

## RECONHECIMENTO INTELIGENTE DE LOCUTORES NA MODALIDADE INDEPENDENTE DO DISCURSO

MATHEUS YOSHIMITSU TAMASHIRO PIRES LANZO, IBILCE, São José do Rio Preto,  
yoshimitsu.lanzo@unesp.br

Apresentado no XXXV Congresso de Iniciação Científica da Unesp – CIC 2023  
“Desafios na produção do conhecimento: democratização e diversidade”

**INTRODUÇÃO:** As técnicas de identificação de indivíduos por voz têm se tornado cada vez mais frequentes, principalmente em aplicações de controle de acesso, substituindo tradicionais sistemas baseados em senhas. O potencial exponencialmente crescente dos computadores constitui um forte fator motivador para que os sistemas de reconhecimento de padrões, inclusive aqueles baseados em voz, estejam presentes em aplicações diversas. O objetivo desta pesquisa é de em meio a discursos variados, identificar diferentes locutores utilizando de algoritmos de análise de sinais, de forma a possibilitar a comparação de características temporais e espectrais e de classificadores diversos, analisando os resultados por meio de validações cruzadas e matrizes de confusão.

**MATERIAL E MÉTODOS:** 10 sinais de voz cedidos por diferentes locutores.

São extraídas as características, ou seja, valores numéricos representativos dos sinais-modelo de voz, visando caracterizar os locutores-alvo independentemente do discurso. Para tal, são delimitadas regiões que estejam na fronteira entre segmentos vozeados e não-vozeados, das quais serão computadas a frequência fundamental ( $F_0$ ) e as frequências formantes ( $F_1$ ,  $F_2$ ,  $F_3$  e  $F_4$ ) utilizando de algoritmos que implementem Análise Cepstral, Taxa de Cruzamento por Zeros e Energia. Uma vez que as melhores características sejam escolhidas, são definidas e testadas configurações distintas de classificadores *knowledge-based* do tipo *deep neural network* (DNN). Método *Support Vector Machine* (SVM): é um modelo mais elaborado, sendo do tipo *knowledge-based*. É um algoritmo de aprendizado de máquina.

Método da Distância Euclidiana: é um critério modesto do tipo *pattern matching*, sem maiores possibilidades de incorporar aprendizado estatístico.

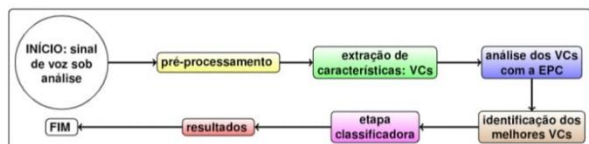


Figura 1. Resumo da Metodologia

**RESULTADOS E DISCUSSÃO:** Tendo sido consideradas 10 sinais de vozes totais, de diferentes locutores e discursos, foram analisadas as acurácias dos dois métodos em identificar os diferentes locutores.

Tabela 1. Resultados do Método Distância Euclidiana

Distância Euclidiana (Treino vs Teste)	Precisão (Acurácia)
10 vs 1	70,0%
9 vs 2	65,0%
8 vs 3	63,3%
7 vs 4	55,0%

Tabela 2. Resultados do Método SVM

SVM (Treino vs Teste)	Precisão (Acurácia)
10 vs 1	85,0%
9 vs 2	80,3%
8 vs 3	73,3%
7 vs 4	65,0%

Ao comparar os dois métodos, é possível perceber também as divergências de resultados, fazendo com que hajam também diferentes aplicações para ambos, como por exemplo, um cenário industrial, menos preciso, para o primeiro método e um cenário acadêmico, mais preciso, para o segundo. Embora a precisão mais alta não esteja próxima de 100%, pode-se considerar que melhorias nos algoritmos de escolhas de características, mas principalmente um maior número de amostras impactaria positivamente a performance dos métodos, sendo o *Support Vector Machine*, aquele mais beneficiado.

**CONCLUSÕES:** O estudo realizado evidencia a importância do reconhecimento inteligente de locutores, na modalidade *text-independent*, ao comprovar que mesmo sem um alto número de sinais de voz disponíveis para obter um resultado ótimo em *machine learning*, é possível extrair altas porcentagens de precisão na identificação de locutores, mostrando então uma área de pesquisa em expansão e que demonstra grande utilidade em diversos cenários, percorrendo desde o uso em assistentes de voz, até mesmo em níveis de segurança como senhas.

**AGRADECIMENTOS:** Agradeço ao Conselho Nacional de Pesquisa e Desenvolvimento (CNPq) pelo incentivo, à Unesp pela oportunidade e ao meu orientador, Rodrigo Capobianco Guido, por guiar o projeto e se mostrar sempre disposto a sanar, com sabedoria, as dúvidas advindas durante o percurso.

### REFERÊNCIAS:

- DENG, L.; O'SHAUGHNESSY, D. Speech processing: a dynamic and optimization-oriented approach. [s.l.] New York Marcel Dekker, 2003.
- GUIDO, R.C. ZCR-aided neurocomputing: a study with applications. Knowledge-based Systems, v. 105, pp.248-269, 2016.
- RABINER, L. R.; SCHAFER, R. W. Theory and applications of digital speech processing. Upper Saddle River: Pearson, 2011.