

# Identificação biométrica com EEG pela comparação de matrizes de conectividade calculadas através de recorrências espaço-temporais

**Palavras-Chave:** biometria, eletroencefalografia, recorrências espaço-temporais

**Autores/as:**

Marina Cabral de Paulo [IFGW - UNICAMP]  
Manuela Von Ah Davanço [IFGW - UNICAMP]  
Paula Gabrielly Rodrigues [CECS - UFABC]  
Prof. Dr. Diogo Coutinho Soriano [CECS - UFABC]  
Prof.<sup>a</sup> Dr.<sup>a</sup> Gabriela Castellano (orientadora) [IFGW - UNICAMP]

## INTRODUÇÃO:

A eletroencefalografia (EEG) é uma das técnicas mais antigas utilizadas para o estudo da dinâmica cerebral [1]. Recentemente, esta técnica tem sido explorada para distinguir indivíduos e implementar um possível sistema biométrico [6] – [8]. Para isso, vários trabalhos têm usado sinais de EEG obtidos no estado de repouso [9], [10]. A vantagem de utilizar este tipo de paradigma (“estado de repouso”) é que pode ser aplicado a qualquer tipo de indivíduo (mesmo pacientes com deficiências, que teriam dificuldade para realizar algum experimento que requeresse uma resposta motora ou cognitiva) e, além disso, diminui o problema dos artefatos de movimento, já que a pessoa deve ficar imóvel durante a aquisição.

Dentre os trabalhos que visam obter informação biométrica a partir de sinais de EEG, estão os que têm usado características extraídas de eletrodos específicos, e aqueles que têm explorado as relações entre os sinais de EEG obtidos por diferentes eletrodos, método conhecido como conectividade cerebral. A ideia por trás das abordagens baseadas em conectividade é que muitas funções cerebrais são executadas por um conjunto de regiões (ou rede cerebral) e, portanto, entender como estas interagem entre si pode fornecer informação adicional sobre o indivíduo.

Os métodos para avaliar conectividade cerebral podem ser agrupados em conectividade anatômica (ou estrutural), conectividade funcional e conectividade efetiva [11]. O objetivo deste trabalho foi utilizar medidas baseadas em conectividade funcional, obtidas a partir de dados de EEG no estado de repouso, para tentar identificar diferentes indivíduos. Em 2021, apresentamos um trabalho piloto [12] com dados de 20 indivíduos, obtidos de uma base de dados disponível online. No presente trabalho, a análise foi estendida para 50 indivíduos, e foram alterados os métodos de pré-processamento e classificação dos sinais.

## METODOLOGIA:

Os dados de EEG utilizados foram obtidos do banco de dados online “*EEG Motor Movement/Imagery Dataset*” (<https://physionet.org/content/eegmmidb/1.0.0/>) [12]-[14]. Este banco contém dados de 109 indivíduos, mas para este trabalho, foram utilizados dados de 50 indivíduos. Cada indivíduo do banco possui 14 aquisições de EEG, sendo que cada aquisição foi feita com diferentes atividades. Em particular, as duas primeiras aquisições foram feitas no estado de repouso (uma com olhos abertos e outra com olhos fechados), e foram as utilizadas neste trabalho.

Após o *download* dos dados no formato EDF, o pré-processamento foi realizado na toolbox EEGlab do MATLAB, e consistiu nos seguintes passos: remoção dos artefatos por simples inspeção;

decomposição dos dados usando o ICA (*Independent Component Analysis*) e remoção de componentes indesejados; remoção de frequências da banda alfa (7 a 13 Hz) e da rede elétrica (60 Hz); e por fim, re-referenciamento CAR (*Common Average Referencing*).

Para remover os artefatos mais perceptíveis, foi utilizada a ferramenta “*Inspect/Reject data by eye*”, na qual seções dos sinais podem ser selecionadas para remoção. As seções que tinham amplitude muito maior (cinco vezes ou mais) do que o restante do sinal foram removidas.

