

# Decodificação de Atenção Auditiva com ear-EEG: Análise de Viabilidade do uso da Abordagem Correlacional

Isabela Mary Sasaki e Silva

Universidade Tecnológica Federal do Paraná (UTFPR)  
Departamento Acadêmico de Eletrotécnica (DAELT)  
Curitiba, Paraná, BRA

Fábio Pires Itturriet

Universidade Tecnológica Federal do Paraná (UTFPR)  
Programa de Pós-Graduação em Engenharia Biomédica  
(PPGEB), Curitiba, Paraná, BRA

Wemerson Delcio Parreira

Pontifícia Universidade Católica de Campinas  
(PUC-Campinas), Escola Politécnica  
Campinas, São Paulo, BRA

Renata Coelho Borges

Universidade Tecnológica Federal do Paraná (UTFPR)  
Programa de Pós-Graduação em Engenharia Biomédica  
(PPGEB), Curitiba, Paraná, BRA

## RESUMO

Ear-EEG has emerged as a compact, low-energy alternative to conventional EEG, enabling brain activity monitoring in naturalistic settings. This technology shows potential for applications such as auditory attention decoding, sleep monitoring, epilepsy detection, and brain-computer interfaces. However, ear-EEG signals are characterized by low spatial resolution, high susceptibility to artifacts, and considerable interindividual variability, which challenge reliable signal interpretation. In this study, we investigated the feasibility of using Pearson's linear correlation between the acoustic envelope and EEG signals as a simple, low-complexity method to discriminate attended from ignored auditory stimuli. While individual trials occasionally exhibited clear separation between attended and ignored conditions, global patterns were modest, with strong overlap between conditions and high variability across participants and trials. Comparisons with non-parametric metrics (such as Kendall and Spearman) revealed no consistent improvement over Pearson correlation, suggesting that more complex approaches may not necessarily lead to substantially better results. These findings align with previous studies, where methods like canonical correlation analysis (CCA) and multivariate Temporal Response Functions (mTRF) yielded similar accuracy levels. Although linear correlation can capture markers of auditory attention in certain cases, robust decoding of attention in ear-EEG requires methods capable of modeling temporal dynamics and integrating information across multiple channels. Despite these limitations, Pearson correlation remains a computationally efficient tool for preliminary analyses, particularly in scenarios with limited resources, providing a useful baseline for future methodological advancements.

## KEYWORDS

ear-EEG; Decodificação da atenção auditiva; Correlação de Pearson

## 1 INTRODUÇÃO

O ear-EEG tem se consolidado como uma alternativa compacta e de baixo consumo energético ao EEG convencional, permitindo o monitoramento contínuo da atividade cerebral por meio de eletrodos discretamente posicionados na orelha [1]. Sua configuração intra-auricular possibilita a aquisição de sinais neurais com hardware minimalista, favorecendo o uso prolongado em ambientes naturais e

aplicações do cotidiano. Essas características tornam o ear-EEG particularmente atrativo para áreas como avaliação do sono, detecção de epilepsia, interfaces cérebro-computador e sistemas auditivos inteligentes [2-5].

Entre suas aplicações, destaca-se a decodificação da atenção auditiva (*Auditory Attention Decoding* - AAD), cujo objetivo é identificar, a partir dos sinais neurais, qual estímulo sonoro recebe o foco atencional do indivíduo em ambientes acústicos competitivos [6]. Essa capacidade é especialmente relevante em cenários como o problema da festa de coquetel, nos quais múltiplas fontes sonoras competem simultaneamente pela atenção do ouvinte. A AAD abre caminho para o desenvolvimento de próteses auditivas neurodirecionadas e interfaces cérebro-computador mais responsivas, capazes de se adaptar dinamicamente ao estado atencional do usuário.

Apesar de suas vantagens, como maior estabilidade mecânica e potencial redução da impedância de interface, o ear-EEG apresenta desafios significativos em termos de qualidade do sinal [7, 8]. Em comparação ao EEG de escalpo, os sinais obtidos na região auricular possuem menor amplitude e maior suscetibilidade a artefatos periféricos, o que impacta negativamente a relação sinal-ruído [9]. Ainda assim, a possibilidade de aquisição contínua, discreta e confortável torna o ear-EEG particularmente adequado para sistemas embarcados, com restrições de processamento, consumo energético e integração com dispositivos auditivos vestíveis.

O uso do ear-EEG para AAD enfrenta, adicionalmente, dificuldades associadas à elevada variabilidade inter e intraindivíduo, bem como à instabilidade do sinal ao longo do tempo [10-12]. Métodos tradicionais empregados na literatura, como regressão multivariada e redes neurais profundas, têm demonstrado bom desempenho na decodificação da atenção auditiva, porém à custa de alto custo computacional e dependência de grandes volumes de dados rotulados, o que dificulta sua implementação em sistemas portáteis e de baixo consumo energético [13].

Nesse contexto, torna-se relevante investigar métricas de baixa complexidade, capazes de operar em tempo real e com reduzida demanda computacional. Uma alternativa particularmente simples é a estimação direta da correlação entre o envelope acústico e o sinal neural, abordagem amplamente empregada em modelos estímulo-resposta. Embora essa técnica seja recorrente como ferramenta auxiliar na análise da relação entre estímulo e EEG, seu uso como métrica primária para a decodificação da atenção auditiva ainda é pouco explorado, especialmente em cenários de ear-EEG [14, 15].

Este trabalho investiga a viabilidade do uso da correlação linear direta como ferramenta para discriminar estímulos atendidos e ignorados a partir de sinais de ear-EEG. A análise é conduzida utilizando um banco de dados público amplamente empregado na literatura de AAD, permitindo a avaliação sistemática do comportamento das correlações sob diferentes condições de atenção. Os resultados oferecem uma análise quantitativa das potencialidades e limitações dessa abordagem, contribuindo para a compreensão das fronteiras operacionais de métricas lineares simples aplicadas ao ear-EEG.

## 2 METODOLOGIA

Neste estudo, adotou-se uma abordagem quantitativa para avaliar a decodificação da atenção auditiva a partir de sinais ear-EEG. O processo metodológico envolveu a seleção e pré-processamento de dados, filtragem e extração de envelopes, alinhamento temporal com os estímulos de áudio e análise estatística das correlações.

