

# Autenticação Biométrica de Locutores Drasticamente Disfônicos Aprimorada pela *Imagined Speech*

André Furlan - [ensismoebius@gmail.com](mailto:ensismoebius@gmail.com)

Universidade Estadual Paulista Júlio de Mesquita Filho

2024

# Introdução

A autenticação biométrica tem sido amplamente adotada por fornecer individualização baseada em traços únicos dos indivíduos. No entanto, características físicas que diferem das esperadas pelo sistema podem dificultar ou impedir o acesso de certas pessoas, especialmente aquelas pertencentes a grupos étnicos minorizados ou com deficiências. No caso da autenticação por voz, a exigência de uma palavra-chave ou outro tipo de fonação pode criar barreiras adicionais. Para mitigar esse problema, este trabalho propõe um sistema de autenticação biométrica aprimorado pela fala imaginada.

- ▶ Criar um sistema de autenticação por voz aprimorado pela fala imaginada.
- ▶ Comparar o desempenho das estratégias de extração de características automatizadas com o uso de autoencoders com as que usam *Wavelet-Packet* de tempo discreto e engenharia paraconsistente de características.
- ▶ Criar uma base com dados de autenticação voz, fala imaginada e ambos os sinais simultâneos fornecendo a comunidade científica mais uma fonte de dados.

## Estrutura da apresentação

- ▶ Revisão de conceitos utilizados.
- ▶ Estado-da-arte em autenticação por voz e fala imaginada.
- ▶ Abordagem proposta.

## Revisão de conceitos utilizados

Para desenvolver o trabalho, é necessário definir os conceitos de *downsampling*, filtros digitais *wavelets* e energia, que serão usados para a extração de características dos sinais. No processamento dos sinais de voz, serão revisadas as frequências fundamentais da voz, além das escalas BARK e MEL. Também será revisada a engenharia paraconsistente de características, método utilizado para verificar a qualidade das características extraídas. Em seguida, serão estudadas as bases das interfaces Humano-Máquina para eletroencefalograma, que coletarão dados das regiões do cérebro. Concluída essa etapa, serão revisadas as redes neurais de pulso, autoencoders e redes neurais residuais.



# Sinais Digitais e Sub-amostragem (Downsampling)



Figura: Sub-amostragem

Mesmo depois de amostrados e quantizados (HAYKIN; MOHER, 2011) os sinais de EEG e voz, na maioria das vezes, precisam ser sub-amostrados para que se sejam viáveis o processamento e o armazenamento (POLIKAR et al., 1996).

Além disso o *downsampling* será fundamental no uso das transformadas *wavelet*.

# Caracterização dos processos de produção da voz humana

## Áreas de estudo

- ▶ fisiológica ou “fonética articulatória”.
- ▶ acústica ou “fonética acústica”.
- ▶ perceptual.

Neste trabalho, o foco será apenas na questão acústica, pois não serão analisados aspectos da fisiologia relacionada à voz, mas sim os sinais sonoros propriamente ditos.

# Caracterização dos processos de produção da voz humana

## Vozeada versus não-vozeada

- ▶ vozeada: Pregas vocais.
- ▶ não vozeada: Sem pregas vocais.

# Caracterização dos processos de produção da voz humana

## Frequência fundamental da voz

- ▶ conhecida como  $F_0$  representa o tom ou *pitch* da voz (KREMER; GOMES, 2014).
- ▶ reflete a excitação pulmonar moldada pelas pregas vocais.
- ▶ componente periódico resultante da vibração das pregas vocais.
- ▶ geralmente são apresentadas em Hz (FREITAS, 2013)

A alteração desta frequência (jitter) e/ou intensidade (shimmer) do *pitch* durante a fala é definida como entonação, porém, também pode indicar algum distúrbio ou doença relacionada ao trato vocal (WERTZNER; SCHREIBER; AMARO, 2005).

# Caracterização dos processos de produção da voz humana

## Formantes

Se referem as modificações feitas em  $F_0$  pelas estruturas do sistema fonador (VALENÇA et al., 2014):

- ▶  $F_1 \rightarrow$  amplificação sonora na cavidade oral posterior, posição da língua no plano vertical.
- ▶  $F_2 \rightarrow$  cavidade oral anterior, posição da língua no plano horizontal.
- ▶  $F_3 \rightarrow$  cavidades à frente e atrás do ápice da língua.
- ▶  $F_4 \rightarrow$  formato da laringe e da faringe.

# Bandas críticas de energia

## Definições

Definem os intervalos em que serão calculadas as energias.

A energia de um sinal digital  $s[\cdot]$  com  $M$  amostras é definida como

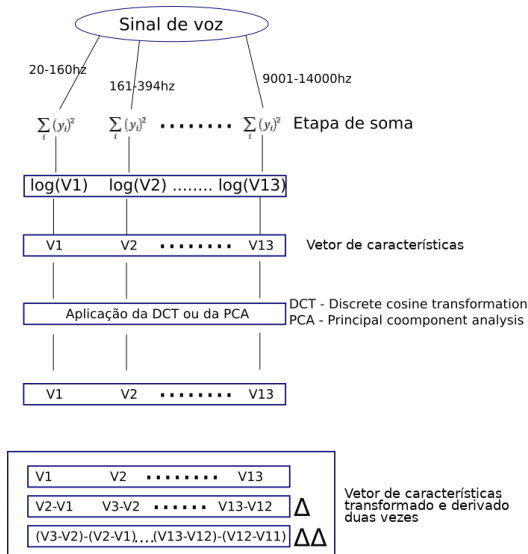
$$E = \sum_{i=0}^{M-1} (s_i)^2 \quad . \quad (1)$$

- ▶ **BARK:** 20, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500 (bandas de audição humana) (ZWICKER, 1961).
- ▶ **MEL:** 20, 160, 394, 670, 1000, 1420, 1900, 2450, 3120, 4000, 5100, 6600, 9000, 14000 (bandas de audição adaptada para voz) (BERANEK, 1949)



# Bandas críticas de energia

## Cálculo de vetores de características com MEL





- ▶ suporte compacto.
- ▶ análise multirresolução.
- ▶ wavelet regular e wavelet packet.
- ▶ variadas funções-filtro não-periódicas.
- ▶ análise detalhada em altas e baixas frequências.

# Filtros digitais *wavelet*

## Restrição de escopo

- ▶ wavelet packet
- ▶ domínio discreto.
- ▶ apenas transformadas diretas.
- ▶ não haverá reconstrução do sinal.
- ▶ construção dos vetores de características.

# Filtros digitais *wavelet*

Porquê e *wavelet packet*?