Em seguida, o sinal foi decomposto em componentes independentes (ICA), utilizando a ferramenta “*Decompose Data by ICA*”. Esta ferramenta exibe as componentes independentes obtidas através de mapas de projeções da atividade auferida no exame de EEG da região do escalpo. Aquelas componentes relacionadas a movimentos musculares da face, como por exemplo, piscar de olhos e demais movimentos oculares podem ser, assim, facilmente reconhecidas e removidas.

Então, a banda alfa foi removida, utilizando um filtro de *stopband* (com a ferramenta “*Basic FIR Filter*” do EEGlab, considerando um intervalo de 7 Hz a 13 Hz, e selecionando a opção “*Notch filter the data instead of pass band*”). Isso foi feito porque queríamos comparar sinais obtidos em aquisições feitas com olhos fechados e com olhos abertos, e essa banda é notável por caracterizar movimentos de piscadas de olhos, sendo bem diferente para olhos abertos e fechados.

Por fim, o sinal foi filtrado com um filtro *passband*, ou seja, escolhemos uma banda de frequências e rejeitamos as partes do sinal que estão fora dela (novamente com a mesma ferramenta). Utilizamos o intervalo entre 4 Hz e 50 Hz, para eliminar artefatos de baixa frequência e ruído de alta frequência.

A etapa final de pré-processamento consistiu no re-referenciamento dos sinais usando CAR. Este método consiste em calcular a média dos sinais sobre os eletrodos e então subtrair esse valor de cada sinal de eletrodo e foi implementado diretamente no MATLAB (sem uso de ferramentas adicionais).

Todo o banco de dados foi pré-processado nas quatro etapas citadas, porém, neste trabalho, foram utilizados apenas os dados de 50 sujeitos. Esses sujeitos foram selecionados considerando a duração das aquisições após o pré-processamento. Indivíduos com aquisições com menos de 45 segundos foram descartados.

Da aquisição R1 (olhos abertos), foram extraídas quatro épocas, iniciando nos segundos 10, 20, 30 e 40. Da aquisição R2 (olhos fechados), apenas uma época foi extraída, iniciando no segundo 30. Durações de 1 segundo e 5 segundos foram testadas para essas épocas. Assim, as matrizes de conectividade funcional foram calculadas para as épocas R1 e R2. Foi usada uma abordagem de casamento com uma referência, na qual a média das matrizes R1 foi calculada para fornecer uma matriz de conectividade de referência por sujeito, enquanto a matriz R2 foi usada como amostra de teste.

O método de recorrências espaço-temporais [16],[17], que será detalhado a seguir, foi implementado para calcular essas matrizes de conectividade que foram usadas posteriormente na comparação. Tal método consiste basicamente em uma verificação de se o sistema dinâmico (que evolui ao longo do tempo, como são as aquisições de EEG) retornou a uma configuração em que já esteve anteriormente.

A recorrência espaço-temporal entre duas séries temporais  $x_i$  e  $x_j$  é, por definição:

$$STR_{i,j}(\varepsilon, n) = \Theta[\varepsilon - |x_i(n) - x_j(n)|] \quad (1)$$

A estrutura  $STR$  é chamada de matriz de recorrência espaço-temporal: uma estrutura de dados tridimensional de  $N \times N \times N_s$ , sendo  $N$  o número de canais (ou eletrodos, neste caso,  $N = 64$ ); e  $N_s$  o

número total de amostras no período de tempo escolhido (por exemplo,  $N_s = 160$  para épocas de 1 segundo ou  $N_s = 800$  para épocas de 5 segundos, uma vez que a taxa de amostragem foi de 160 Hz).  $\Theta$  é a função Heaviside, portanto:  $\Theta(x) = 0$  se  $x < 0$  e  $\Theta(x) = 1$  se  $x \geq 0$ . Finalmente,  $\varepsilon$  é um limiar de distância arbitrário. Neste trabalho, escolhemos  $\varepsilon = 50\%$  da distância máxima ( $|x_i(n) - x_j(n)|$ ) entre as séries temporais dos eletrodos.

