

Monografia apresentada para cumprimento da disciplina de estudos especiais II

André Furlan - ensismoebius@gmail.com

Universidade Estadual Paulista Júlio de Mesquita Filho

2023

Estrutura da apresentação

- Introdução.
- Revisão de conceitos utilizados.
- Trabalhos correlatos.
- Cronograma.

Introdução

Autenticação Biométrica de Locutores com Distúrbios Vocais

- A demanda por mecanismos de autenticação biométrica de locutores (ABLS) é evidente, conforme evidenciado na dissertação [Fur21].
- Estudos importantes, como [BB11] e [NP12], além de publicações em revistas de prestígio, como [HH15], [WMWX22] e [LSLR20], comprovam essa demanda.

Duas Linhas de Pesquisa Relevantes

- Autenticação Biométrica de Locutores (ABLS):
 - Essa linha de pesquisa tem sido amplamente explorada e conta com obras científicas importantes, como [Fur21], [BB11] e [NP12].
 - Métodos e técnicas têm sido propostos para aprimorar a autenticação de locutores por meio do processamento de sinais de voz, como as redes neurais profundas exemplificadas por [MS21] e [MPP21].
- Análise de Sinais de Voz para Pré-diagnóstico e Classificação de Distúrbios Vocais:
 - Essa linha de pesquisa se concentra em estratégias acústico-computacionais para identificar e classificar alterações laríngeas e outras irregularidades na fonação, como descrito em [CSBY22] e [FKO⁺22].
 - Metodologias avançadas, como redes neurais profundas [GBC16], e a lógica paraconsistente, como descrito em [FPM⁺17], têm sido utilizadas nessa área.

Desafio: Autenticar Locutores com Distúrbios Vocais

- O desafio é autenticar locutores afetados por distúrbios vocais que alteram a impressão acústica da fala, como mencionado em [LHA05a, LHA05b].
- Distúrbios vocais podem ter origem orgânica, funcional ou orgânico-funcional, e afetam a qualidade e características da voz.
- O projeto de pesquisa visa investigar e desenvolver algoritmos biométricos para autenticar locutores com fala severamente degradada devido a esses distúrbios, conforme discutido em [GPP⁺21].
- Uma abordagem promissora é combinar as informações da voz com os sinais cerebrais durante a fonação, ou seja, provenientes da fala imaginada [MTGVPC19].

- Ao pesquisar o tema nos registros do Web of Science e outras bases de dados, observa-se uma escassez de artigos científicos sobre autenticação de locutores com distúrbios vocais.
- A maioria dos artigos encontrados são propostas genéricas que envolvem metodologias para viabilizar interfaces cérebro-computador (BCIs), conforme descrito em [RG21] e [BK10].
- Por outro lado, os artigos mais relevantes e dedicados à biometria são encontrados nas referências [MTGVPC19, MM18, JCA16, JCA17, DPBATRT14, RBJL16].

Objetivos e metas

O objetivo geral deste trabalho é projetar e implementar algoritmos biométricos capazes de autenticar, por meio da fala, indivíduos que produzem apenas locuções potencialmente degradadas, adicionando informações dos sinais cerebrais durante a fonação (imagined speech).

- 1 Realizar um levantamento bibliográfico sobre ABLs com disfonias laríngeas severas (DLSs) e estratégias BCI para imagined speech [CGR17].
- 2 Estudar bases de dados públicas de imagined speech para experimentos iniciais.
- 3 Criar uma base de dados contendo sinais de voz de indivíduos com DLS e os respectivos sinais de imagined speech.