- ▶ decompõe a aproximação e os detalhes
- ▶ proporciona um nível de decomposição maior

# Filtros digitais *wavelet*

Resposta em frequência e linearidade

**Tabela:** Algumas das *wavelets* mais usadas e suas propriedades

| Wavelet    | Resposta em frequência  | Resposta em fase |
|------------|---|------------------|
| Haar       | Pobre   | Linear           |
| Daubechies | mais próxima da ideal à medida que o suporte aumenta; <i>maximally-flat</i>     | Não linear       |
| Symmlets   | mais próxima da ideal à medida que o suporte aumenta; não <i>maximally-flat</i> | Quase linear     |
| Coiflets   | mais próxima da ideal à medida que o suporte aumenta; não <i>maximally-flat</i> | Quase linear     |

Fonte: Elaborado pelo autor, 2022.

# Filtros digitais *wavelet*

## Algoritmo de Malat

- ▶ *Wavelet* Haar:  $h[\cdot] = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ .
- ▶ Par ortogonal:  $g[\cdot] = [\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}]$ .
- ▶ sinal:  $s[\cdot] = [1, 2, 3, 4]$ .

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} \frac{3}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ \frac{7}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix} \Rightarrow \left[ \frac{3}{\sqrt{2}}, \frac{7}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right] .$$

# Filtros digitais *wavelet*

## Exemplo de *wavelet* regular e *packet*

### Packet-wavelet para filtro haar = {0.5, 0.5}

|          |         |        |         |         |        |         |         |         |        |        |        |        |        |         |        |        |
|----------|---------|--------|---------|---------|--------|---------|---------|---------|--------|--------|--------|--------|--------|---------|--------|--------|
| Sinal    | 32      | 10     | 20      | 38      | 37     | 28      | 38      | 34      | 18     | 24     | 24     | 9      | 23     | 24      | 28     | 34     |
|          | h       |        |         |         |        |         |         |         | g      |        |        |        |        |         |        |        |
| Nível 01 | 21      | 29     | 32,5    | 36      | 21     | 16,5    | 23,5    | 31      | 11     | -9     | 4,5    | 2      | -3     | 7,5     | -0,5   | -3     |
|          | h       |        |         |         | g      |         |         |         | g      |        |        |        | h      |         |        |        |
| Nível 02 | 25      | 34,25  | 18,75   | 27,25   | -4     | -1,75   | 2,25    | -3,75   | 10     | 1,25   | -5,25  | 1,25   | 1      | 3,25    | 2,25   | -1,75  |
|          | h       |        | g       |         | g      |         | h       |         | h      |        | g      |        | g      |         | h      |        |
| Nível 03 | 29,62   | 23     | -4,625  | -4,25   | -1,125 | 3       | -2,875  | -0,75   | 5,625  | -2     | 4,375  | -3,25  | -1,125 | 2       | 2,125  | 0,25   |
|          | h       |        | g       |         | g      |         | h       |         | h      |        | g      |        | h      |         | g      |        |
| Nível 04 | 26,3125 | 3,3125 | -0,1875 | -4,4375 | 0,9375 | -2,0625 | -1,0625 | -1,8125 | 1,8125 | 3,8125 | 3,8125 | 0,5625 | 0,4375 | -1,5625 | 0,9375 | 1,1875 |

|  |                                 |
|--|---------------------------------|
|  | ← resultado do passa-baixas (h) |
|  | ← resultado do passa-altas (g)  |

### Comparação

#### Transformada wavelet regular

|          |         |        |        |       |    |       |      |       |    |    |     |   |    |     |      |    |
|----------|---------|--------|--------|-------|----|-------|------|-------|----|----|-----|---|----|-----|------|----|
| Sinal    | 32      | 10     | 20     | 38    | 37 | 28    | 38   | 34    | 18 | 24 | 24  | 9 | 23 | 24  | 28   | 34 |
| Nível 01 | 21      | 29     | 32,5   | 36    | 21 | 16,5  | 23,5 | 31    | 11 | -9 | 4,5 | 2 | -3 | 7,5 | -0,5 | -3 |
| Nível 02 | 25      | 34,25  | 18,75  | 27,25 | -4 | -1,75 | 2,25 | -3,75 | 11 | -9 | 4,5 | 2 | -3 | 7,5 | -0,5 | -3 |
| Nível 03 | 29,62   | 23     | -4,625 | -4,25 | -4 | -1,75 | 2,25 | -3,75 | 11 | -9 | 4,5 | 2 | -3 | 7,5 | -0,5 | -3 |
| Nível 04 | 26,3125 | 3,3125 | -4,625 | -4,25 | -4 | -1,75 | 2,25 | -3,75 | 11 | -9 | 4,5 | 2 | -3 | 7,5 | -0,5 | -3 |

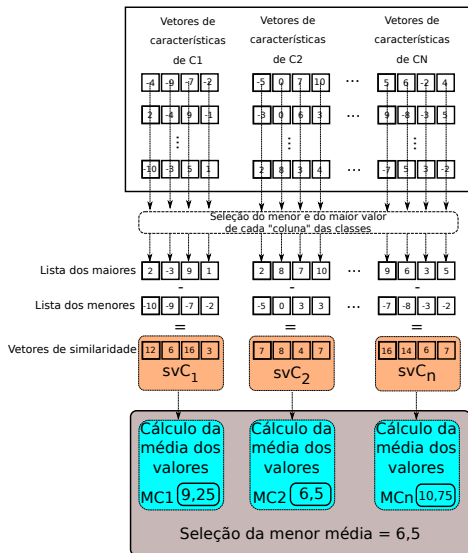
#### Transformada packet-wavelet

|          |         |        |         |         |        |         |         |         |        |        |        |        |        |         |        |        |
|----------|---------|--------|---------|---------|--------|---------|---------|---------|--------|--------|--------|--------|--------|---------|--------|--------|
| Sinal    | 32      | 10     | 20      | 38      | 37     | 28      | 38      | 34      | 18     | 24     | 24     | 9      | 23     | 24      | 28     | 34     |
| Nível 01 | 21      | 29     | 32,5    | 36      | 21     | 16,5    | 23,5    | 31      | 11     | -9     | 4,5    | 2      | -3     | 7,5     | -0,5   | -3     |
| Nível 02 | 25      | 34,25  | 18,75   | 27,25   | -4     | -1,75   | 2,25    | -3,75   | 10     | 1,25   | -5,25  | 1,25   | 1      | 3,25    | 2,25   | -1,75  |
| Nível 03 | 29,62   | 23     | -4,625  | -4,25   | -1,125 | 3       | -2,875  | -0,75   | 5,625  | -2     | 4,375  | -3,25  | -1,125 | 2       | 2,125  | 0,25   |
| Nível 04 | 26,3125 | 3,3125 | -0,1875 | -4,4375 | 0,9375 | -2,0625 | -1,0625 | -1,8125 | 1,8125 | 3,8125 | 3,8125 | 0,5625 | 0,4375 | -1,5625 | 0,9375 | 1,1875 |



# Engenharia paraconsistente de características

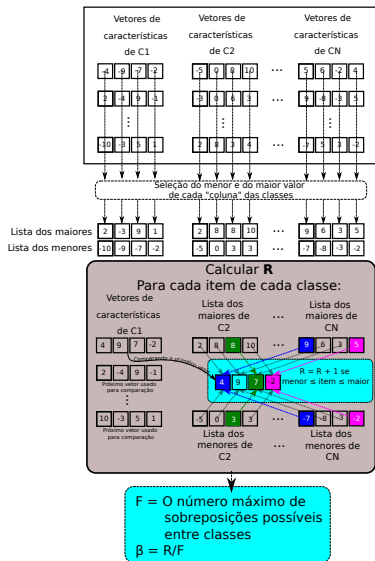
## Cálculo de $\alpha$



Menor similaridade intraclasse,  $\alpha$

# Engenharia paraconsistente de características

## Cálculo de $\beta$



Razão de sobreposição interclasse,  $\beta$ .  
$$F = N.(N - 1).X.T$$

- ▶ N a quantidades de classes;
- ▶ X a quantidade de vetores de características por classe;
- ▶ T o tamanho do vetor de características.