### 2.1 Banco de dados e aquisição dos sinais

Para o desenvolvimento deste trabalho, foi utilizado o banco de dados disponibilizado por Thornton et al. [16] e empregado em Thornton et al. [12], conforme ilustrado na Figura 1. O conjunto de dados foi obtido a partir de um experimento envolvendo 19 participantes saudáveis. Cada participante escutou 16 capítulos de dois audiolivros narrados simultaneamente por um locutor masculino e uma locutora feminina, sendo instruído a direcionar sua atenção a apenas um dos fluxos de fala. A cada quatro capítulos, os participantes eram solicitados a alternar o foco de atenção entre os dois narradores. Os estímulos auditivos foram apresentados sem separação espacial, isto é, os dois áudios foram reproduzidos simultaneamente em ambos os ouvidos. Os capítulos apresentavam durações variáveis.

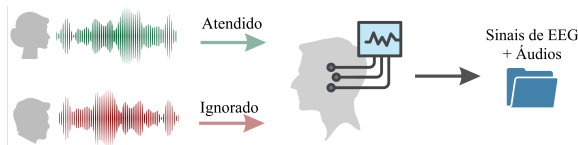


Figura 1: Representação geral da aquisição de dados realizada por Thornton et al. [16].

A aquisição dos sinais neurais foi realizada por meio de um sistema de ear-EEG sem fio, com taxa de amostragem de 256 Hz, composto por dois canais intra-auriculares (orelha esquerda e direita), um eletrodo de referência posicionado na região frontal e um eletrodo de terra localizado no lóbulo da orelha. O banco de dados é disponibilizado em formato .h5, organizado por participante e por *trial* (capítulo), enquanto os sinais de áudio são fornecidos em arquivos .wav, separados em estímulos atendidos e ignorados para cada *trial*. O sistema fornece dois canais unilaterais: Canal 1 (CH1), correspondente à orelha esquerda, e Canal 2 (CH2), correspondente à orelha direita. As análises individuais foram realizadas separadamente para cada canal, uma vez que a captação local pode apresentar variações entre os participantes devido a ruído, assimetrias anatômicas e fisiológicas, bem como à geometria e ao posicionamento dos eletrodos [12, 17]. Observou-se que o participante 3 não possui dados referentes à primeira *trial*, provavelmente devido a

problemas técnicos durante a aquisição. Por esse motivo, apenas 18 participantes foram considerados nas análises apresentadas neste trabalho.

### 2.2 Pré-processamento dos sinais

Todo o processamento foi realizado no Matlab. Inicialmente, os arquivos de áudio e EEG foram lidos e organizados em estruturas de dados apropriadas para manipulação no software. Os áudios foram convertidos de estéreo para mono e reamostrados de 44,1 kHz para 256 Hz, igualando a frequência de amostragem do EEG, de modo a permitir a sincronização precisa entre os sinais acústicos e neurais durante as análises.

**2.2.1 Cálculo do envelope via Transformada de Hilbert.** O envelope dos sinais de áudio foi obtido por meio da Transformada Analítica de Hilbert, técnica amplamente utilizada para extrair a modulação de amplitude da fala. Dado um sinal de áudio  $x(t)$ , calcula-se sua transformada de Hilbert  $\mathcal{H}\{x(t)\}$ , definida como

$$\mathcal{H}\{x(t)\} = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau, \quad (1)$$

em que p.v. denota o valor principal da integral no sentido de Cauchy. O sinal analítico é então construído como

$$x_a(t) = x(t) + j\mathcal{H}\{x(t)\}, \quad (2)$$

e o envelope, utilizado nas análises de correlação, é obtido pelo módulo desse sinal complexo:

$$e(t) = |x_a(t)| = \sqrt{x(t)^2 + \mathcal{H}\{x(t)\}^2}. \quad (3)$$

A informação relevante do envelope de fala encontra-se predominantemente em modulações de baixa frequência (tipicamente abaixo de 10 Hz), conforme demonstrado em estudos clássicos de percepção e processamento neural da fala [18, 19]. Assim, a reamostragem do áudio para 256 Hz não acarreta perda significativa de informação para análises baseadas no envelope, uma vez que essas frequências de modulação permanecem integralmente preservadas após o processamento [18–20].

### 2.3 Janelamento

O artigo de referência [12], que utiliza o mesmo banco de dados empregado neste estudo, realiza uma análise focada no efeito do tamanho das janelas. Durante a validação cruzada, as *trials* de teste foram divididas em janelas não sobrepostas de diferentes durações. Para cada tamanho de janela, a acurácia foi calculada como a proporção de janelas em que o método conseguiu identificar corretamente o áudio correspondente. Além disso, foi ajustado o “nível de chance” para cada tamanho de janela, levando em consideração o número total de janelas disponíveis (uma vez que janelas mais curtas geram um maior número de decisões), permitindo assim uma comparação com o desempenho aleatório.

Os resultados indicaram que a acurácia aumentava com o tamanho da janela, seguindo uma relação aproximada do tipo raiz-quadrada entre a duração da janela e a acurácia (com ganhos rápidos no início, seguidos de uma saturação posterior), sugerindo maior estabilidade estatística à medida que mais amostras eram incorporadas no cálculo da correlação. Adicionalmente, o decodificador baseado no *onset*-envelope apresentou acurácias populacionais mais

altas e menor variabilidade ao longo dos diferentes tamanhos de janela, quando comparado ao baseado no envelope temporal.

Com base nessa abordagem, foi implementada uma etapa de varredura de janelas, com durações variando de 5 a 40 segundos, com o objetivo de investigar se a variabilidade observada nos resultados estava relacionada à escolha do tamanho da janela. Janelas mais curtas tendem a produzir correlações ruidosas e instáveis, enquanto janelas mais longas suavizam flutuações momentâneas, evidenciando padrões atencionais mais consistentes. A avaliação de múltiplos tamanhos de janela permite analisar a robustez do método em diferentes escalas temporais, identificar a duração que maximiza a acurácia sem comprometer a resolução temporal e comparar diretamente os resultados com a curva de estabilidade reportada no estudo de referência. Esse procedimento possibilita, ainda, separar os efeitos específicos do banco de dados de limitações metodológicas.