Através da *STR* calculada pela Eq. 1, podemos encontrar a matriz de conectividade  $A_{i,j}$ , que consiste na normalização da soma dos valores de cada par de eletrodos da *STR*:

$$A_{i,j} = \frac{1}{N} \sum_{n=1}^N STR_{i,j}(\varepsilon, n). \quad (2)$$

Dessa forma, é possível reduzir a dimensão do problema em questão, já que  $A_{i,j}$  é uma estrutura bidimensional  $N \times N$ .

Uma vez calculadas as matrizes  $A_{i,j}$  correspondentes a R1 e R2, usamos o coeficiente de correlação de Pearson para compará-las, com a finalidade de avaliar a semelhança entre os sinais e, assim, identificar um determinado sujeito. Se o maior valor de correlação foi para R1 e R2 do mesmo indivíduo, foi possível identificar a pessoa, porque a comparação indicou maior similaridade entre diferentes aquisições de um mesmo sujeito. Caso contrário, não foi possível identificar a pessoa corretamente.

## RESULTADOS E DISCUSSÃO:

Ao realizar a comparação descrita anteriormente, foram obtidos os resultados expressos na Tabela 1 para a porcentagem de acertos.

Duração da época	11 sujeitos (teste)	50 sujeitos
1 segundo	81%	36%
5 segundos	90%	38%

*Tabela 1: Porcentagens de acertos para 11 e 50 sujeitos para cada época testada.*

Primeiro, foi feito um teste utilizando apenas 11 sujeitos dentre os 50 sujeitos selecionados da base de dados, e as taxas de acerto foram muito boas, com 80% para épocas de 1 s e 90% para épocas de 5 s. Todavia, ao aumentar a base e, portanto, o número de testes, a quantidade de acertos caiu significativamente, para 36% para épocas de 1 s e 38% para 5 s. Devemos ressaltar, contudo, que quanto maior a base de sujeitos diferentes usada na comparação, mais significativo é o resultado estatisticamente. Ou seja, na comparação com 11 sujeitos, uma taxa obtida de forma totalmente aleatória seria de 9%; já na comparação com 50 sujeitos, a taxa aleatória seria de 2%. Por isso, podemos argumentar que os resultados obtidos são maiores do que se esperaria se fizéssemos comparações aleatórias, logo, o método pode ser viável, mas precisa ser aprimorado, uma vez que em situações reais de identificação biométrica a base seria muito maior.

Outra característica determinante foi a duração da época. Conforme mostrado na Tabela 1, os resultados com trechos (épocas) cuja duração era de 5 segundos obtiveram resultados um pouco melhores do que as de 1 segundo.

## CONCLUSÕES:

Os resultados obtidos estão abaixo do desejado para a identificação biométrica de sujeitos. Entretanto, existem etapas do método implementado que podem ser melhoradas e devem ser consideradas.

Por exemplo, a etapa de pré-processamento poderia se beneficiar de um algoritmo de remoção automática de artefatos como SOUND [18], que tiraria a subjetividade de escolher quais trechos e componentes de ICA remover do sinal. Além disso, outros tipos de métodos de referência, como REST [19], podem ser explorados em vez do CAR. A implementação da matriz de recorrências espaço-temporais pode ser explorada para diferentes escolhas de limiar  $\epsilon$ .

Finalmente, na etapa de classificação, foi utilizado um método de classificação muito simples, a saber, o coeficiente de correlação de Pearson, mas podemos explorar outras abordagens de classificação como a distância geodésica [20] ou ainda abordagens comparativamente mais sofisticadas como *Linear Discriminant Analysis* (LDA), *Support Vector Machine* (SVM) ou mesmo redes neurais.