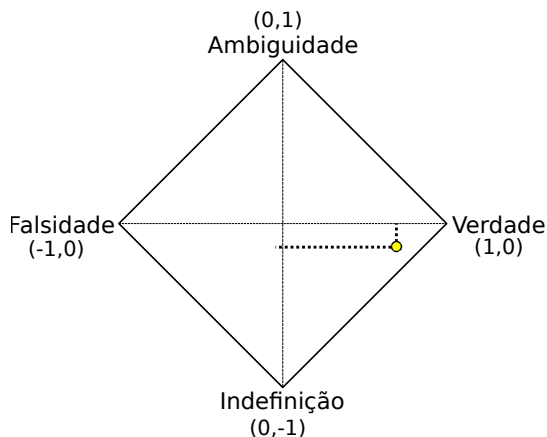
Objetivos Específicos (continuação)

- 4 Extrair características representativas dos sinais de voz e imagined speech de cada locutor utilizando técnicas de handcrafted extraction e feature learning.
- 5 Autenticar os locutores na base de dados usando estruturas de aprendizagem profunda como Redes Neurais Residuais (RNNs) e Deep Spiking Neural Networks (DSNNs).
- 6 Implementar os algoritmos em C/C++ para funcionamento off-line e, se possível, em tempo real.
- 7 Disseminar os resultados por meio de apresentações em congressos e publicações em periódicos de alto impacto.

Revisão de conceitos utilizados

Engenharia paraconsistente de características

Os vetores de características proporcionam uma boa separação interclasses?



- Verdade:
 $\alpha = 1$ e $\beta = 0$.
- Ambigüidade:
 $\alpha = 1$ e $\beta = 1$.
- Falsidade:
 $\alpha = 0$ e $\beta = 1$.
- Indefinição:
 $\alpha = 0$ e $\beta = 0$.

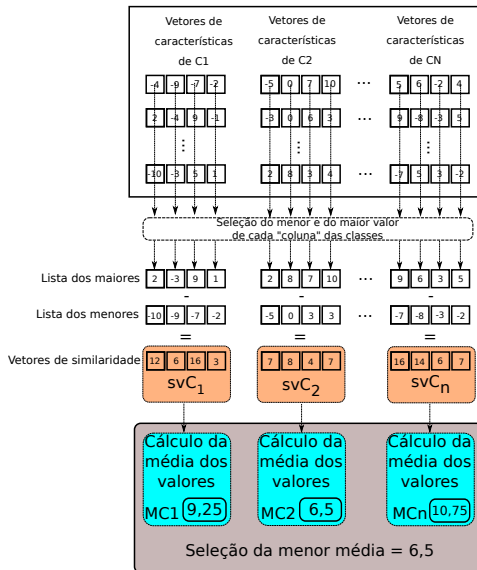
Figura: Plano paraconsistente



unesp

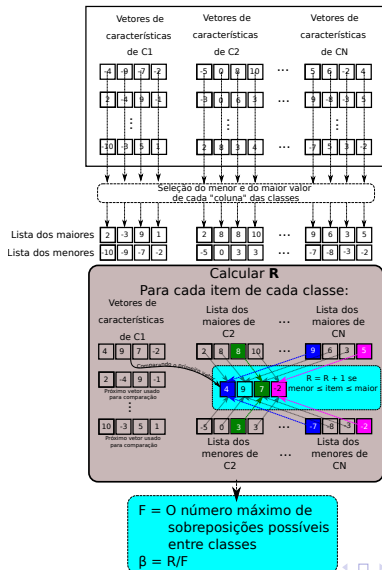
Engenharia paraconsistente de características

Cálculo de α



Engenharia paraconsistente de características

Cálculo de β



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 $P = (G_1, G_2)$

- $(-1,0) \rightarrow$ Falsidade;
- $(1,0) \rightarrow$ Verdade;
- $(0,-1) \rightarrow$ Indefinição;
- $(0,1) \rightarrow$ Ambiguidade.

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:

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

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

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

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

- Tamanho de janelas variável.
- Análise multiresolução.
- Análise detalhada em altas e baixas frequências.

- Apenas transformadas diretas.
- Não haverá reconstrução do sinal.
- Construção de vetores de características.
- Domínio discreto.