## 2.4 Atraso neural

Para investigar a influência de fatores temporais nas correlações entre o EEG e o estímulo, além do efeito do tamanho das janelas, é essencial considerar o atraso neural (*lag*). Embora o janelamento nos permita explorar como a duração das janelas de análise afeta a acurácia, o atraso temporal no processamento neural também desempenha um papel importante, já que as respostas cerebrais não são instantâneas e apresentam variações individuais e experimentais. Esse fator é crucial para o alinhamento preciso dos sinais de EEG aos estímulos auditivos, uma vez que o processamento neural apresenta latência variável entre indivíduos e *trials*, influenciado por diferentes condições, como atenção e fadiga [11].

No estudo de Thornton et al. [12], o atraso foi tratado por meio do modelo TRF (*Temporal Response Functions*), utilizando uma faixa contínua de *lags* (-64 a +128 amostras a 64 Hz), cobrindo aproximadamente 1 s antes até 2 s após o envelope. No presente trabalho, o *lag* foi selecionado manualmente com base na acurácia global, definida como a proporção de janelas em que a correlação com o áudio atendido supera a do áudio ignorado. Foram avaliados *lags* variando de -400 ms a 1000 ms.

## 3 ANÁLISE DE CORRELAÇÃO

Para investigar a relação entre os sinais de EEG e os áudios apresentados, utilizou-se uma abordagem baseada em correlações estatísticas. O objetivo principal é quantificar até que ponto a atividade neural registrada nas orelhas acompanha os envelopes dos estímulos auditivos atendidos, em comparação com os ignorados. Essa análise permite avaliar de forma objetiva o rastreamento da atenção auditiva pelos participantes.

Foram consideradas três métricas complementares de correlação: Pearson, Kendall e Spearman, cada uma com características distintas. O coeficiente de correlação de Pearson é uma medida paramétrica que quantifica associações lineares entre duas variáveis contínuas sob a suposição de normalidade das distribuições marginais — ou seja, captura a força e a direção de uma relação linear entre variáveis quantificáveis [21]. Em contraste, as correlações de Spearman e Kendall são medidas não paramétricas de postos que avaliam associações monotônicas, independentemente da forma funcional exata da relação entre as variáveis, e são mais

robustas a não linearidades, violações de normalidade e valores extremos [22, 23].

Todos os cálculos foram realizados em janelas temporais, tanto para os envelopes de EEG, quanto para os envelopes dos áudios atendido e ignorado. Essa abordagem garante comparações consistentes e possibilita a análise da atenção auditiva em escala temporal adequada, permitindo ainda explorar a influência do tamanho da janela sobre a estabilidade e acurácia da detecção.

### 3.1 Correlação de Pearson

A correlação de Pearson é uma medida estatística que avalia a intensidade e a direção da relação linear entre duas variáveis quantitativas [18, 21]. Seu coeficiente, chamado  $r$ , varia entre -1 e 1, onde valores próximos de 1 indicam forte correlação positiva (quando uma variável aumenta, a outra também aumenta), valores próximos de -1 indicam forte correlação negativa (quando uma variável aumenta, a outra diminui) e valores próximos de 0 sugerem ausência de relação linear. A equação matemática do coeficiente de Pearson é dada por:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

em que  $x_i$  e  $y_i$  representam os valores observados das variáveis  $X$  e  $Y$ ,  $\bar{x}$  e  $\bar{y}$  são suas respectivas médias, e  $n$  é o número de observações.

Neste estudo, a correlação de Pearson foi utilizada para avaliar a similaridade entre os sinais de EEG e os áudios. Para cada janela, foram calculadas as correlações entre os envelopes de cada canal de EEG e os envelopes dos dois áudios (atendido e ignorado).

### 3.2 Correlação de Kendall

A correlação de Kendall (ou  $\tau$  de Kendall) é uma medida não paramétrica que quantifica a associação monotônica entre duas variáveis com base na concordância entre pares de observações [23]. Ao invés de depender diretamente dos valores das amostras, Kendall avalia se pares  $(x_i, y_i)$  são concordantes (variam na mesma direção) ou discordantes (variam em direções opostas). Seu coeficiente,  $\tau$ , varia entre -1 e 1, sendo que valores próximos de 1 indicam forte concordância monotônica positiva e valores próximos de -1 indicam forte concordância negativa. A equação de Kendall é dada por:

$$\tau = 2 \frac{C - D}{n(n - 1)} \quad (5)$$

em que  $C$  é o número de pares concordantes,  $D$  o número de pares discordantes e  $n$  o número de observações. Neste trabalho,  $\tau$  foi calculado entre o envelope do EEG e os envelopes de áudio atendido e ignorado em cada janela, permitindo detectar relações monotônicas mesmo quando a associação linear não é evidente.

### 3.3 Correlação de Spearman

A correlação de Spearman baseia-se nos *ranks* (ordenações) dos dados, sendo mais robusta a não-linearidades e valores extremos [22]. Seu coeficiente, denominado  $\rho$ , varia entre -1 e 1, seguindo a mesma escala de Pearson e Kendall. A equação é dada por:

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (6)$$

em que  $d_i = \text{rank}(x_i) - \text{rank}(y_i)$  representa a diferença entre os ranks das observações  $x_i$  e  $y_i$ , e  $n$  é o número de observações. Neste estudo, a correlação de Spearman foi utilizada como alternativa à Pearson para avaliar associações monotônicas entre os envelopes de EEG e os áudios atendido e ignorado, considerando janelas temporais segmentadas.

## 4 RESULTADOS

Nesta seção, são apresentados os principais achados do estudo, organizados em diferentes níveis de complexidade. Inicialmente, os dados foram analisados de forma individual, com foco na variabilidade interindividual no EEG bruto, considerando participantes específicos e *trials* isoladas. Em seguida, foi realizada uma análise populacional (agregada), com o cálculo das métricas de correlação e acurácia para identificar padrões gerais de separação entre as condições de áudio atendido e ignorado. O efeito do tamanho da janela temporal e da latência (*lag*) na separação entre essas condições também foi investigado. Por fim, o impacto do pré-processamento foi discutido e diferentes métricas de correlação foram comparadas, a fim de avaliar quais abordagens oferecem maior robustez na decodificação da atenção auditiva a partir do EEG.

