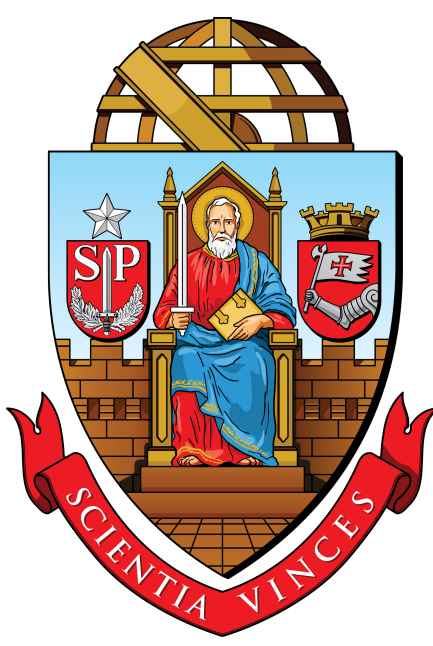




Supressão de ruído acústico no sinal de áudio da fala durante exames de ressonância magnética

Aluno: Vinícius Pessoa Duarte
Supervisor: Prof^o Dr. Marcel Parolin Jackowski
IME-USP, São Paulo - 2018
vinipd@ime.usp.br



Introdução

- Crianças com transtornos residuais ou persistentes da fala (TRPF) acabam por ter dificuldades, tanto para comunicação social, quanto para sua alfabetização.
- Alterações do trato vocal podem observadas em crianças com TRPF utilizando imageamento por ressonância magnética (RM) durante exercícios vocais.
- No entanto, o áudio adquirido durante sessões de RM é normalmente poluído por ruídos acústicos gerados pela vibração das bobinas fazem parte do equipamento de RM [3].
- Como esses ruídos dificultam a compreensão e análise da fala original, pois a intensidade destes é alta, além de terem em sua composição frequências que intersectam o conjunto de frequências que compõem a fala, acabam por impor a necessidade de supressão de ruído acústico (SRA) no áudio obtido, de modo a melhorar a compreensão da fala do paciente.
- Dado que tais ruídos possuem características periódicas, em conjuntos de frequências específicas, SRA pode ser concretizada por algoritmos que visam identificar essas características e, a partir dos padrões encontrados, extrair a fala "limpa" do paciente, i.e., a fala com diminuição do ruído acústico proveniente da máquina de RM.

Objetivos

- Estudar e implementar três diferentes métodos para a realização de SRA em áudios semelhantes aos provenientes de exames de RM, sendo eles: Subtração espectral; método dos mínimos quadrados no domínio das frequências (FLMS) e supressão de ruído baseada no aprendizado de dicionários e análise *wavelet*.
- Implementação de duas versões de um *software* para realização de supressão de ruído acústico.
- Realizar análises de tempo de execução, da diferença de energia total dos espectros dos áudios obtidos e da capacidade de supressão de acordo com a amplitude da fala em relação ao ruído.
- Apresentar subsídios para a escolha de método de SRA a partir dos resultados obtidos.