Filtros digitais *wavelet*

Wavelets de Daubechies

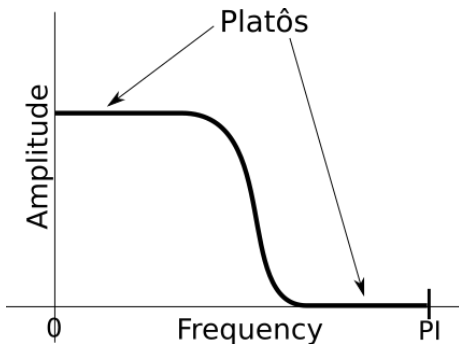


Figura: platôs maximamente planos em um filtro digital

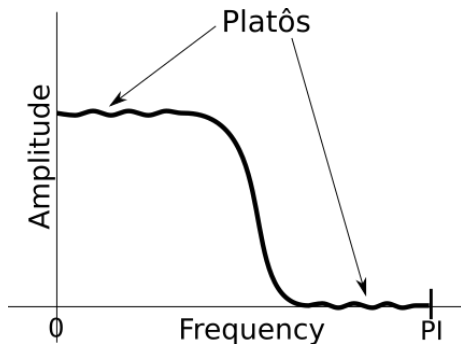


Figura: platôs não maximamente planos de um filtro digital

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: $s[\cdot] = [1, 2, 3, 4]$.

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

Caracterização dos processos de produção da voz humana

Áreas de estudo

- Fisiológica ou fonética articulatória.
- Acústica ou fonética acústica.
- Perceptual.

Foco apenas na acústica

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 .
- Componente periódico resultante da vibração das pregas vocais.

Caracterização dos processos de produção da voz humana

Formantes

- $F_1 \rightarrow$ 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.

- Autoencoders são redes neurais treinadas para reconstruir seus dados de entrada.
- Consistem em uma função codificadora $h = f(x)$ e uma função decodificadora $r = g(h)$.
- A camada oculta h representa uma representação comprimida dos dados de entrada.
- O objetivo é aprender uma representação compactada dos dados e reconstruir os dados de entrada com a maior precisão possível.
- Os autoencoders são limitados para apenas aproximar a entrada e priorizar certos aspectos dos dados.
- Podem ser treinados usando técnicas como gradient descent com minibatch ou estocástico.

- As ResNets incluem conexões de salto ou mapeamentos de identidade.
- Essas conexões permitem que a saída de uma camada seja adicionada diretamente à entrada da camada subsequente.
- Permitem que redes mais profundas possam aprender e evitam o problema de vanishing gradients.
- As conexões de salto criam "highway connections" que contornam as camadas intermediárias.
- Melhoram o desempenho e a capacidade de aprendizado das redes.

- As SNNs são redes neurais inspiradas na maneira como os neurônios se comunicam no cérebro.
- Utilizam picos ou potenciais de ação para representar informações.
- As informações são representadas pelo tempo e pela taxa de picos, não pela força das conexões entre os neurônios.
- O tempo dos picos e a dinâmica temporal da rede são cruciais no processamento da informação.
- Cada pico carrega informações de tempo, permitindo uma codificação mais precisa de informações.
- As SNNs são promissoras para a classificação de sinais de voz e EEG devido à natureza temporal desses sinais.

Trabalhos correlatos

Avaliação de um sistema ASR para falantes de língua inglesa com deficiências na fala [GMJ⁺21]

- Fonte dos dados ou participantes: 432 falantes da língua inglesa de variadas etnias e deficiências na fala.
- Informações ou características extraídas: Energias calculadas a partir de um banco de filtros de 80 dimensões.
- Classificadores usados: Rede extratora de características com 8 camadas e rede classificadora LSTM com 2 camadas.
- Resultados: Precisão média de reconhecimento de 9% a 80%, superando transcritores humanos. Taxa de erro de palavras (WER) média de 4,6%.