# Engenharia paraconsistente de características

## Graus de certeza e contradição

- ▶ Grau de certeza  $\rightarrow G_1 = \alpha - \beta$ .
- ▶ Grau de contradição  $\rightarrow G_2 = \alpha + \beta - 1$ .

Onde:  $-1 \leq G_1 \leq 1$  e  $-1 \leq G_2 \leq 1$  .

Seja  $P = (G_1, G_2)$

- ▶ Verdade  $\rightarrow$  fé total ( $\alpha = 1$ ) e nenhum descrédito ( $\beta = 0$ )
- ▶ Ambiguidade  $\rightarrow$  fé total ( $\alpha = 1$ ) e descrédito total ( $\beta = 1$ )
- ▶ Falsidade  $\rightarrow$  fé nula ( $\alpha = 0$ ) e descrédito total ( $\beta = 1$ )
- ▶ Indefinição  $\rightarrow$  fé nula ( $\alpha = 0$ ) e nenhum descrédito ( $\beta = 0$ ) .

# Engenharia paraconsistente de características

## Distancias no plano paraconsistente

As distâncias( $D$ ) do ponto  $P = (G_1, G_2)$  dos limites supracitados. Tal cálculo pode ser feito da seguinte forma:

$$D_{-1,0} = \sqrt{(G_1 + 1)^2 + (G_2)^2} \quad ,$$

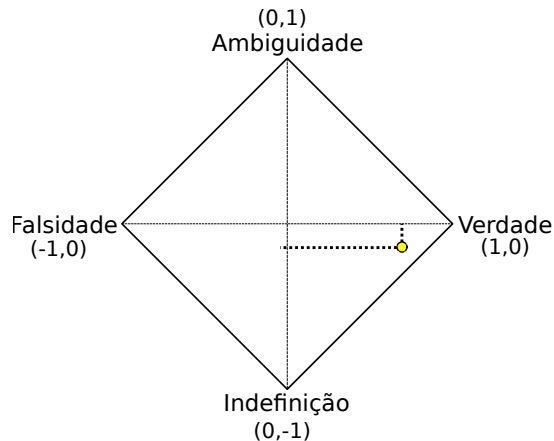
$$D_{1,0} = \sqrt{(G_1 - 1)^2 + (G_2)^2} \quad ,$$

$$D_{0,-1} = \sqrt{(G_1)^2 + (G_2 + 1)^2} \quad ,$$

$$D_{0,1} = \sqrt{(G_1)^2 + (G_2 - 1)^2} \quad ,$$

# Engenharia paraconsistente de características

Os vetores de características proporcionam uma boa separação interclasses?(GUIDO, 2019)



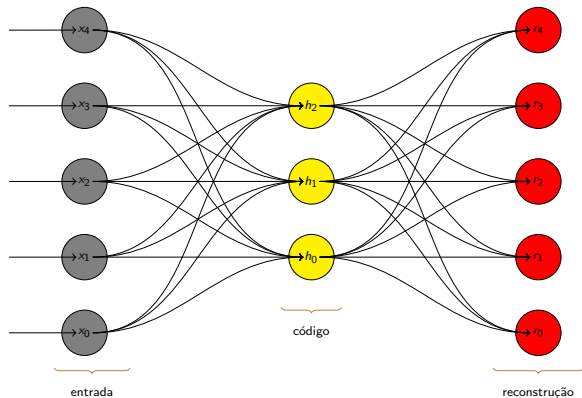
- ▶ **Verdade:**  
 $\alpha = 1$  e  $\beta = 0$ .
- ▶ **Ambiguidade:**  
 $\alpha = 1$  e  $\beta = 1$ .
- ▶ **Falsidade:**  
 $\alpha = 0$  e  $\beta = 1$ .
- ▶ **Indefinição:**  
 $\alpha = 0$  e  $\beta = 0$ .

- ▶ serão usados os subcompletos.
- ▶ a saída é uma aproximação da entrada.
- ▶ seleciona características em sua camada de código.

# Autoencoders

## Subcompletos

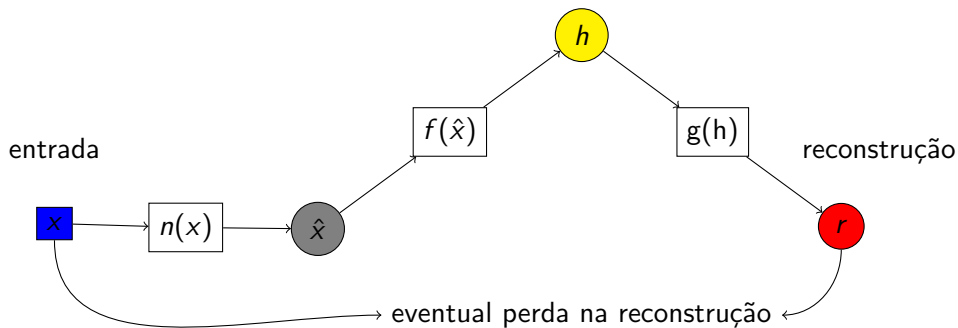
Exemplo esquemático de um *autoencoder*: Os nós de entrada estão em cinza, a camada de código em amarelo contêm as características codificadas e finalmente a camada de reconstrução em vermelho contém uma cópia aproximada da entrada.



# Autoencoders

## Denoising

Representação funcional de um *denoising autoencoder*: Sendo o vetor  $x$  uma entrada e  $n$  um função que adiciona um ruído aleatório então a informação com ruído  $\hat{x}$  é definida como  $\hat{x} = n(x)$ , portanto o vetor  $h$  é resultado da aplicação de uma função codificadora  $f$  sobre  $x$ :  $h = f(x)$ , finalmente  $r$  é a reconstrução de  $x$  a partir  $h$  através de uma função decodificadora  $g$ :  $r = g(h)$ . É importante notar que  $r$  é comparada com  $x$  e não com  $\hat{x}$ .