### 4.1 Análise Individual de Participantes

A análise inicial foi conduzida de forma individual, considerando participantes específicos e uma *trial* isolada, com o objetivo de examinar detalhadamente a distribuição das correlações entre canais do EEG bruto. Os boxplots das correlações evidenciaram padrões altamente heterogêneos entre os sujeitos, indicando que a atividade elétrica cerebral apresenta grande variabilidade interindividual mesmo sob condições experimentais idênticas. Como exemplo, os participantes 9 (Figura 2) e 16 (Figura 3), na *trial* 9, exibiram separação visível entre correlação com o áudio atendido e com o áudio ignorado. Entretanto, o sinal dessas correlações mostrou-se invertido entre os indivíduos, sendo positivo em um participante e negativo em outro. Tal fenômeno pode ser interpretado como uma demonstração de que a polaridade das correlações não é consistente entre sujeitos, refletindo diferenças individuais na topografia, na fase ou na sincronização das respostas neurais. Essa observação ressalta a importância de análises que considerem tanto a variabilidade intra- quanto interindividual e sugere que métricas agregadas, como médias populacionais, devem ser interpretadas com cautela, especialmente quando se busca generalizar padrões observados em EEG bruto.

### 4.2 Análise Agregada

Diante da elevada variabilidade observada nos resultados individuais, a análise foi estendida para uma abordagem consolidada, agregando as métricas de todas as *trials* de todos os participantes. O objetivo dessa etapa foi identificar padrões mais consistentes e avaliar a existência de uma relação sistemática entre o EEG bruto e os estímulos auditivos. Para isso, foram calculadas, para cada canal, a média da diferença entre as correlações com o áudio atendido e ignorado ( $\Delta r$ ), a acurácia, definida como a proporção de janelas em que a correlação com o áudio atendido foi maior, e os respectivos desvios padrão. Essa análise permitiu uma avaliação populacional da associação entre o EEG e os estímulos, reduzindo o impacto da

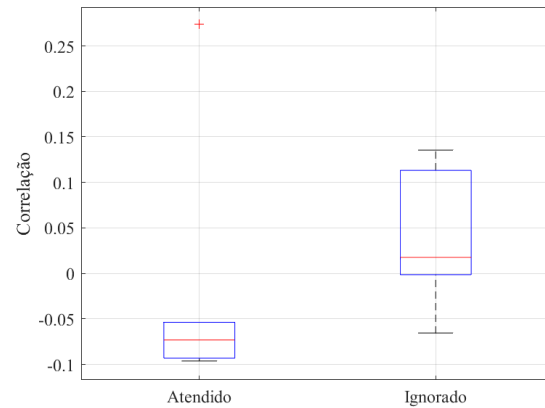


Figura 2: Correlação entre EEG e áudio atendido e áudio ignorado para o participante 9 na *trial* 9, obtida no Canal 1.

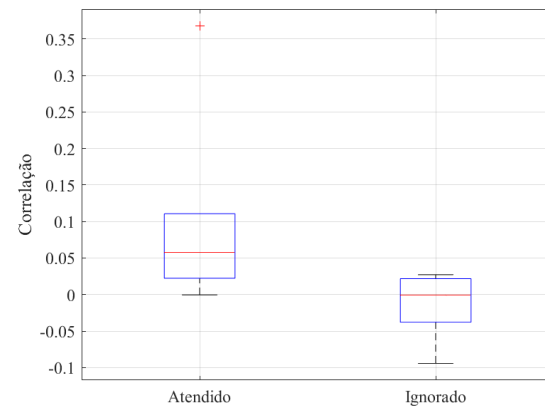


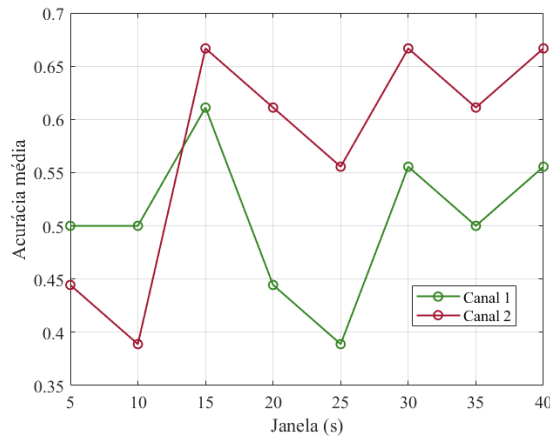
Figura 3: Correlação entre EEG e áudio atendido e áudio ignorado para o participante 16 na *trial* 9, obtida no Canal 1.

variabilidade interindividual e destacando tendências gerais presentes no conjunto de dados. Essas métricas foram obtidas apenas para as *trials* estáveis (2, 3, 6, 7, 10, 11, 14 e 15), isto é, trechos do experimento em que a instrução de atenção permaneceu fixa do início ao fim da *trial*, sem mudança do falante atendido. Em contraste, nas demais *trials* ocorre o estado de alternância de atenção, no qual o participante é induzido a trocar o foco auditivo entre os dois fluxos de fala ao longo da mesma *trial*. Como essa alternância introduz transientes neurais e períodos de adaptação logo após a troca, ela pode distorcer a estimativa de correlação entre EEG e envelope do áudio. Por isso, limitar a análise às *trials* estáveis reduz interferências desse processo de troca e permite avaliar de forma mais limpa o rastreamento de envelope associado à atenção sustentada.

### 4.3 Efeito do Tamanho da Janela

Uma vez que a correlação foi calculada por janelas temporais, foi investigada ainda a influência do tamanho da janela sobre a estabilidade das métricas globais. Foram testadas janelas de 5 seg a

40 seg (em passos de 5 seg), para cada condição. A Figura 5 mostra a variação da acurácia média por janela no Canal 1 e no Canal 2.