Aumento de dados para reconhecimento de fala disfônica utilizando redes adversariais generativas de convolução profunda (DCGAN) [JGX⁺21]

- Fonte dos dados ou participantes: UASpeech.
- Informações ou características extraídas: *GAN generator*.
- Classificadores usados: *GAN discriminador*.
- Resultados: Redução da taxa de erro de palavra (WER) de até 3,05% em comparação com sistemas sem aumento de dados.

Detecção de segmentos de palavras imaginadas em sinais de EEG contínuos [HDTRGVP21]

- Fonte de dados: Próprias: duas com 27 pessoas e 1 com 20.
- Características: Transformada Discreta de Wavelet (DWT), Decomposição Empírica de Modos (EMD), características de energia, dimensões fractais e medidas de caos.
- Classificadores: Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) e Logistic Regression (LR).
- Desempenho: Pontuação média de F1 score de 0,75.

Combinação de áudio e EEG para aprimorar o reconhecimento de fala em sistemas de Interação Humano-Máquina (HMI) [MTG21]

- Fonte de dados: Kara One [ZR15]
- Características: Múltiplas modalidades (áudio e EEG) e técnicas de Transformada Wavelet (WT).
- Resultados: Taxas de precisão de até 74,48% no reconhecimento de fala.

Uso de técnicas de inteligência artificial (IA) para decodificar a fala a partir de sinais cerebrais humanos [SAM⁺22]

- Modalidade de Dados e Técnicas de IA: Dados de EEG e estímulos de palavras/frases. Técnicas de IA incluem aprendizado de máquina e aprendizado profundo.
- Extração de Características e Processamento de Sinais: Técnicas de normalização e extração de características foram aplicadas devido ao ruído nos sinais de EEG. Filtragem de banda passante e técnicas de wavelet foram comumente utilizadas.
- Conjunto de Dados e Equipamentos de Gravação: Dispositivos de EEG com 64 canais foram amplamente utilizados, com alguns estudos usando 32 canais ou menos.

Sistema de verificação automática de falas (ASV) para pacientes com disartria [SSA23]

- Recursos Utilizados: Pitch, volume e probabilidade de vocalização. Aumento de dados fora do domínio.
- Conjunto de Dados: Dyarthric Speech Database (DSD) e SpeechDat-Car.
- Extração de Características: Vetores de características i-vector e x-vector usando MFCC, variáveis prosódicas e suas combinações.
- Resultados: EER de 11,09 para disartria leve, 13,26 na média e 11,97 para disartria grave.

Uso de redes neurais profundas (DNNs) como criadores de modelos acústicos [HDY⁺12]

- Método: DNNs para geração de vetores de características, posteriormente classificados usando Hidden Markov Models (HMM).
- Base de Dados: TIMIT, com mais de 630 falantes da língua inglesa.
- Resultados: WER de 18,5% usando a combinação DNN + HMM.

Uso de modelo de Rede Neural Convolucional (CNN) para classificação de fala [AFG⁺20]

- Base de Dados: BREF (fala francesa) e C2SI-LEC (pacientes com falas disfuncionais).
- Características Extraídas: Coeficientes Cepstrais de Frequência Mel (MFCCs) e suas derivadas.
- Pré-processamento: Normalização dos dados de entrada.
- Técnica de Treinamento: Subamostragem aleatória e taxa de aprendizado decaído exponencialmente.
- Resultados: Acurácia de 0,68 na base BREF e 0,71 na base C2SI-LEC, superando ouvintes humanos.

Uso de Rede Neural Recorrente de Memória de Curto Prazo (LSTM-RNN) para decodificação de sinais de EEG [PD22]

- Dados de EEG: 8 canais de medição dos sinais.
- Extração de Características: Transformadas Wavelets nos canais.
- Comandos de Áudio: Para cima, para baixo, para a esquerda e para a direita.
- Resultados: Acurácia de classificação geral de 92,50% e métricas adicionais (precisão, *recall* e *F1-score*) com valores próximos.