# Fala imaginada

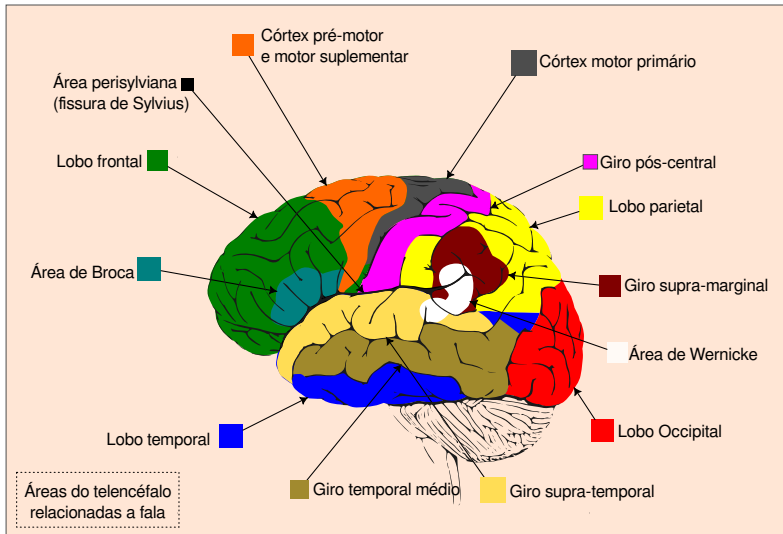
A fala imaginada é um fenômeno em que uma pessoa "ouve" a si mesma falando internamente, sem produzir som. As regiões cerebrais usadas na fala imaginada são similares às que estão envolvidas na fala verbal, englobando áreas do córtex motor e pré-motor, além de outras áreas associadas à linguagem e a fala.

A ativação cerebral durante processos linguísticos (incluindo a fala imaginada) envolve as seguintes regiões (PINTO, 2012), (VANDERAH, 2020), (VASKOVIĆ ALEXANDRA OSIKA, 2024):

- ▶ **área de Broca:** Produção da fala.
- ▶ **área de Wernicke:** Compreensão da linguagem.
- ▶ **córtex motor e pré-motor:** Articulação da fala.
- ▶ **giro supramarginal:** Planejamento motor da fala.
- ▶ **giro supratemporal e médio:** Compreensão linguística e articulação durante a repetição de sílabas.

# Fala imaginada

## Visão geral do telencéfalo





# Interfaces Humano-Máquina e EEG

## EEG e as frequências do cérebro

### Eletroencefalograma:

- ▶ não invasivo
  - ▶ mais sujeito a ruídos
  - ▶ exige tolerância a ruído
  - ▶ mais econômico e simples de implementar
  - ▶ eletrodos secos (reutilizáveis, mais interferência)
  - ▶ eletrodos úmidos (não reutilizáveis, menos interferência)
- ▶ **Delta (1–4Hz):** A onda mais lenta. Observada em bebês e durante o sono profundo em adultos.
  - ▶ **Theta (4–8Hz):** Observada em crianças, adultos sonolentos e durante a recordação de memórias.
  - ▶ **Alpha (8–12Hz):** Geralmente a banda de frequência dominante, aparecendo durante a consciência relaxada ou quando os olhos estão fechados.
  - ▶ **Beta (12–25Hz):** Associada ao pensamento, concentração ativa e atenção focada.
  - ▶ **Gamma (acima de 25Hz):** Observada durante o processamento sensorial múltiplo.

# Interfaces Humano-Máquina e EEG

Posicionamento dos eletrodos de acordo com o padrão 10-20. Números ímpares são atribuídos aos eletrodos no hemisfério esquerdo, e números pares são atribuídos aos eletrodos no hemisfério direito (VICENTE, 2023), (BIDGOLY; BIDGOLY; AREZOUMAND, 2020).

# Redes neurais de pulso (RNP)

## Características

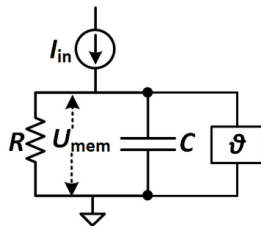
- ▶ funciona segundo limiares de ativação.
- ▶ redes esparsas (baixa ativação sináptica).
- ▶ uso de pulsos em vez de valores contínuos.
- ▶ processamento orientado ao tempo (ESHRAGHIAN et al., 2023).
- ▶ tolerantes a ruídos (atuam como um filtro passa baixa).
- ▶ não um simulação de um-para-um como o textbfneurônio Hodgkin-Huxley. (GERSTNER et al., 2014) ou outros (JONES; KORDING, 2020).

# Redes neurais de pulso (RNP)

## *Leaky Integrate and Fire Neurons (LIF)*

- ▶ se assemelha com circuitos Resistor-Capacitor.
- ▶ os pulsos são representados como **uns** dispersamente distribuídos em uma sequência de **zeros**
- ▶ simples, mais eficientes e atualmente generalizam melhor para a maioria dos problemas (GOODMAN et al., 2022).

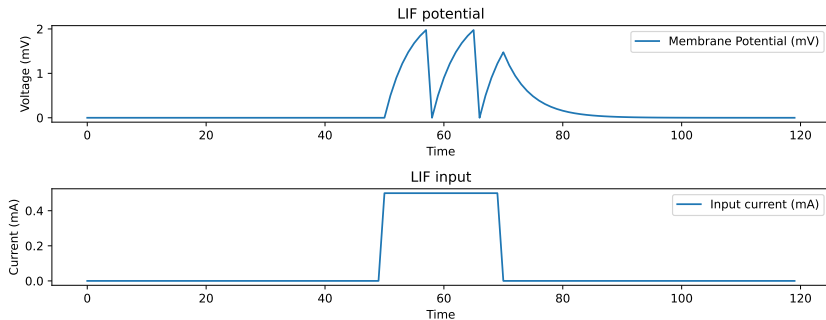
Figura: O modelo RC



# Redes neurais de pulso (RNP)

## *Leaky Integrate and Fire Neurons (LIF)*

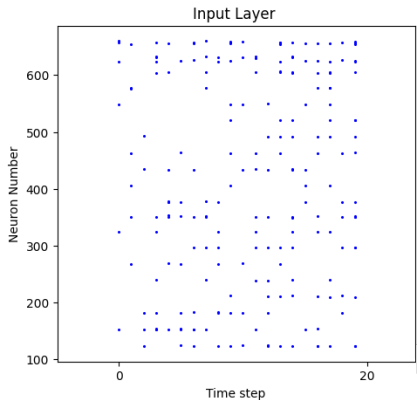
Gráfico do LIF simulado completo.



# Redes neurais de pulso (RNP)

## Esparsidade

Atividade dispersa de uma RNP: O eixo horizontal representa o momento no qual os dados estão sendo processados e o vertical representa o número do neurônio (índice) na RNP. Note que, na maior parte do tempo, muito poucos neurônios são ativados.