Materiais

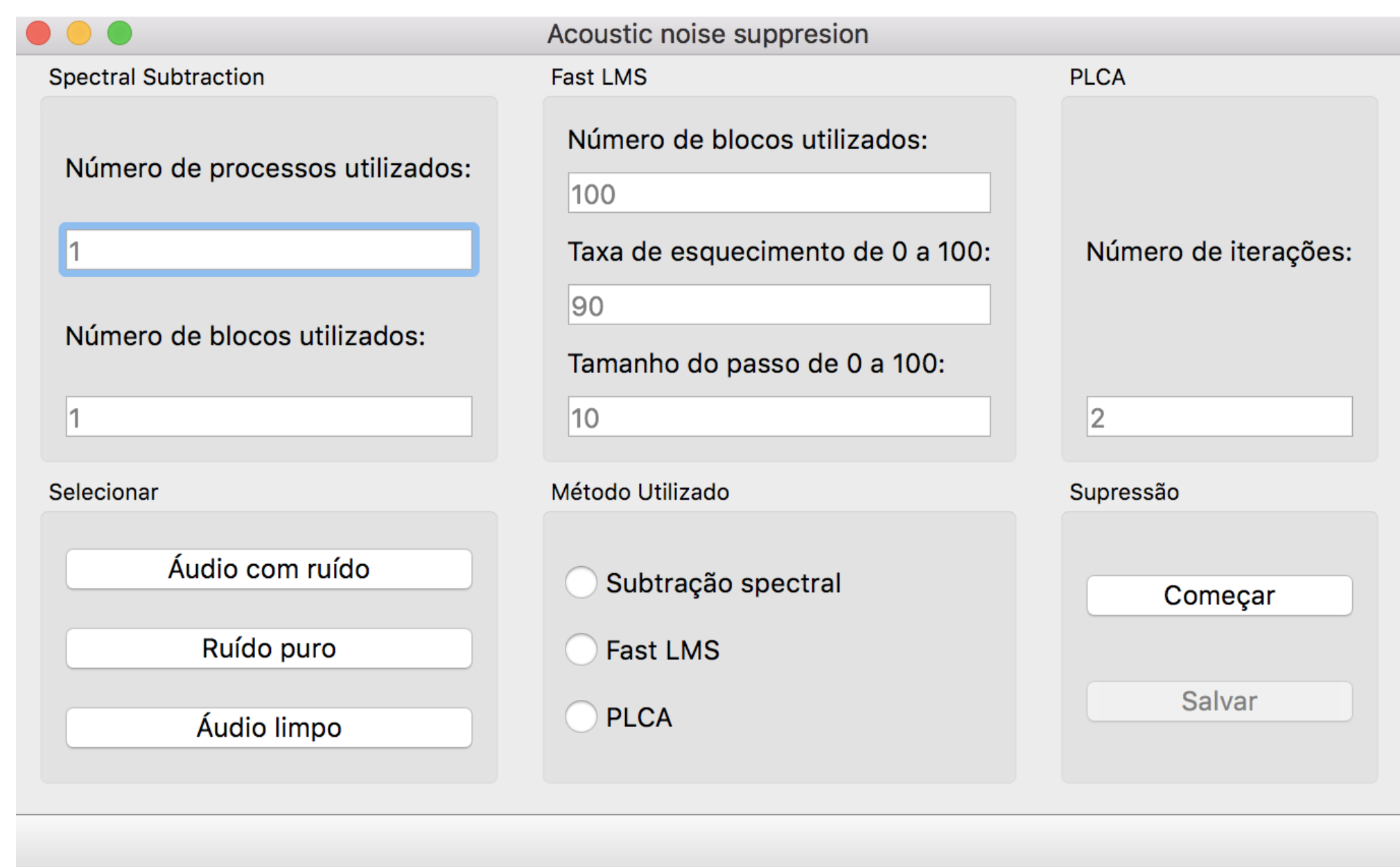


Figura 1: Software com interface gráfica

- Áudios utilizados para as análises, compostos por trecho de diálogo combinado com três diferentes tipos de ruído acústico: *Echo planar*; *diffusion* e *fast spin echo*.
- Software*, em *Python*, implementando os métodos de SRA estudados neste trabalho. Há uma versão por linha de comando e uma versão com interface gráfica. A implementação do *software*, bem como a implementação dos métodos estudados, encontra-se em <https://github.com/Vinyxr4/MAC0499>.
- Implementação dos métodos de SRA, descritos a seguir:

Subtração espectral: Supondo aditividade entre o ruído e o sinal de fala, i.e., no domínio das frequências, a fala e o ruído podem ser separados em componentes distintas que, quando somadas, resultam no espectrograma do áudio com ruído original [3], o método consiste na subtração entre a energia do ruído acústico e a energia do sinal que contém fala e ruído.

FLMS: Esse método consiste no treinamento de um filtro adaptativo com base em pesos de ativação, os quais são atualizados iterativamente de acordo com o áudio desejado e a estimativa do áudio suprimido calculado na etapa anterior, uma função de passo, uma amostra do áudio corrompido original e o próprio valor dos pesos durante a iteração anterior [1].