Uso de EEG e espectroscopia de infravermelho próximo funcional (fNIRS) para coleta de dados [CFC21]

- Extração de Características: Tempo (média, variância, assimetria e curtose) e frequência (densidade espectral de potência e potência de banda).
- Classificadores Utilizados: Análise Discriminante Linear (LDA), Máquina de Vetores de Suporte (SVM) e Rede Neural Convolucional (CNN).
- Resultados: Precisão de 87,18% para fala aberta e 53% para fala imaginada, com destaque para estímulos imagéticos.

Uso de sinais de EEG para fala imaginada e detecção de interações entre regiões do cérebro [BKEM22]

- Dados de EEG: Sinais correspondentes a fala imaginada de quatro vogais.
- Extração de Características: Índice de localização (LI) baseado em conectividade funcional entre regiões do cérebro.
- Classificadores Utilizados: SVM e Análise Discriminante Linear (LDA).
- Resultados: Precisão média da classificação de 81,1%.

Estrutura baseada em aprendizado métrico profundo para decodificação de *imagined speech* usando interfaces cérebro-máquina (BCI) [LLL21]

- Bases de Dados: *Coretto* (6 classes) e *BCI Competition* (5 classes).
- Extração de Características: Frequência instantânea e entropia espectral dos sinais de EEG.
- Estrutura Proposta: Rede Neural Siamesa com aprendizagem métrica profunda.
- Resultados: Precisão de $45,00 \pm 3,13\%$ para 6 classes e $48,10 \pm 3,68\%$ para 5 classes.

Classificação de sinais de EEG correspondentes a falas imaginadas de vogais e palavras [TMM20]

- Conjunto de Dados EEG: 15 sujeitos imaginando vogais e palavras.
- Pré-processamento: Filtragem passa-faixa, subamostragem, segmentação em lotes e normalização.
- Classificador Utilizado: Rede Neural Convolucional com transferência de aprendizado.
- Resultados: Acurácia de 23,98% ($\pm 3,08$).

Uso de *Hashing* sensível a locus (LSH) para reconhecimento de locutor [AKY⁺18]

- Método Proposto: Combinação de MFCC e LSH para identificação do locutor.
- Base de Dados: TIMIT 2018.
- Procedimentos: Extração de MFCC e criação de tabela *hash* com LSH.
- Resultados: Acurácia de 92,66%.

Fechamento

Com base nos estudos mencionados nesta pesquisa, as seguintes abordagens serão consideradas:

Camada Extratora de Características e Decomposição de Sinais:

- Utilizar uma camada extratora de características semelhante à abordada por [GMJ⁺21].
- Aplicar filtros wavelet packet para a decomposição dos sinais.

Técnicas de Aumento de Dados e Protocolo de Coleta de Dados:

- Recorrer a técnicas de aumento de dados, como aquelas empregadas por [JGX⁺21], devido à escassez de locutores disfônicos.
- Adaptar o protocolo de obtenção de dados apresentado por [TMM20].
- Investigar a base de dados UASpeech mencionada por [TMM20], que pode fornecer informações valiosas para a verificação de padrões.

Utilização das Transformadas Wavelet e Normalização de Dados:

- Utilizar diretamente as Transformadas Wavelet para a criação dos vetores de características, seguindo o trabalho de [MTG21].
- Considerar técnicas de normalização de dados discutidas por [SAM⁺22].

Além disso, os seguintes aspectos interessantes, mas que não necessariamente serão utilizados, foram observados nos artigos revisados:

Modelagem Acústica e Conectividade Funcional do Cérebro:

- O artigo de [HDY⁺12] introduz o conceito de Modelagem Acústica, que pode ser relevante para o problema desta pesquisa.
- Destaca-se o uso de matrizes que representam a conectividade funcional entre diferentes regiões do cérebro durante a fala imaginada, conforme mencionado por [BKEM22].

Aprendizado Métrico Profundo:

- O estudo de [LLL21] propõe uma estrutura baseada em aprendizado métrico profundo para decodificar a fala imaginada, embora essa abordagem difira significativamente da proposta desta pesquisa.