**Figura 4: Variação da acurácia de acordo com o tamanho da janela para Canal 1 (esquerda) e Canal 2 (direita).**

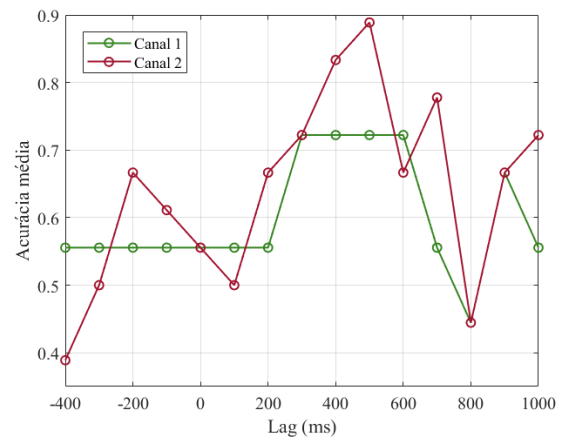
A análise do efeito do tamanho da janela sobre a acurácia revelou variações significativas, embora sem uma tendência monotônica clara de melhora com o aumento da duração. No Canal 1, os valores de acurácia oscilaram entre 0,389 e 0,611, com uma média aproximada de 0,506, indicando desempenho ligeiramente superior ao nível de acaso. No Canal 2, os índices variaram de 0,389 a 0,667, com média em torno de 0,573, mostrando um desempenho consistentemente superior ao acaso e ligeiramente mais elevado do que o observado no Canal 1. Apesar das flutuações entre janelas consecutivas, ambos os canais mantiveram acurácias acima de 50% na maior parte das medições, refletindo a presença de informação discriminativa no EEG bruto. Esses resultados sugerem que o efeito do tamanho da janela é moderado, com algumas durações (10-20 s para Canal 2 e 25-30 s para Canal 1) produzindo picos de desempenho, mas sem induzir melhorias contínuas. A análise evidencia ainda a importância de considerar a variabilidade inter-janelas e inter-canais ao interpretar a robustez de métricas baseadas em EEG bruto, reforçando a necessidade de abordagens que capturem padrões consistentes em diferentes escalas temporais.

Na literatura sobre decodificação da atenção auditiva (AAD) com ear-EEG, é comum observar melhora de desempenho com o aumento da duração das janelas, uma vez que segmentos mais longos fornecem maior número de amostras, estabilizam a estimativa de reconstrução e melhoram a relação sinal-ruído, especialmente em sensores de baixa densidade como os ear-EEG [9]. Além disso, o próprio banco de dados utilizado por Thornton et al. [12] evidencia que marcadores de atenção podem ser detectados mesmo em segmentos curtos, embora a estabilidade não seja garantida em todas as condições experimentais. Nesse contexto, o fato de os resultados obtidos neste estudo não apresentarem ganho contínuo com o aumento da janela sugere que, para a métrica direta de correlação empregada, o desempenho é limitado principalmente pela variabilidade intersujeitos e pela sensibilidade do método. Considerando esses fatores, a escolha de uma janela de 30 s se justifica como um

compromisso prático: neste ponto, o Canal 2 apresentou o melhor desempenho global, com picos de acurácia mais elevados e consistentes, sem evidências de benefício adicional ao prolongar ainda mais a duração das janelas.

#### 4.4 Influência da Latência (*Lag*)

Para investigar de maneira mais detalhada o efeito da latência na capacidade de distinguir entre áudio atendido e ignorado, realizamos uma varredura abrangente de *lags* utilizando o EEG bruto segmentado em janelas. A análise foi conduzida ao longo de um intervalo de *lags* que variava de -400 ms a +1000 ms, o que permitiu explorar uma ampla gama de possíveis defasagens temporais entre a resposta neural e os estímulos auditivos. A acurácia foi calculada para cada valor de *lag* e os resultados são apresentados na Figura 5, que mostra a variação da acurácia global para os Canais 1 e 2.



**Figura 5: Curvas globais da acurácia vs. lag (-400 ms a +1000 ms, passo 100 ms) para Canal 1 (esquerda) e Canal 2 (direita).**

Nos dois canais, a análise revelou que as acurácias globais mantiveram-se predominantemente próximas ao nível de chance, com uma variação suave ao longo do intervalo de *lags* considerado. No Canal 1, as acurácias oscilaram em uma faixa relativamente estreita entre 0,55 e 0,72, sem indicar um pico pronunciado em nenhum ponto do intervalo de latência. Embora tenha sido observada uma leve melhoria de desempenho entre 200 e 600 ms, o aumento não foi substancial o suficiente para sugerir uma latência ótima para a separação entre as condições atendida e ignorada.

Por outro lado, o Canal 2 exibiu um comportamento mais irregular. A variação nas amplitudes das acurácias foi mais proeminente, com flutuações visíveis ao longo do intervalo de *lags*. No entanto, observou-se um aumento gradual da acurácia entre 200 e 500 ms, que, embora perceptível, não foi consistente o suficiente para indicar uma latência de resposta ideal. O desempenho do Canal 2, de forma geral, apresentou maior instabilidade em comparação ao Canal 1, refletindo possíveis diferenças na qualidade do sinal ou na sensibilidade do canal em capturar a dinâmica da atenção auditiva.

Em conjunto, os resultados sugerem que, dentro das condições experimentais adotadas, a escolha do valor de *lag* tem um efeito

modesto sobre o desempenho, sem indicar uma latência ótima claramente definida. A falta de uma separação substancial entre as condições atendida e ignorado é um fator importante que contribuiu para a modesta variação nas acurácias. Quando a separação entre as condições é fraca, como no presente estudo, as curvas de acurácia ao longo do intervalo de *lags* se mantêm relativamente planas, o que implica que a latência, por si só, não tem um impacto decisivo na melhoria da distinção entre as condições.

#### 4.5 Impacto do Pré-processamento do EEG e do Áudio

Como análise adicional, avaliou-se o impacto de duas escolhas de pré-processamento. Primeiro, repetiu-se a etapa global com o envelope suavizado por um filtro passa-baixa (Caso 1) e com o envelope do áudio sem filtragem adicional (Caso 2). Observou-se que apesar das diferenças entre as duas condições serem pequenas, ao aplicar o passa-baixa houve um leve aumento de  $\Delta r$  e da acurácia (Tabela 1), sendo  $\Delta r$  a diferença entre correlação com o áudio atendida ( $r_{att}$ ) e o áudio ignorado ( $r_{ign}$ ) e SD (*standard deviation*) os valores de desvio padrão. A acurácia foi definida como a proporção de janelas em que  $r_{att} > r_{ign}$ . Isso sugere que o filtro, ao atenuar componentes rápidas do envelope, pode influenciar ligeiramente na correlação entre estímulo e resposta.