# Redes neurais de pulso (RNP)

## Treinamento

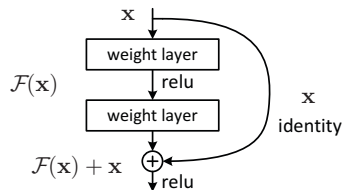
- ▶ **Plasticidade Dependente do Tempo de Pulsos (STDP):** Se um neurônio pré-sináptico dispara **antes** do pós-sináptico, há um fortalecimento na conexão, mas se o neurônio pós-sináptico disparar antes, então há um enfraquecimento.
- ▶ **Descida de Gradiente Emprestada:** Aproxima a função de passo usando outra função matemática, que é diferenciável (como uma sigmoide), para treinar a rede. Essas aproximações são usadas apenas na fase de *backpropagation*, enquanto mantêm a função de passo na fase do *feed-forward*.
- ▶ **Algoritmos Evolutivos:** Usam a seleção dos mais aptos ao longo de muitas gerações de redes.
- ▶ **Reservatório/Computação Dinâmica:** Redes de estado de eco ou Máquinas de estado líquido, respectivamente.

# Redes neurais residuais (ResNets)

Redes neurais profundas podem sofrer com desaparecimento ou explosão do gradiente. A solução é reproduzir o comportamento de redes neurais mais rasas usando conexões de salto. Essa é ideia das redes neurais residuais.



# Redes neurais residuais (ResNets)

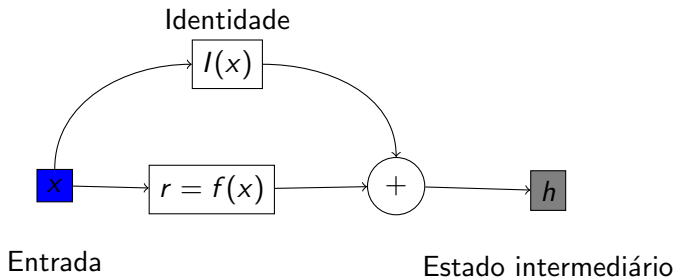


Segundo (HE et al., 2015) a ideia-chave por trás das *ResNets* é a inclusão de conexões de salto como ilustrado na figura ao lado, também conhecidas como mapeamentos de identidade.

Bloco Residual:  $x$  contorna as camadas intermediárias  $F(x)$  via uma função identidade somando-se a  $F(x)$  ao final do bloco.

# Redes neurais residuais (ResNets)

Obtenção do resíduo:  $x$  um vetor de entrada na rede,  $I(\cdot)$  é uma função identidade cujo valor é diretamente somado  $\oplus$  com o resultado produzido pelas várias camadas representadas por  $f(\cdot)$ ,  $h$  é um vetor representando um estado intermediário de uma rede neural.



## Estado-da-arte em autenticação por voz e fala imaginada

## ► Classificação e tratamento de EEG:

- O estudo de (PARK; LEE, 2023) explora o uso de EEG para fala imaginada com NA-MEMD e DTWPT, usando MRF-CNN para classificar sinais decompostos. A acurácia média foi 80,41%.
- Em (ABDULGHANI; WALTERS; ABED, 2023), utilizou-se Wavelet Scattering Transform (WST) e uma Rede Neural LSTM para classificar expressões imaginadas, alcançando 92,50% de acurácia.
- O estudo (MAHAPATRA; BHUYAN, 2023) usou DTWT para processar sinais EEG e alcançou acurácias de 98,18% e 71,60% nas bases MUSE e EPOC, respectivamente.
- A revisão em (SHAH et al., 2022) destacou o uso de filtros Wavelet e algoritmos SVM e CNN, sugerindo a inclusão de métricas como Precisão e Recall para uma visão mais completa.
- O estudo (MAHAPATRA; BHUYAN, 2022) aplicou uma combinação de TCN e CNN com FastICA e DTWT para sinais EEG, alcançando 96,49% de acurácia.
- (AGARWAL; KUMAR, 2022) usou DTWT e dois algoritmos (RF e SVM) para classificar letras imaginadas, obtendo 77,97% de acurácia. Utilizou também CAR para melhorar a SNR.

- ▶ O trabalho em (HERNÁNDEZ-DEL-TORO; REYES-GARCÍA; PINEDA, 2021) usou DTWT, EMD e características baseadas em teoria do caos, alcançando melhores resultados com RF, KNN, SVM e Regressão Logística.
- ▶ Em (MOCTEZUMA et al., 2019), CAR foi combinado com DTWT e TEO para melhorar a SNR, alcançando uma acurácia de 97% com 5 voluntários.
- ▶ (PANACHAKEL; RAMAKRISHNAN, 2021) usou janelas deslizantes e matrizes tridimensionais para classificar sinais EEG, alcançando acurácias de 79,7% a 95,5%.
- ▶ (TAMM; MUHAMMAD; MUHAMMAD, 2020) simplificou uma CNN com transferência de aprendizado, alcançando uma acurácia de 23,98% em uma base de dados reduzida.
- ▶ O estudo de (PANACHAKEL; RAMAKRISHNAN; ANANTHAPADMANABHA, 2019) aplicou transformadas Wavelet de nível 7 e obteve 57,15% de acurácia ao tratar separadamente os vetores de características.
- ▶ (PANACHAKEL; RAMAKRISHNAN; ANANTHAPADMANABHA, 2020) usou uma rede neural profunda e transformada wavelet para duas palavras, alcançando 71,8% de acurácia.

- ▶ A revisão em (SEN et al., 2023) mostrou que SVM, RF, HMM e GMM foram comuns, com técnicas profundas como CNN e RNN emergindo desde 2020.
- ▶ **Classificação e tratamento de voz:**
  - ▶ (WANG et al., 2023) combinou Res2Net com PWPE e ECA, alcançando menos de 8% de EER em contextos variados.
  - ▶ O estudo de (ALI et al., 2022) focou na classificação do gênero usando MFCC e LPC, alcançando 97,07% de acurácia com redes neurais artificiais.
  - ▶ (MENG; ALTAF; JUANG, 2020) introduziu o sistema AVA para autenticação ativa, com uma taxa média de WEER de 3% a 4%.
  - ▶ O estudo (GAO et al., 2021) explorou microfones em fones para criar impressões digitais auditivas, alcançando um EER de 3,64%.
  - ▶ (JIANG et al., 2023) usou sons de estalos para prevenção de spoofing, alcançando uma classificação baseada em vetores de características de 3 dimensões.
  - ▶ O estudo (ANDRÉ, 2021) comparou métodos baseados em distâncias e SVM, alcançando mais de 99% de precisão usando a escala BARK e wavelet Haar.

Os estudos analisados contribuíram para a construção do protocolo de coleta de dados, abordando a importância de filtrar interferências da rede elétrica local e a separabilidade dos canais EEG usando o algoritmo *FastICA*. A fase de extração de características é crucial para o desempenho dos algoritmos de classificação. Também é importante a separação dos indivíduos por sexo, destros e canhotos, além da consideração das ondas cerebrais  $\alpha$ ,  $\beta$  e  $\theta$  isoladamente ou em conjunto. Técnicas como DTWT e CAR são comuns na literatura. Estudos demonstram que redes neurais profundas não necessariamente superam redes rasas, e o uso de sinais EEG separados por canal para aumentar a quantidade de dados é uma abordagem interessante. Além disso, técnicas para extração de características continuam como uma das partes mais importantes no projeto de sistemas de análise e reconhecimento de sinais.