Fechamento (Conclusão)

Em resumo, nenhum dos artigos analisados utiliza diretamente as técnicas propostas neste estudo, como autoencoders, redes neurais residuais, spike neural networks e vetores de características "hand-crafted". Além disso, nenhum deles tem como objetivo comparar esses métodos, o que torna esta tese potencialmente relevante para a comunidade acadêmica.

Cronograma para conclusão do doutorado

Até a presente data foram realizados os levantamentos bibliográficos e o cumprimento dos créditos necessários em disciplinas para o Doutorado.

Fases do Projeto:

- **E1:** Levantamento bibliográfico sobre aquisição e análise de *imagined speech* por meio de EEG e reconhecimento de voz em pacientes com disfonia.
- **E2:** Diálogos com profissionais de fonoaudiologia e neurologia, estudo de base de dados pública e criação de uma base de dados própria.
- **E3:** Pré-processamento e extração de características representativas dos sinais de voz e *imagined speeches*.
- **E4:** Integração das etapas de extração de características e autenticação utilizando estruturas de aprendizado profundo.
- **E5:** Apresentação dos experimentos em eventos científicos, publicação dos resultados e elaboração da monografia de qualificação e tese de doutorado.

Cronograma para conclusão do doutorado

Mês/Ano	E1	E2	E3	E4	E5
09/22	X				
10/22	X				
11/22	X				
12/22	X				
01/23	X	X			
02/23	X	X			
03/23	X	X			
04/23	X	X			
05/23	X	X			X
06/23	X	X			X
07/23		X	X		X
08/23		X	X		X
09/23		X	X		
10/23		X	X		
11/23		X	X		
12/23		X	X		

Mês/Ano	E1	E2	E3	E4	E5
01/24		X	X	X	X
02/24		X	X	X	X
03/24			X	X	
04/24			X	X	
05/24			X	X	
06/24			X	X	
07/24			X	X	
08/24			X	X	
09/24			X	X	
10/24			X	X	
11/24			X	X	
12/24			X	X	
01/25			X	X	
02/25			X	X	
03/25			X	X	
04/25			X	X	

Mês/Ano	E1	E2	E3	E4	E5
05/25			X	X	
06/25			X	X	
07/25			X	X	
08/25			X	X	
09/25			X	X	
10/25			X	X	
11/25			X	X	
12/25			X	X	
01/26			X	X	
02/26			X	X	
03/26			X	X	
04/26			X	X	
05/26					X
06/26					X
07/26					X
08/26					X



unesp



Sondes Abderrazek, Corinne Fredouille, Alain Ghio, Muriel Lalain, Christine Meunier, and Virginie Woisard.

Towards Interpreting Deep Learning Models to Understand Loss of Speech Intelligibility in Speech Disorders — Step 1: CNN Model-Based Phone Classification.

In *Proc. Interspeech 2020*, pages 2522–2526, 2020.



A. Awais, S. Kun, Y. Yu, S. Hayat, A. Ahmed, and T. Tu.

Speaker recognition using mel frequency cepstral coefficient and locality sensitive hashing.

In *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pages 271–276, May 2018.



Homayoon Beigi and Homayoon Beigi.

Speaker recognition.

Springer, 2011.



Katharine Brigham and BVK Vijaya Kumar.

Imagined speech classification with eeg signals for silent communication: a preliminary investigation into synthetic telepathy. In *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, pages 1–4. IEEE, 2010.



Mohamad Amin Bakhshali, Morteza Khademi, and Abbas Ebrahimi-Moghadam.

Investigating the neural correlates of imagined speech: An eeg-based connectivity analysis.

Digital Signal Processing, 123:103435, 2022.



Ciaran Cooney, Raffaella Folli, and Damien Coyle.

A bimodal deep learning architecture for eeg-fnirs decoding of overt and imagined speech.