Na sequência, avaliou-se o impacto da extração do envelope do EEG bruto. A obtenção da envoltória por meio da transformada de Hilbert (Caso 1) atenua a contribuição das oscilações rápidas e da fase instantânea do sinal, preservando predominantemente a modulação lenta da amplitude do sinal, uma grandeza conceitualmente mais compatível com o envelope temporal do estímulo auditivo. Quando essa etapa é removida e o EEG filtrado é utilizado diretamente, sem extração da envoltória (Caso 3), observa-se uma redução da acurácia. Esse resultado sugere que, na ausência da envoltória, a correlação passa a ser mais sensível a variações de fase, componentes de alta frequência e ruído, o que pode comprometer a estabilidade da associação entre EEG e envelope do áudio. Assim, a extração do envelope do EEG parece desempenhar um papel importante na redução da variabilidade e na preservação de componentes temporais relevantes para a discriminação entre os estados de atenção atendida e ignorada dentro desta abordagem baseada em correlação direta.

Sendo assim, considerando as comparações realizadas, a melhor acurácia global foi obtida na condição com filtragem adicional do áudio e extração do envelope do EEG (Caso 1). Ainda assim, mesmo nessa configuração, as acurácias permaneceram em torno de 60%, valor apenas moderadamente superior ao nível de acaso, e os valores médios de  $\Delta r$  continuaram reduzidos. Esses resultados indicam a ausência de um efeito robusto e evidenciam a elevada variabilidade entre participantes e *trials*. Essa limitação torna-se particularmente clara nos gráficos de blocos (Figuras 6 e 7), que mostram uma forte sobreposição entre as distribuições das correlações associadas às condições atendida e ignorada, reforçando a dificuldade de separação consistente entre os dois estados atencionais com base nessa abordagem.

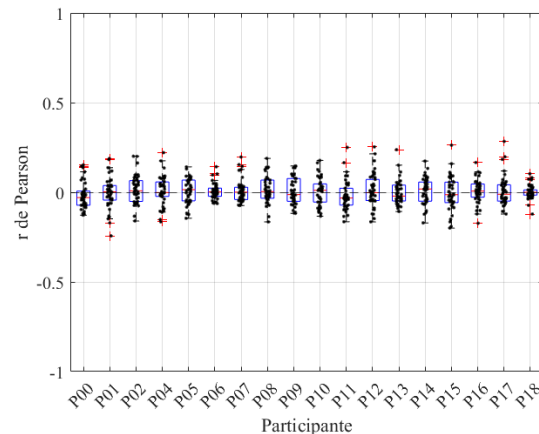


Figura 6: Distribuição global dos coeficientes de correlação de Pearson (Canal 1) entre o EEG e o áudio atendida.

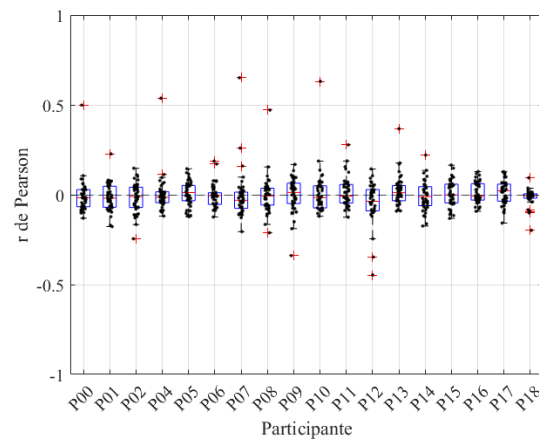


Figura 7: Distribuição global dos coeficientes de correlação de Pearson (Canal 1) entre o EEG e o áudio ignorado.

#### 4.6 Comparação de Métricas de Correlação

Além das análises com correlação de Pearson, calculou-se também Kendall e Spearman nas mesmas condições de pré-processamento que apresentaram melhor desempenho para Pearson (janela de 30 segundos, com filtragem adicional do envelope do áudio e com extração do envelope do EEG). O objetivo foi verificar se métricas monotônicas, menos dependentes de linearidade, seriam capazes de capturar de forma mais robusta o alinhamento entre o envelope de fala e o EEG. A Figura 8 resume as acurácias globais obtidas para ambos os canais.

Observa-se que a correlação de Pearson apresentou as maiores acurácias em ambos os canais, enquanto Spearman e Kendall tiveram resultados similares entre si. Entretanto, tais resultados devem ser interpretados com cautela, uma vez que as condições foram otimizadas com base na Correlação de Pearson. Portanto, esta comparação indica apenas que, nesta configuração específica,

Tabela 1: Resultados das medições para diferentes abordagens.

Caso	Canal 1				Canal 2			
	Média $\Delta r$	SD $\Delta r$	Acurácia	SD Acurácia	Média $\Delta r$	SD $\Delta r$	Acurácia	SD Acurácia
Caso 1	0.0714	0.0231	0.6111	0.5016	0.0704	0.0200	0.6667	0.4851
Caso 2	0.0669	0.0213	0.5556	0.5113	0.0658	0.0192	0.6667	0.4851
Caso 3	0.0658	0.0188	0.5556	0.5113	0.0658	0.0197	0.5000	0.5145

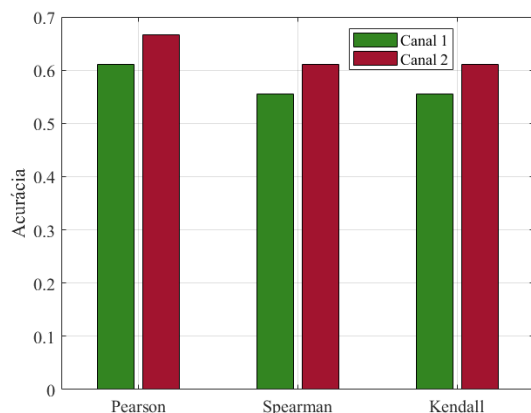


Figura 8: Comparação gráfica entre os resultados globais das correlações de Pearson, Kendall e Spearman para o Canal 1 e Canal 2.

a correlação linear foi a que melhor separou as respostas ao áudio atendido e ao ignorado, enquanto Kendall e Spearman não trouxeram melhora adicional.