## Escalas de energia usadas

- ▶ Energia de Teager (TEO): Usada para extração de características em (MOCTEZUMA et al., 2019).
- ▶ Energia de bandas: Obtida usando DTWT para cada banda de frequência em (HERNÁNDEZ-DEL-TORO; REYES-GARCÍA; PINEDA, 2021).

## Tipos de classificadores usados

- ▶ Redes Neurais Convolucionais (CNN): Usadas em (PANACHAKEL; RAMAKRISHNAN, 2021), (MAHAPATRA; BHUYAN, 2022).
- ▶ Redes Neurais Recorrentes LSTM (Long-Short Term Memory): Usadas em (ABDULGHANI; WALTERS; ABED, 2023).
- ▶ Support Vector Machines (SVM): Usadas em (SHAH et al., 2022), (HERNÁNDEZ-DEL-TORO; REYES-GARCÍA; PINEDA, 2021).



- ▶ Random Forest (RF): Usado em (AGARWAL; KUMAR, 2022), (HERNÁNDEZ-DEL-TORO; REYES-GARCÍA; PINEDA, 2021), (MOCTEZUMA et al., 2019).
- ▶ Bidirectional Recurrent Neural Networks: Usadas em (MAHAPATRA; BHUYAN, 2023).
- ▶ Deep Neural Networks (DNN): Usadas em (PANACHAKEL; RAMAKRISHNAN; ANANTHAPADMANABHA, 2019), (PANACHAKEL; RAMAKRISHNAN; ANANTHAPADMANABHA, 2020).
- ▶ Hidden Markov Model (HMM): Usado em (MENG; ALTAF; JUANG, 2020).
- ▶ Análise Discriminante e redes neurais artificiais: Usadas em (ALI et al., 2022).
- ▶ ResNet50 com transferência de aprendizado: Usado em (PANACHAKEL; RAMAKRISHNAN, 2021).

### Técnicas de extração de características

## Comparativo - Referências III

- ▶ Transformada Wavelet Discreta (DTWT): Usada em (MAHAPATRA; BHUYAN, 2023), (HERNÁNDEZ-DEL-TORO; REYES-GARCÍA; PINEDA, 2021).
- ▶ Empirical Mode Decomposition (EMD): Usada em (PARK; LEE, 2023).
- ▶ Wavelet Scattering Transform (WST): Usada em (ABDULGHANI; WALTERS; ABED, 2023) e (PANACHAKEL; RAMAKRISHNAN; ANANTHAPADMANABHA, 2019).
- ▶ FastICA (Independent Component Analysis): Usada em (MAHAPATRA; BHUYAN, 2022).
- ▶ Mel-Frequency Cepstral Coefficients (MFCC): Usada em (ALI et al., 2022), (MENG; ALTAF; JUANG, 2020).
- ▶ Linear Prediction Coefficients (LPC): Usados em (ALI et al., 2022).
- ▶ Teager Energy Operator (TEO): Usado em (MOCTEZUMA et al., 2019).
- ▶ Fourier Transform: Usada para alguns dados em (SEN et al., 2023).

- ▶ Common Average Referencing (CAR): Usado em (AGARWAL; KUMAR, 2022), (HERNÁNDEZ-DEL-TORO; REYES-GARCÍA; PINEDA, 2021), (MOCTEZUMA et al., 2019).
- ▶ Dimensionamento fractal e métodos da teoria do Caos: Usados em (HERNÁNDEZ-DEL-TORO; REYES-GARCÍA; PINEDA, 2021).

## Escalas de energia usadas

- ▶ energia das bandas na escala BARK
- ▶ energia das bandas na escala MEL
- ▶ energia das bandas  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\theta$  e  $\delta$

## Tipos de classificadores usados

- ▶ Redes neurais de pulso (SNN)
- ▶ Redes neurais residuais (RNN)

## Técnicas de extração de características

- ▶ Transformada Wavelet Discreta (DTWT)
- ▶ Mel-Frequency Cepstral Coefficients (MFCC)
- ▶ Auto-encoders

## Abordagem proposta

# A base de sinais

## Coleta dos dados

- ▶ questionário preliminar.
- ▶ ciclo silencioso e ruidoso.
- ▶ coleta da fala, EEG, EEG e fala.
- ▶ coleção de sentenças específicas.
- ▶ quantização de 16 bits para voz e EEG.
- ▶ taxa de amostragem 44100Hz para voz e 800Hz para EEG.

# A base de sinais

## Questionário preliminar

- ▶ ingeriu ou ingere algum medicamento? Há quanto tempo?
- ▶ usou alguma droga legal ou ilegal? Há quanto tempo?
- ▶ ingeriu alguma bebida energética? Há quanto tempo?
- ▶ é canhota, destra ou ambidestra?
- ▶ qual seu gênero?
- ▶ qual sua etnia?
- ▶ qual seu sexo?

# A base de sinais

## Protocolo de coleta da fala fonada

### 1. ciclo silencioso.

- 1.1 em uma tela é exibido ao participante por 5 segundos a sentença que deve ser pronunciada.
- 1.2 um sinal é tocado por 1 segundo
- 1.3 por 5 segundos a sentença deve ser **pronunciada** uma única vez.
- 1.4 volta para o item 1.1 até que todas as sentenças sejam pronunciadas.

### 2. ciclo ruidoso.

- 2.1 inicia-se a reprodução do ruído
- 2.2 em uma tela é exibido ao participante por 5 segundos a sentença que deve ser pronunciada.
- 2.3 um sinal é tocado por 1 segundo
- 2.4 por 5 segundos a sentença deve ser **pronunciada** uma única vez.
- 2.5 volta para o item 2.2 até que todas as sentenças sejam pronunciadas.
- 2.6 a reprodução do ruído é finalizada.



# A base de sinais

## Protocolo de coleta da fala imaginada

1. Coleta "pré-protocolar" de cinco segundos.
2. **ciclo silencioso.**
  - 2.1 em uma tela é exibido ao participante por 5 segundos a sentença que deve ser imaginada.
  - 2.2 um sinal é tocado por 1 segundo
  - 2.3 Por 5 segundos a sentença deve ser **imaginada** uma única vez.
  - 2.4 volta para o item 2.1 até que todas as sentenças sejam imaginadas.
3. **ciclo ruidoso.**
  - 3.1 inicia-se a reprodução do ruído
  - 3.2 em uma tela é exibido ao participante por 5 segundos a sentença que deve ser imaginada.
  - 3.3 um sinal é tocado por 1 segundo
  - 3.4 por 5 segundos a sentença deve ser **imaginada** uma única vez.
  - 3.5 volta para o item 3.1 até que todas as sentenças sejam pronunciadas.
  - 3.6 a reprodução do ruído é finalizada.