IEEE Transactions on Biomedical Engineering, 69(6):1983–1994, 2022.



Germán A. Pressel Coretto, Iván E. Gareis, and H. Leonardo Rufiner. Open access database of eeg signals recorded during imagined speech. 2017.



Mounira Chaiani, Sid Ahmed Selouani, Malika Boudraa, and Mohammed Sidi Yakoub.

Voice disorder classification using speech enhancement and deep learning models.

Biocybernetics and Biomedical Engineering, 42(2):463–480, 2022.



Marcos Del Pozo-Banos, Jesús B Alonso, Jaime R Ticay-Rivas, and Carlos M Travieso.

Electroencephalogram subject identification: A review.

Expert Systems with Applications, 41(15):6537–6554, 2014.



Shintaro Fujimura, Tsuyoshi Kojima, Yusuke Okanoue, Kazuhiko Shoji, Masato Inoue, Koichi Omori, and Ryusuke Hori.

Classification of voice disorders using a one-dimensional convolutional neural network.

Journal of Voice, 36(1):15–20, 2022.



Everthon Silva Fonseca, Denis César Mosconi Pereira, Luís Fernando Castilho Maschi, Rodrigo Capobianco Guido, and Katia Cristina Silva Paulo.

Linear prediction and discrete wavelet transform to identify pathology in voice signals.

In *2017 Signal Processing Symposium (SPSymposium)*, pages 1–4. IEEE, 2017.



André Furlan.

Caracterização de voice spoofing para fins de verificação de locutores com base na transformada wavelet e na análise paraconsistente de características.

Dissertação de mestrado, Universidade Estadual Paulista - campus de São José do Rio Preto-SP, São José do Rio Preto, Brazil, 2021.

Orientador: Prof Dr Rodrigo Capobianco Guido.



Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

Deep learning.

MIT press, 2016.



Jordan R. Green, Robert L. MacDonald, Pan-Pan Jiang, Julie Cattiau, Rus Heywood, Richard Cave, Katie Seaver, Marilyn A. Ladewig, Jimmy Tobin, Michael P. Brenner, Philip C. Nelson, and Katrin Tomanek.

Automatic speech recognition of disordered speech: Personalized models outperforming human listeners on short phrases.

In *INTERSPEECH 2021*, Interspeech, pages 4778–4782, 2021.

Interspeech Conference, Brno, CZECH REPUBLIC, AUG 30-SEP 03, 2021.



Siddhant Gupta, Ankur T Patil, Mirali Purohit, Mihir Parmar, Maitreya Patel, Hemant A Patil, and Rodrigo Capobianco Guido.

Residual neural network precisely quantifies dysarthria severity-level based on short-duration speech segments.

Neural Networks, 139:105–117, 2021.



Tonatiuh Hernandez-Del-Toro, Carlos A. Reyes-Garcia, and Luis Villasenor-Pineda.

Toward asynchronous eeg-based bci: Detecting imagined words segments in continuous eeg signals.

BIOMEDICAL SIGNAL PROCESSING AND CONTROL, 65, MAR 2021.



Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury.

Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups.

IEEE Signal Processing Magazine, 29(6):82–97, 2012.



John HL Hansen and Taufiq Hasan.

Speaker recognition by machines and humans: A tutorial review.
IEEE Signal processing magazine, 32(6):74–99, 2015.



Isuru Jayarathne, Michael Cohen, and Senaka Amarakeerthi.

Brainid: Development of an eeg-based biometric authentication system.

In *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 1–6. IEEE, 2016.



Isuru Jayarathne, Michael Cohen, and Senaka Amarakeerthi.

Survey of eeg-based biometric authentication.

In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)*, pages 324–329. IEEE, 2017.



Zengrui Jin, Mengzhe Geng, Xurong Xie, Jianwei Yu, Shansong Liu, Xunying Liu, and Helen Meng.

Adversarial Data Augmentation for Disordered Speech Recognition.
In *Proc. Interspeech 2021*, pages 4803–4807, 2021.