## 5 DISCUSSÕES

Os resultados deste trabalho indicam que é possível extrair informação relacionada à atenção auditiva a partir de sinais de ear-EEG utilizando métricas simples de correlação, com acurácias que atingem valores de até aproximadamente 66% nas melhores configurações. Esse desempenho, embora consistentemente superior ao nível de acaso, mostrou-se fortemente dependente do participante, da *trial* e das escolhas metodológicas, não se sustentando de forma uniforme em todas as condições analisadas.

Na análise individual, alguns participantes apresentaram padrões claros de separação entre as condições atendida e ignorada; contudo, esses efeitos não se generalizaram no nível populacional. Na consolidação global, os valores médios de  $\Delta r$  permaneceram reduzidos e as distribuições de correlação associadas às duas condições exibiram forte sobreposição, indicando que os ganhos observados em acurácia decorrem de diferenças sutis e instáveis entre as respostas neurais aos estímulos auditivos.

Diversos fatores contribuem para esse comportamento. Primeiramente, a elevada variabilidade intersujeitos — associada a diferenças anatômicas, fisiológicas e comportamentais — influencia diretamente a forma como a atenção auditiva se manifesta nos

sinais de EEG [13]. Em segundo lugar, a própria natureza do ear-EEG, caracterizada por um número reduzido de eletrodos e maior suscetibilidade a ruído e artefatos, limita a relação sinal-ruído disponível para detectar efeitos atencionais sutis [11, 17]. Além disso, a abordagem baseada em correlação direta pressupõe uma relação predominantemente linear e instantânea entre o envelope da fala e o EEG, o que pode representar uma simplificação excessiva para um fenômeno que envolve múltiplas escalas temporais e atrasos neurais variáveis.

As análises complementares reforçam essa interpretação. A varredura de tamanhos de janela não revelou um ganho monotônico de desempenho com o aumento da duração, como frequentemente reportado na literatura, mas sim oscilações com picos localizados em determinadas janelas. Isso sugere que, nesta abordagem, o desempenho não é limitado apenas pelo número de amostras disponíveis, mas principalmente pela combinação entre variabilidade intersujeitos, instabilidade inter-*trials* e sensibilidade reduzida da métrica utilizada. De modo semelhante, a comparação entre diferentes coeficientes de correlação indicou que, nas condições avaliadas, a correlação de Pearson apresentou desempenho igual ou superior às métricas não paramétricas, sem evidência de ganho relevante ao empregar Kendall ou Spearman.

De forma consistente com resultados recentes na literatura de ear-EEG, nos quais métodos lineares e não lineares apresentaram desempenhos semelhantes em janelas de curta duração, os achados deste trabalho sugerem que o gargalo do desempenho não reside exclusivamente na técnica de decodificação, mas principalmente na qualidade e na representatividade do sinal neural capturável por dispositivos posicionados na orelha. Em particular, Thornton et al. [12] demonstraram que marcadores de atenção auditiva podem ser extraídos mesmo de segmentos curtos de ear-EEG, e que abordagens como CCA, regressão linear e métodos não lineares tendem a apresentar diferenças de desempenho pequenas e, em muitos casos, não significativas. Esses resultados reforçam a ideia de que ganhos substanciais em acurácia exigem não apenas algoritmos mais complexos, mas também melhorias no sinal disponível ou no próprio paradigma experimental.

## 6 CONCLUSÕES

Este estudo apresentou uma análise detalhada da relação entre EEG e estímulos auditivos em uma tarefa de atenção auditiva utilizando ear-EEG. Mesmo com uma abordagem simples baseada na correlação de Pearson, foi possível alcançar uma discriminação moderada entre as condições atendida e ignorada, com acurácias médias entre 61–66% nas melhores configurações. Embora esse desempenho não seja elevado, ele está em consonância com valores reportados na literatura para sistemas de ear-EEG e reflete limitações estruturais

associadas à qualidade do sinal, à cobertura espacial reduzida e à menor relação sinal-ruído em comparação com montagens convencionais de EEG.

A comparação com métricas não paramétricas e com abordagens mais sofisticadas de decodificação não revelou ganhos substanciais em relação ao método simples adotado. Esse resultado sugere que, no contexto avaliado, o aumento da complexidade algorítmica, por si só, não é suficiente para superar as restrições impostas pela qualidade do sinal. A elevada variabilidade intersujeitos e entre *trials* destacou-se como um fator central na interpretação dos resultados, evidenciando que diferenças individuais na morfologia anatômica, no posicionamento dos eletrodos e na dinâmica neural da atenção auditiva influenciam significativamente o desempenho dos classificadores.

Ainda assim, modelos como o mTRF permanecem conceitualmente relevantes, pois oferecem uma modelagem mais realista da dinâmica temporal entre estímulo e resposta neural, permitindo estimar funções de resposta temporal que caracterizam explicitamente a relação estímulo-córtex [24]. Mesmo quando não resultam em ganhos imediatos de acurácia, tais abordagens contribuem para uma compreensão mais profunda dos mecanismos subjacentes à codificação cortical da fala e mantêm potencial para aplicações em ear-EEG, como discutido também em [16].

Resultados semelhantes foram relatados por Thornton et al. [12], que também utilizaram ear-EEG e obtiveram acurácias comparáveis com métodos como CCA e mTRF. Em conjunto, esses achados reforçam a ideia de que, no contexto do ear-EEG, o principal gargalo não está necessariamente na sofisticação do algoritmo, mas na qualidade e estabilidade do sinal neural capturado.

Diante desse cenário, as implicações futuras apontam para múltiplas frentes de desenvolvimento. Primeiramente, melhorias no hardware, incluindo otimização do design dos eletrodos, melhor adaptação ao canal auditivo e redução de artefatos de movimento, podem contribuir de forma decisiva para elevar a relação sinal-ruído. Em segundo lugar, estratégias avançadas de pré-processamento, como técnicas adaptativas de remoção de artefatos e métodos de normalização individualizados, podem reduzir a variabilidade intersujeitos. Além disso, abordagens personalizadas de modelagem, com treinamento específico por participante ou técnicas de adaptação de domínio, podem ser particularmente promissoras em cenários com alta heterogeneidade neural.