Aprendizado de dicionários e análise *wavelet*: Esse método se baseia em um algoritmo de separação de componentes e limiarização [4]. O método possui duas fases: a aplicação do algoritmo *probabilistic latent component analysis* (PLCA) e realização de análise *wavelet*. Na primeira fase, um dicionário de ruído é descoberto, o qual é utilizado para encontrar uma aproximação do sinal de fala sem ruído original. Já na segunda fase, o sinal obtido é limiarizado de modo a aproximar ainda mais o sinal resultante do sinal de fala original.

Resultados

- Os resultados a seguir são referentes a um sinal contendo um diálogo com duração de um minuto, ruído acústico *echo planar* de mesma duração e sinal referente à SRA da combinação dos sinais anteriores para cada um dos métodos implementados.

Tempo de execução, em segundos, para realizar SRA

Tabela 1: Ruído <i>echo planar</i>			
	Subtração espectral	FLMS	Dicionário e wavelet
Média	0.543656	1.000183	6.842135
Desvio padrão	0.02294	0.022398	0.320086
Variância	0.000526	0.000502	0.102455

Energia do diálogo e do diálogo com ruído *echo planar*

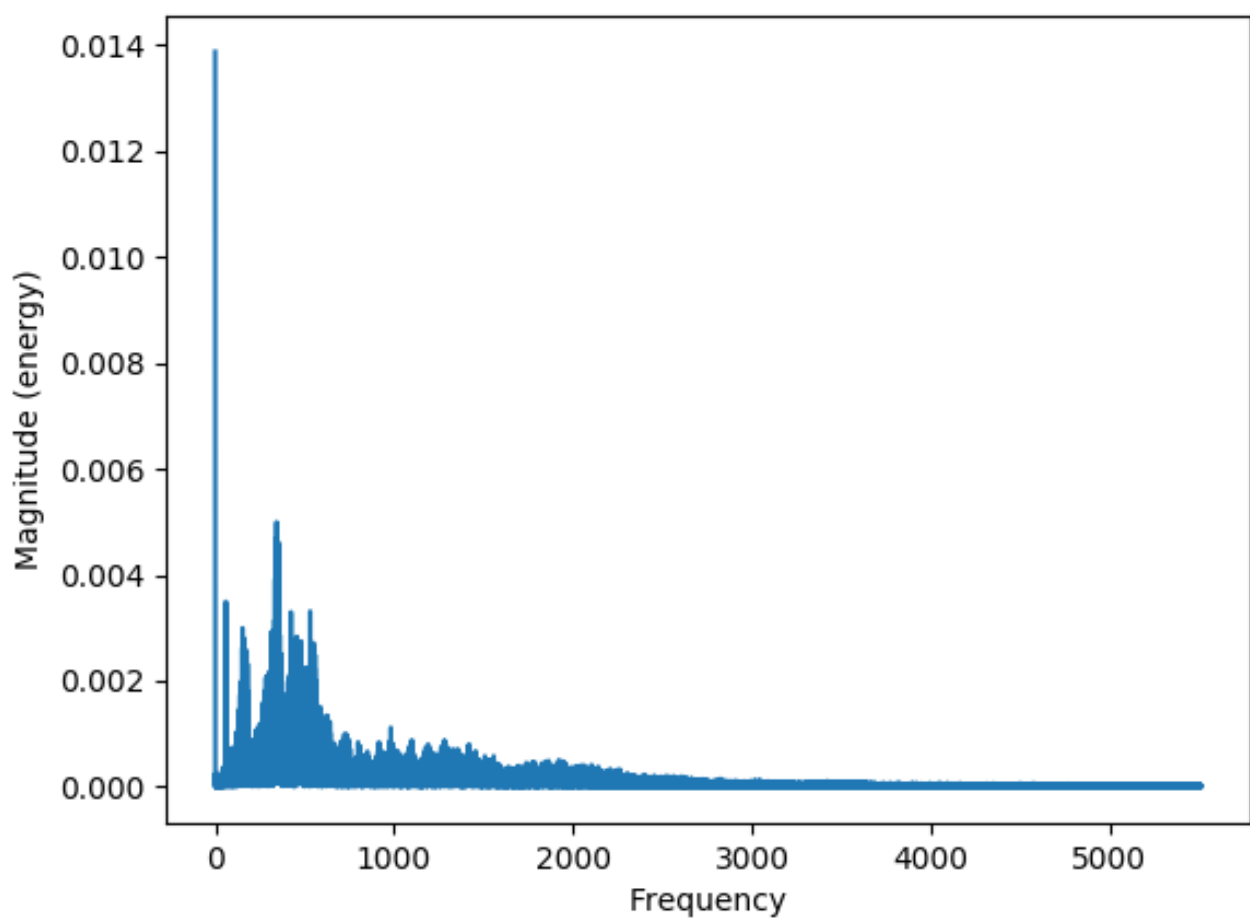


Figura 2: Espectro do diálogo limpo, i.e., sem ruído acústico.

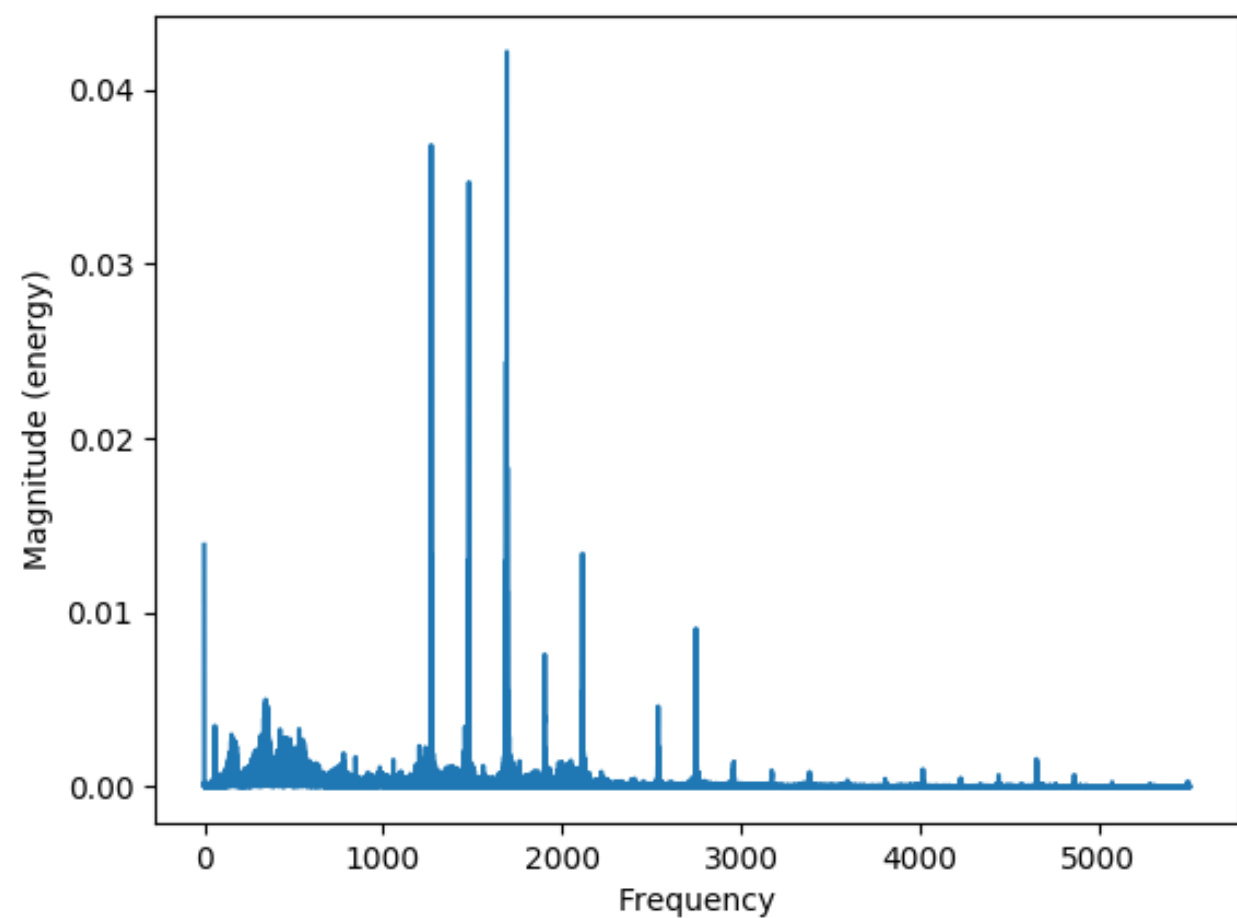


Figura 3: Espectro do diálogo com ruído *echo planar*.

