and J. G. V. Miranda, "Motif-Synchronization: A new method for analysis of dynamic brain networks with EEG," *Phys. A Stat. Mech. its Appl.*, vol. 439, pp. 7–19, Dec. 2015.
- [15] E. Olofsen, J. W. Sleight, and A. Dahan, "Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect," *Br. J. Anaesth.*, vol. 101, no. 6, pp. 810–821, Dec. 2008.
- [16] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, Jun. 2000.
- [17] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, 2004.
- [18] H. Berger, "Über das elektroencephalogramm des menschen," *Arch. Für Psychiatr.*, vol. 87, no. 1, pp. 527–570, 1929.
- [19] R. S. do Rosário, "Sincronização por Motifs: uma proposta para análise de redes cerebrais variantes no tempo," Universidade Federal da Bahia, 2014.