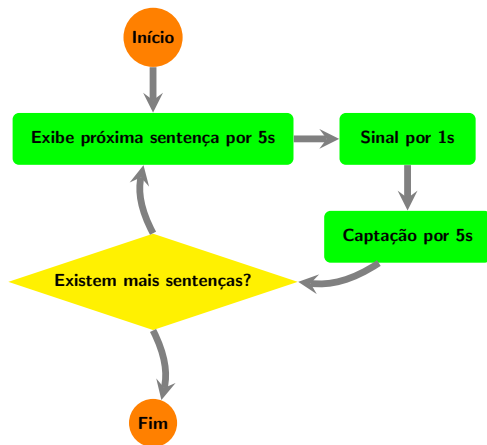
# A base de sinais

## Protocolo de coleta da fala mista

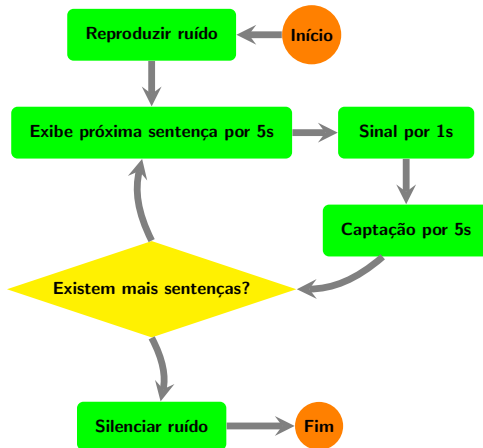
Deve seguir os mesmos protocolos para o ciclo silencioso e ruidoso anteriormente descritos porém, captando ambos os sinais fonados e imaginados simultaneamente;

# A base de sinais

## Protocolo de coleta de dados



(a) Protocolo de coleta silenciosa



(b) Protocolo de coleta ruidosa

# A base de sinais

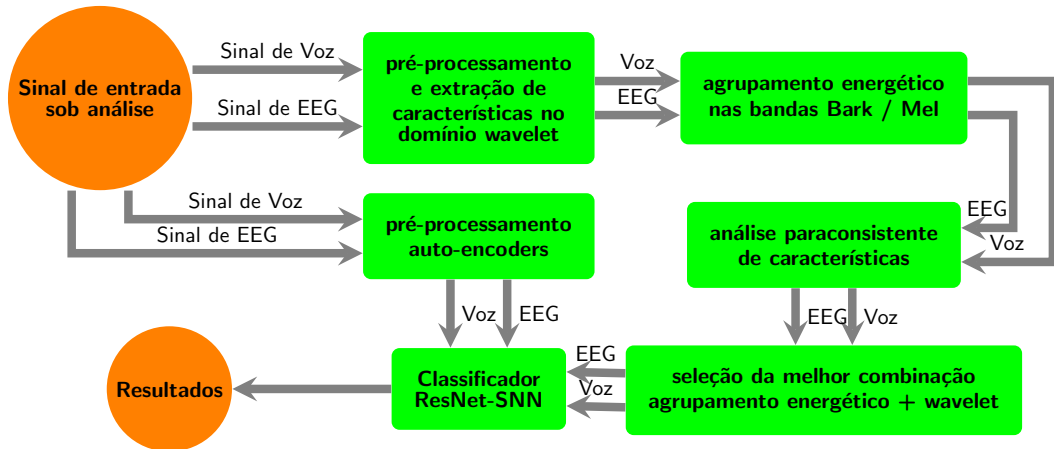
## Organização

- ▶ fonadas
  - ▶ com ruído
  - ▶ sem ruído
- ▶ imaginadas
  - ▶ com ruído
  - ▶ sem ruído
- ▶ mista (fonadas e imaginadas simultaneamente)
  - ▶ com ruído
  - ▶ sem ruído

Cada registro deve ser vinculado a um código que deve referenciar os dados da pessoa que o gerou, a data, hora, minuto e segundo em que a fala iniciou e finalizou, a sentença e o tipo de fala (fonada ou imaginada).

# Estrutura da Estratégia Proposta

## Diagrama



# Estrutura da Estratégia Proposta


Wavelets inicialmente usadas


- ▶ Haar.
- ▶ Beylkin com suporte 18.
- ▶ Vaidyanathan de suporte 24.
- ▶ Daubechies de suportes 4 até 76.
- ▶ Symmlets com suportes 8, 16 e 32.
- ▶ Coiflets com suportes 6, 12, 18, 24 e 30.


# Estrutura da Estratégia Proposta

## Métricas

- ▶ Tabela de confusão.
  - ▶ EER (Equal Error Rate).
  - ▶ Acurácia e seu respectivo desvio padrão.
- ▶ pontuação F1
- ▶ precisão
- ▶ recall
- ▶ Erro Médio Quadrático
- ▶ área sob a Curva ROC
- ▶ sensibilidade
- ▶ especificidade

 ABDULGHANI, M. M.; WALTERS, W. L.; ABED, K. H. Imagined speech classification using eeg and deep learning. *Bioengineering (Basel)*, MDPI AG, Switzerland, v. 10, n. 6, p. 649–, 2023. ISSN 2306-5354.

 AGARWAL, P.; KUMAR, S. Electroencephalography based imagined alphabets classification using spatial and time-domain features. *International journal of imaging systems and technology*, John Wiley and Sons, Inc, Hoboken, USA, v. 32, n. 1, p. 111–122, 2022. ISSN 0899-9457.

 ALI, Y. M. et al. Speech-based gender recognition using linear prediction and mel-frequency cepstral coefficients. *Indonesian J Electric Eng Comput Sci*, v. 28, n. 2, p. 753–761, 2022.



📄 ANDRÉ, F. *Caracterização de voice spoofing para fins de verificação de locutores com base na Transformada Wavelet e na Análise Paraconsistente de Características*. Dissertação (Mestrado) — Universidade Estadual Paulista - campus de São José do Rio Preto-SP, São José do Rio Preto, Brasil, 2021.


📄 BERANEK, L. L. *Acoustic Measurements*. [S.l.]: J. Wiley, 1949.

📄 BIDGOLY, A. J.; BIDGOLY, H. J.; AREZOUMAND, Z. A survey on methods and challenges in eeg based authentication. *Computers and Security*, v. 93, p. 101788, 2020. ISSN 0167-4048. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167404820300730>.


📄 ESHRAGHIAN, J. K. et al. Training spiking neural networks using lessons from deep learning. *Proceedings of the IEEE*, v. 111, n. 9, p. 1016–1054, 2023.


📄 FREITAS, S. *Avaliação acústica e áudio percetiva na caracterização da voz humana*. Faculdade de Engenharia da Universidade do Porto, 2013.

## Referências III

 GAO, Y. et al. Voice in ear: Spoofing-resistant and passphrase-independent body sound authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Association for Computing Machinery, New York, NY, USA, v. 5, n. 1, mar 2021. Disponível em: [〈https://doi.org/10.1145/3448113〉](https://doi.org/10.1145/3448113).