SRA do diálogo com ruído

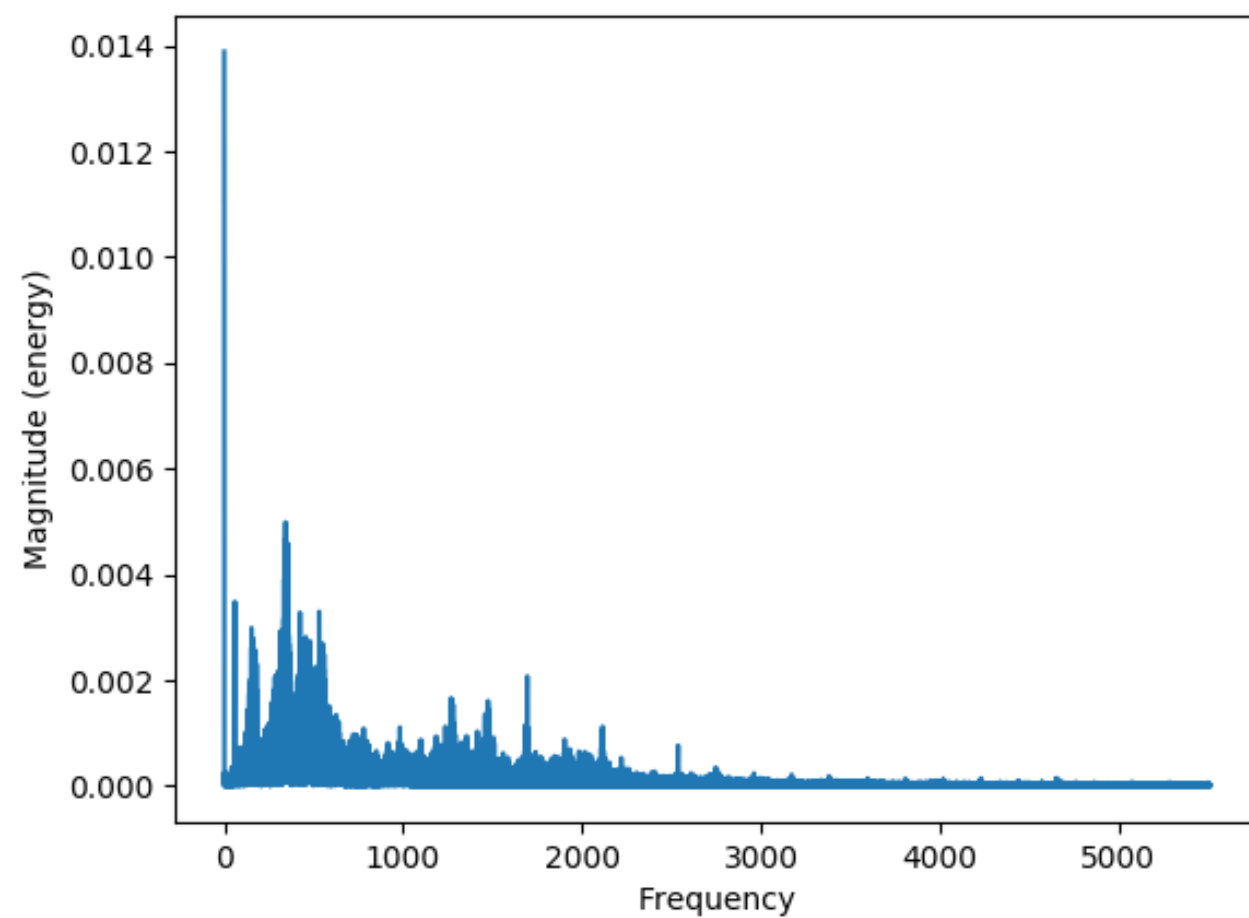


Figura 4: Espectro do diálogo com ruído *echo planar* após a supressão utilizando subtração espectral.