## Agradecimentos:

Agradecemos ao PIBIC/SAE-UNICAMP, CNPq (bolsa 304008/2021-4) e FAPESP (bolsa 2013/07759-3) pelo apoio financeiro.

---

## 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] M. Koutroumanidis *et al.*, "The role of EEG in the diagnosis and classification of the epilepsy syndromes: a tool for clinical practice by the ILAE Neurophysiology Task Force (Part 1)," *Epileptic Disord.*, vol. 19, no. 3, pp. 233–298, Sep. 2017.
- [3] I. G. Campbell, "EEG Recording and Analysis for Sleep Research," *Curr. Protoc. Neurosci.*, vol. 49, no. 1, Oct. 2009.
- [4] 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.
- [5] H. Van Dis, M. Corner, R. Dapper, G. Hanewald, and H. Kok, "Individual differences in the human electroencephalogram during quiet wakefulness," *Electroencephalogr. Clin. Neurophysiol.*, vol. 47, no. 1, pp. 87–94, Jul. 1979.
- [6] 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.
- [7] 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.
- [8] 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.

- [9] D. Mantini, M. G. Perrucci, C. Del Gratta, G. L. Romani, and M. Corbetta, “Electrophysiological signatures of resting state networks in the human brain,” *Proc. Natl. Acad. Sci.*, vol. 104, no. 32, pp. 13170–13175, Aug. 2007.
- [10] P. Campisi and D. La Rocca, “Brain waves for automatic biometric-based user recognition,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 5, pp. 782–800, May 2014.
- [11] K. J. Friston, “Functional and Effective Connectivity: A Review,” *Brain Connect.*, vol. 1, no. 1, pp. 13–36, Jan. 2011.
- [12] M. C. de Paulo, M. V. A. Davanço, M. S. J. Preiss, G. Castellano, “Estudo piloto sobre identificação de indivíduos a partir de dados de EEG por meio de conectividade funcional utilizando recorrências espaço-temporais”, In: XXIX Congresso de Iniciação Científica da UNICAMP, 2021, Campinas. Anais eletrônicos. Campinas, Galoá, 2021. Disponível em: <<https://proceedings.science/pibic-2021/papers/estudo-piloto-sobre-identificacao-de-individuos-a-partir-de-dados-de-eeg-por-meio-de-conectividade-funcional-utilizando-?lang=pt-br>>. Acesso em: 20 jul. 2022.
- [13] A. L. Goldberger *et al.*, “PhysioBank, PhysioToolkit, and PhysioNet,” *Circulation*, vol. 101, no. 23, Jun. 2000.
- [14] 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.
- [15] H. Berger, “Über das elektrenkephalogramm des menschen,” *Arch. Für Psychiatr.*, vol. 87, no. 1, pp. 527–570, 1929.
- [16] P. G. Rodrigues, C. A. S. Filho, R. Attux, G. Castellano, and D. C. Soriano, “Space-time recurrences for functional connectivity evaluation and feature extraction in motor imagery brain-computer interfaces,” *Med. Biol. Eng. Comput.*, vol. 57, no. 8, pp. 1709–1725, Aug. 2019.
- [17] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, “Recurrence Plots of Dynamical Systems,” *Europhys. Lett.*, vol. 4, no. 9, pp. 973–977, Nov. 1987.
- [18] T. P. Mutanen, J. Metsomaa, S. Liljander, and R. J. Ilmoniemi, “Automatic and robust noise suppression in EEG and MEG: The SOUND algorithm,” *Neuroimage*, vol. 166, pp. 135–151, Feb. 2018.
- [19] L. Dong *et al.*, “MATLAB Toolboxes for Reference Electrode Standardization Technique (REST) of Scalp EEG,” *Front. Neurosci.*, vol. 11, Oct. 2017.
- [20] M. Venkatesh, J. Jaja, L. Pessoa, “Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification”, *NeuroImage*, vol. 207, Nov. 2019.