Por fim, a integração do ear-EEG com outras modalidades fisiológicas e o uso de modelos híbridos que combinem simplicidade estatística com representações temporais mais ricas podem representar caminhos viáveis para aplicações práticas, como dispositivos auditivos inteligentes ou interfaces cérebro-computador discretas. Assim, embora os resultados atuais indiquem limitações claras de desempenho, eles também delineiam um panorama realista e construtivo para o avanço da decodificação da atenção auditiva baseada em ear-EEG.

## REFERÊNCIAS

- [1] Netiwit Kaongoen, Jaehoon Choi, Jin Woo Choi, Haram Kwon, Chaeun Hwang, Guebin Hwang, Byung Hyung Kim, and Sungho Jo. The future of wearable EEG: A review of ear-EEG technology and its applications. *Journal of neural engineering*, 2023. doi: 10.1088/1741-2552/acfdca.
- [2] Takashi Nakamura, Yousef D. Alqurashi, Mary J. Morrell, and Danilo P. Mandic. Hearables: Automatic overnight sleep monitoring with standardized in-ear eeg sensor. *IEEE Transactions on Biomedical Engineering*, 67:203–212, 2020. URL <https://api.semanticscholar.org/CorpusID:133605312>.
- [3] Ryan Kaveh, Carolyn Schwendeman, Leslie Pu, Ana C Arias, and Rikky Muller. Wireless ear EEG to monitor drowsiness. *Nature Communications*, 15(1):6520, 2024. doi: 10.1038/s41467-024-48682-7.
- [4] Altyngul T Kamzanova, Almira M Kustubayeva, and Gerald Matthews. Use of eeg workload indices for diagnostic monitoring of vigilance decrement. *Human factors*, 56(6):1136–1149, 2014.
- [5] Miguel Bhagubai, Kaat Vandecasteele, Lauren Swinnen, Jaiver Macea, Christos Chatzichristos, Maarten De Vos, and Wim Van Paesschen. The power of ECG in semi-automated seizure detection in addition to two-channel behind-the-ear EEG. *Bioengineering*, 10(4):491, 2023. doi: 10.3390/bioengineering10040491.
- [6] Siqi Cai, Hongxu Zhu, Tanja Schultz, Haizhou Li, et al. Eeg-based auditory attention detection in cocktail party environment. *APSIPA Transactions on Signal and Information Processing*, 12(3), 2023.
- [7] Akshay Paul, Gopabandhu Hota, Behnam Khaleghi, Yuchen Xu, Tajana Rosing, and Gert Cauwenberghs. Attention state classification with in-ear EEG. In *2021 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–5. IEEE, 2021.
- [8] Martin G. Bleichner. Studying individual noise disturbance using long term ear-ecg (electroencephalography) recordings in everyday life. In *23rd International Congress on Acoustics*, pages 3026–3032, 2019.
- [9] Simon Geirnaert, Simon L. Kappel, and Preben Kidmose. A direct comparison of simultaneously recorded scalp, around-ear, and in-ear eeg for neural selective auditory attention decoding to speech, 2025. URL <https://arxiv.org/abs/2505.14478>.
- [10] Haolin Zhu, Yujie Yan, Xiran Xu, Zhongshu Ge, Pei Tian, Xihong Wu, and Jing Chen. Using Ear-EEG to decode auditory attention in multiple-speaker environment. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2025.
- [11] Björn Holtze, Marc Rosenkranz, Manuela Jaeger, Stefan Debener, and Bojana Mirkovic. Ear-EEG measures of auditory attention to continuous speech. *Frontiers in Neuroscience*, 16:869426, 2022. doi: 10.3389/fnins.2022.869426.
- [12] Mike D. Thornton, Danilo P. Mandic, and Tobias Reichenbach. Comparison of linear and nonlinear methods for decoding selective attention to speech from ear-ecg recordings. *IEEE Access*, 13:127614–127625, 2025. doi: 10.1109/ACCESS.2025.3590490.
- [13] Isabela Mary Sasaki, Wemerson Delcio Parreira, Fábio Pires Itturiet, Renata Coelho Borges, et al. Decodificação de atenção auditiva com ear-ecg: Uma revisão sistemática da literatura. *Anais do Computer on the Beach*, 16:593–595, 2025.
- [14] Paul L Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford university press, 2006.
- [15] Sheng Qu, Xinchun Wu, Yaxiu Tang, Qi Zhang, Laigang Huang, Baojuan Cui, Shengxiu Jiao, Qiangsan Sun, and Fanshuo Zeng. Analyzing brain-activation responses to auditory stimuli improves the diagnosis of a disorder of consciousness by non-linear dynamic analysis of the eeg. *Scientific Reports*, 14(1):17446, 2024.
- [16] Michael Thornton, Tobias Reichenbach, and Danilo Mandic. Imperial college Ear-EEG dataset for auditory attention decoding, July 2024. URL <https://doi.org/10.5281/zenodo.10260082>.
- [17] Martin G. Bleichner, Bojana Mirkovic, and Stefan Debener. Identifying auditory attention with ear-ecg: ceegrid versus high-density cap-ecg comparison. *Journal of Neural Engineering*, 13(6):066004, December 2016. ISSN 1741-2552. doi: 10.1088/1741-2560/13/6/066004.
- [18] Nai Ding and Jonathan Z Simon. Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in human neuroscience*, 8:311, 2014.
- [19] Rob Drullman, Joost M Festen, and Reinier Plomp. Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, 95(5):2670–2680, 1994.
- [20] Alan J Power, Natasha Mead, Lisa Barnes, and Usha Goswami. Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in human neuroscience*, 7:777, 2013.
- [21] Karl Pearson. of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Breakthroughs in Statistics: Methodology and Distribution*, page 11, 1991.
- [22] Joost CF De Winter, Samuel D Gosling, and Jeff Potter. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, 21(3):273, 2016.
- [23] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2): 81–93, 1938.
- [24] Michael J Crosse, Giovanni M Di Liberto, Adam Bednar, and Edmund C Lalor. The multivariate temporal response function (mtrf) toolbox: a matlab toolbox for relating neural signals to continuous stimuli. *Frontiers in human neuroscience*, 10:604, 2016.