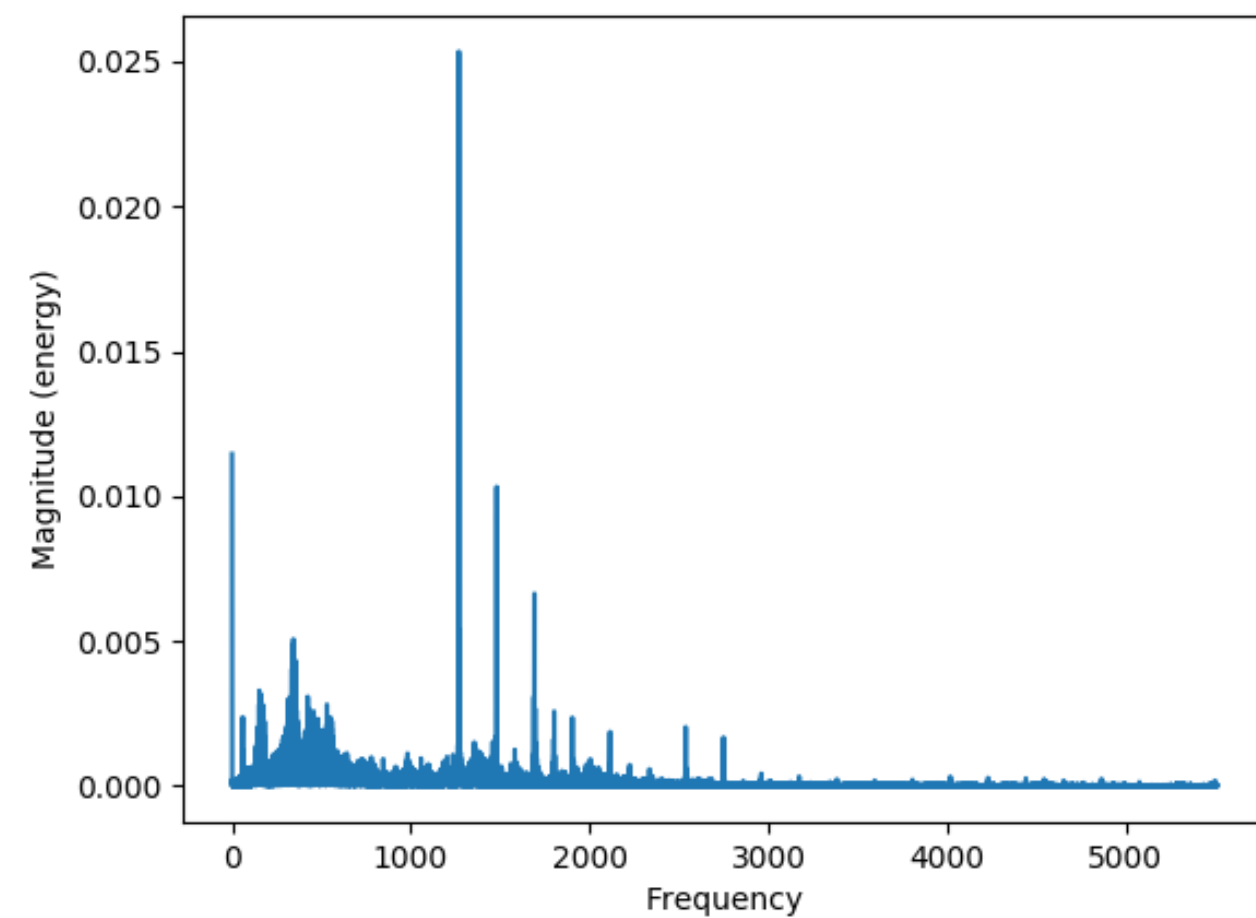


Figura 5: Espectro do diálogo com ruído *echo planar* após a supressão utilizando o método FLMS.