 GERSTNER, W. et al. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge University Press, 2014. ISBN 9781107060838. Disponível em: [〈https://neurondynamics.epfl.ch/online/Ch2.S2.html〉](https://neurondynamics.epfl.ch/online/Ch2.S2.html).

 GOODMAN, D. et al. *Spiking Neural Network Models in Neuroscience - Cosyne Tutorial 2022*. Zenodo, 2022. Disponível em: [〈https://doi.org/10.5281/zenodo.7044500〉](https://doi.org/10.5281/zenodo.7044500).

 GUIDO, R. C. Paraconsistent feature engineering [lecture notes]. *IEEE Signal Processing Magazine*, v. 36, n. 1, p. 154–158, Jan 2019.


 HAYKIN, S.; MOHER, M. *Sistemas de Comunicação-5*. [S.l.]: Bookman Editora, 2011.

## Referências IV

 HE, K. et al. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. Disponível em: [⟨http://arxiv.org/abs/1512.03385⟩](http://arxiv.org/abs/1512.03385).


 HERNÁNDEZ-DEL-TORO, T.; REYES-GARCÍA, C. A.; PINEDA, L. Villaseñor. Toward asynchronous eeg-based bci: Detecting imagined words segments in continuous eeg signals. *arXiv.org*, Cornell University Library, arXiv.org, Ithaca, 2021. ISSN 2331-8422.


 JIANG, P. et al. Securing liveness detection for voice authentication via pop noises. *IEEE Transactions on Dependable and Secure Computing*, v. 20, n. 2, p. 1702–1718, 2023.

 JONES, I. S.; KORDING, K. P. *Can Single Neurons Solve MNIST? The Computational Power of Biological Dendritic Trees*. 2020. Disponível em: [⟨https://arxiv.org/abs/2009.01269⟩](https://arxiv.org/abs/2009.01269).


 KREMER, R. L.; GOMES, M. d. C. A eficiência do disfarce em vozes femininas: uma análise da frequência fundamental. *ReVEL*, v. 12, p. 23, 2014.

## Referências V

 MAHAPATRA, N. C.; BHUYAN, P. Multiclass classification of imagined speech vowels and words of electroencephalography signals using deep learning. *Advances in human-computer interaction*, Hindawi, New York, v. 2022, p. 1–10, 2022. ISSN 1687-5893.

 MAHAPATRA, N. C.; BHUYAN, P. Eeg-based classification of imagined digits using a recurrent neural network. *Journal of neural engineering*, IOP Publishing, England, v. 20, n. 2, p. 26040–, 2023. ISSN 1741-2560.

 MENG, Z.; ALTAF, M. U. B.; JUANG, B.-H. F. Active voice authentication. *DIGITAL SIGNAL PROCESSING*, v. 101, JUN 2020. ISSN 1051-2004.


 MOCTEZUMA, L. A. et al. Subjects identification using eeg-recorded imagined speech. *Expert Systems with Applications*, v. 118, p. 201–208, 2019. ISSN 0957-4174. Disponível em: [〈https://www.sciencedirect.com/science/article/pii/S0957417418306468〉](https://www.sciencedirect.com/science/article/pii/S0957417418306468).


 PANACHAKEL, J.; RAMAKRISHNAN, A. Decoding imagined speech from eeg using transfer learning. *IEEE Access*, PP, 10 2021.

## Referências VI

 PANACHAKEL, J. T.; RAMAKRISHNAN, A.; ANANTHAPADMANABHA, T. Decoding imagined speech using wavelet features and deep neural networks. In: *2019 IEEE 16th India Council International Conference (INDICON)*. IEEE, 2019. Disponível em: <http://dx.doi.org/10.1109/INDICON47234.2019.9028925>.

 PANACHAKEL, J. T.; RAMAKRISHNAN, A. G.; ANANTHAPADMANABHA, T. V. *A Novel Deep Learning Architecture for Decoding Imagined Speech from EEG*. 2020.

 PARK, H.-j.; LEE, B. Multiclass classification of imagined speech eeg using noise-assisted multivariate empirical mode decomposition and multireceptive field convolutional neural network. *Frontiers in human neuroscience*, Frontiers Media S.A, v. 17, p. 1186594–1186594, 2023. ISSN 1662-5161.

 PINTO, F. C. G. *Manual de Iniciação em Neurocirurgia*. 2. ed. São Paulo: Santos, 2012. 384 p. ISBN 978-85-7288-979-7.

 POLIKAR, R. et al. *The wavelet tutorial*. 1996.

## Referências VII

📄 SEN, O. et al. Machine-learning methods for speech and handwriting detection using neural signals: A review. *Sensors*, v. 23, n. 12, 2023. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/23/12/5575>.


📄 SHAH, U. et al. The role of artificial intelligence in decoding speech from eeg signals: A scoping review. *Sensors (Basel, Switzerland)*, MDPI AG, Basel, v. 22, n. 18, p. 6975–, 2022. ISSN 1424-8220.


📄 TAMM, M.-O.; MUHAMMAD, Y.; MUHAMMAD, N. Classification of vowels from imagined speech with convolutional neural networks. *Computers*, MDPI, v. 9, n. 2, p. 46, 2020.


📄 VALENÇA, E. H. O. et al. Análise acústica dos formantes em indivíduos com deficiência isolada do hormônio do crescimento. Universidade Federal de Sergipe, 2014.


📄 VANDERAH, T. W. *Nolte's The Human Brain in Photographs and Diagrams*. 5th. ed. [S.l.: s.n.], 2020.


## Referências VIII

 VASKOVIĆ ALEXANDRA OSIKA, N. M. J. *Neuroanatomia*. 2024. Disponível em: <https://www.kenhub.com/pt/library/anatomia/neuroanatomia>).

 VICENTE, E. *Sistema 10-20 para localizar alvos terapêuticos em EMT*. 2023. Disponível em: <https://www.kandel.com.br/post/como-localizar-os-pontos-de-estimulacao-para-estimulacao-magnetica-transcraniana>).

 WANG, S. et al. Voiceprint recognition under cross-scenario conditions using perceptual wavelet packet entropy-guided efficient-channel-attention-res2net-time-delay-neural-network model. *Mathematics*, v. 11, n. 19, 2023. ISSN 2227-7390. Disponível em: <https://www.mdpi.com/2227-7390/11/19/4205>).

 WERTZNER, H. F.; SCHREIBER, S.; AMARO, L. Análise da frequência fundamental, jitter, shimmer e intensidade vocal em crianças com transtorno fonológico. *Revista Brasileira de Otorrinolaringologia*, SciELO Brasil, v. 71, p. 582–588, 2005.

 ZWICKER, E. Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen). *The Journal of the Acoustical Society of America*, v. 33, n. 2, p. 248–248, 02 1961. ISSN 0001-4966. Disponível em: <https://doi.org/10.1121/1.1908630>.