Francois Le Huche and André Allali.

A voz: patologia vocal de origem funcional.
In *A voz: patologia vocal de origem funcional*, pages 187–187. 2005.



François Le Huche and André Allali.

A voz: patologia vocal de origem orgânica.
In *A voz: patologia vocal de origem orgânica*, pages 154–154. 2005.





Dong-Yeon Lee, Minji Lee, and Seong-Whan Lee.

Decoding imagined speech based on deep metric learning for intuitive bci communication.

IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29:1363–1374, 2021.



Kong Aik Lee, Seyed Omid Sadjadi, Haizhou Li, and Douglas A Reynolds.

Two decades into speaker recognition evaluation-are we there yet?

Comput. Speech Lang., 61:101058, 2020.



Luis Alfredo Moctezuma and Marta Molinas.

Eeg-based subjects identification based on biometrics of imagined speech using emd.

In *Brain Informatics: International Conference, BI 2018, Arlington, TX, USA, December 7–9, 2018, Proceedings 11*, pages 458–467. Springer, 2018.



Ioanna Miliaresi, Kyriakos Poutos, and Aggelos Pikrakis.

Combining acoustic features and medical data in deep learning networks for voice pathology classification.

In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 1190–1194. IEEE, 2021.



unesp



Vikas Mittal and RK Sharma.

Deep learning approach for voice pathology detection and classification.

International Journal of Healthcare Information Systems and Informatics (IJHISI), 16(4):1–30, 2021.



P. P. Mini, Tessamma Thomas, and R. Gopikakumari.

Wavelet feature selection of audio and imagined/vocalized eeg signals for ann based multimodal asr system.

BIOMEDICAL SIGNAL PROCESSING AND CONTROL, 63, JAN 2021.



Luis Alfredo Moctezuma, Alejandro A Torres-García, Luis Villaseñor-Pineda, and Maya Carrillo.

Subjects identification using eeg-recorded imagined speech.

Expert Systems with Applications, 118:201–208, 2019.



Amy Neustein and Hemant A Patil.
Forensic speaker recognition, volume 1.
Springer, 2012.



Dipti Pawar and Sudhir Dhage.
Wavelet-based imagined speech classification using
electroencephalography.
International Journal of Biomedical Engineering and Technology,
38(3):215–224, 2022.



Maria V Ruiz-Blondet, Zhanpeng Jin, and Sarah Laszlo.
Cerebre: A novel method for very high accuracy event-related potential
biometric identification.
IEEE Transactions on Information Forensics and Security,
11(7):1618–1629, 2016.



Ana-Luiza Rusnac and Ovidiu Grigore.

Eeg preprocessing methods for bci imagined speech signals.

In *2021 International Conference on e-Health and Bioengineering (EHB)*, pages 1–4. IEEE, 2021.



Uzair Shah, Mahmood Alzubaidi, Farida Mohsen, Alaa Abd-Alrazaq, Tanvir Alam, and Mowafa Househ.

The role of artificial intelligence in decoding speech from eeg signals: A scoping review.

SENSORS, 22(18), SEP 2022.



Shinimol Salim, Syed Shahnawazuddin, and Waquar Ahmad.

Automatic speaker verification system for dysarthric speakers using prosodic features and out-of-domain data augmentation.

Applied Acoustics, 210:109412, 2023.



Markus-Oliver Tamm, Yar Muhammad, and Naveed Muhammad.
Classification of vowels from imagined speech with convolutional neural networks.

Computers, 9(2):46, 2020.



Xingmei Wang, Jiaxiang Meng, Bin Wen, and Fuzhao Xue.
Racp: A network with attention corrected prototype for few-shot speaker recognition using indefinite distance metric.

Neurocomputing, 490:283–294, 2022.



Shunan Zhao and Frank Rudzicz.
Classifying phonological categories in imagined and articulated speech.
In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 992–996. IEEE, 2015.