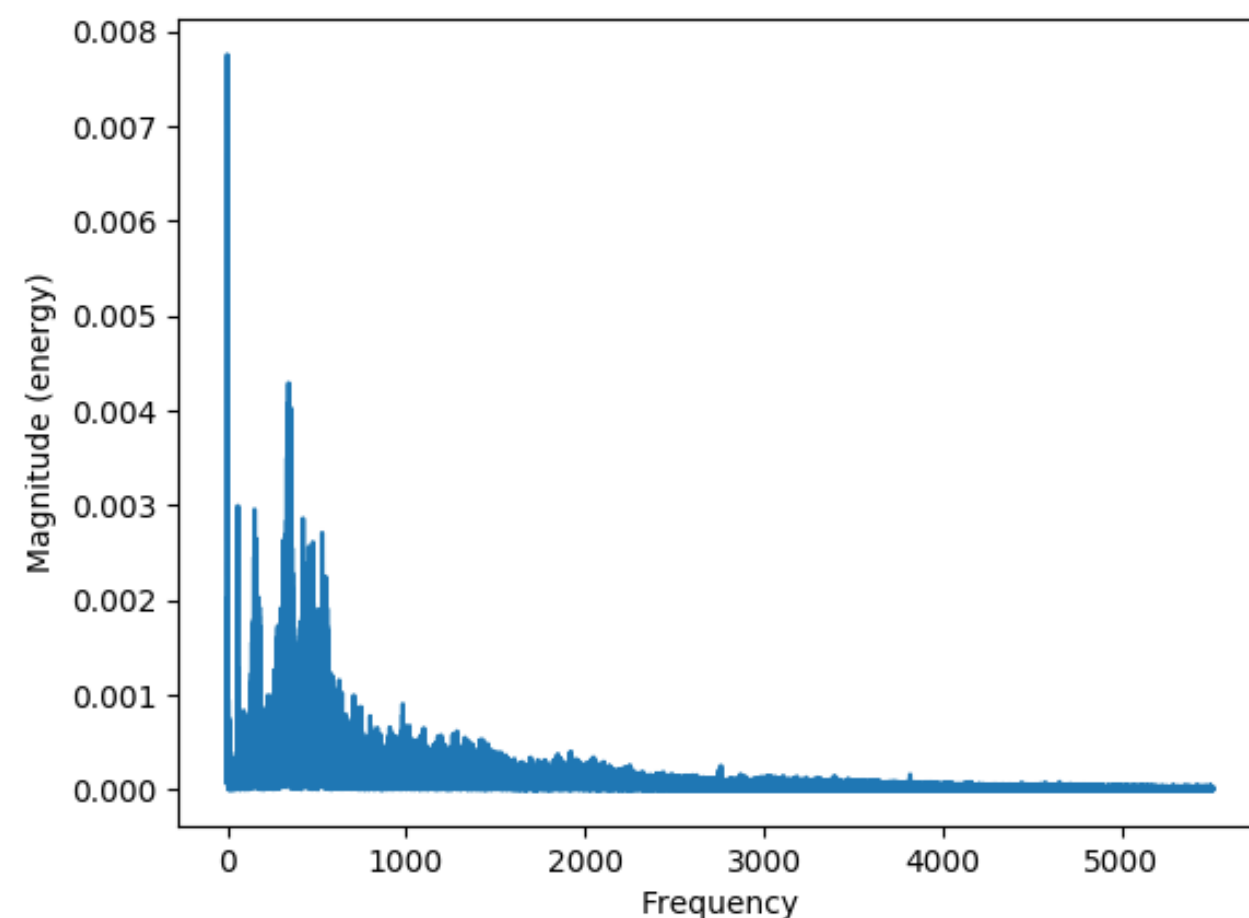


Figura 6: Espectro do diálogo com ruído *echo planar* após a supressão utilizando aprendizagem de dicionários e análise *wavelet*.

Diferença de energia total e comparação entre a amplitude da fala e do ruído

- A seguir, apresentamos a diferença de energia total entre o diálogo sem ruído e o diálogo com ruído e os sinais referentes à aplicação de SRA.
- Quão mais próximo de zero mais próximas são as energias totais dos sinais. Valores maiores que zeros indicam que o sinal resultante tem energia total maior que o sinal original e valores menores que zero indicam que o sinal tem energia total menor que o sinal original.

Tabela 2: Valores das diferenças de área dos espectros em relação ao diálogo sem ruído

	Fala com ruído	Subtração espectral	FLMS	Dicionário e wavelet
Dif.	0.048967	0.002359	-0.042717	-0.067662

- Analizamos a capacidade de realização de SRA para diferentes proporções de amplitude do sinal de diálogo em relação ao ruído acústico utilizando a diferença de energia total para comparar as diferentes proporções.
- A diferença de energia para fala com amplitude 100% corresponde à tabela 2.

Tabela 3: Diferenças de área dos espectros para o ruído *echo planar* com 1% da amplitude de fala

	Fala com ruído	Subtração espectral	FLMS	Dicionário e wavelet
Dif.	0.882254	0.162784	0.832077	-0.0121896

Tabela 5: Diferenças de área dos espectros para o ruído *echo planar* com 25% da amplitude de fala

	Fala com ruído	Subtração espectral	FLMS	Dicionário e wavelet
Dif.	0.207477	0.017063	0.026462	-0.168428

Tabela 4: Diferenças de área dos espectros para o ruído *echo planar* com 10% da amplitude de fala

	Fala com ruído	Subtração espectral	FLMS	Dicionário e wavelet
Dif.	0.413564	0.037801	0.322014	-0.209526

Tabela 6: Diferenças de área dos espectros para o ruído *echo planar* com 50% da amplitude de fala

	Fala com ruído	Subtração espectral	FLMS	Dicionário e wavelet
Dif.	0.106133	0.008061	-0.062564	-0.112022

Conclusão

Avaliação de complexidade de tempo

- O método de subtração espectral, o mais básico dentre os implementados, obteve bons resultados de supressão, sendo o método com menor complexidade de tempo. Já o método FLMS tem uma complexidade de tempo cerca de duas vezes maior que o primeiro método, o que ainda é pequeno, se comparado à duração dos áudios suprimidos.
- O método de aprendizado de dicionário de ruído e análise *wavelet* tem complexidade de tempo que pode chegar a cerca de vinte vezes o tempo para subtração espectral. Para o áudio de um minuto, esse método demorou mais de doze vezes o tempo gasto pelo método de subtração espectral.

Avaliação de qualidade

- O sinal resultante do método de subtração espectral tende a ter energia maior que o sinal que contém apenas fala.
- O método FLMS tende a produzir sinal de energia menor que o sinal de fala, porém, utilizamos o próprio sinal de fala como resultado desejado, o que dificulta a utilização do método de forma prática, pois esse sinal é justamente o que tentamos encontrar.
- O método de aprendizado de dicionário de ruído e análise *wavelet* se mostrou como aquele com melhor resultado, em termos de aproximação do sinal com fala e ruído ao sinal de fala.
- Quanto mais próxima a amplitude do sinal de fala em relação ao sinal referente ao ruído acústico, melhor é a aproximação ao sinal de fala original resultante da aplicação dos métodos de supressão implementados.

Observações finais

- O método de aprendizado de dicionários de ruído e análise *wavelet* se mostrou o mais eficiente qualitativamente para realização de supressão, podendo ser indicado para supressão *offline*.
- Já para supressão *online*, indicamos o método de subtração espectral, dada sua baixa complexidade de tempo para execução associada a uma boa capacidade de supressão.

Referências

- BRESCH, E., NIELSEN, J., NAYAK, K., AND NARAYANAN, S. Synchronized and noise-robust audio recordings during realtime magnetic resonance imaging scans. *The Journal of the Acoustical Society of America* 120, 4 (2006), 1791–1794.
- HAYKIN, S. *Adaptive Filter Theory (3rd Ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.
- RAVICZ, M. E., MELCHER, J. R., AND KIANG, N. Y.-S. Acoustic noise during functional magnetic resonance imaging. *The Journal of the Acoustical Society of America* 108, 4 (2000), 1683–1696.
- VAZ, C., RAMANARAYANAN, V., AND NARAYANAN, S. A two-step technique for mri audio enhancement using dictionary learning and wavelet packet analysis. In *INTERSPEECH* (2013), pp. 1312–1315.