



UFRJ

Universidade Federal Do Rio de Janeiro
Instituto de Computação
Computação Científica e Análise de dados

Professor: João Paixão
Aluno: Igor Pires Antunes

Projeto final

Compressão otimizada de espectrogramas através da análise dos componentes principais

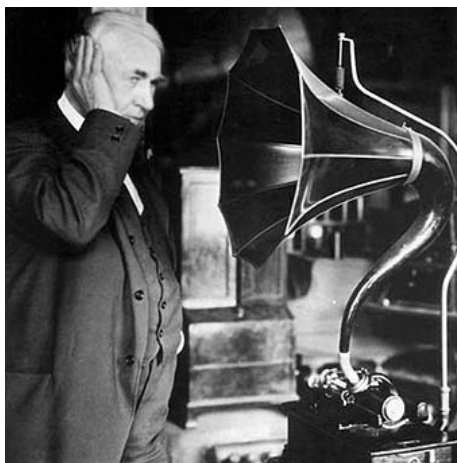
Rio de janeiro, 18 de julho de 2024

Sumário

Introdução.....	3
Representação.....	4
A entrada.....	5
Análise de componentes principais e redução de posto.....	6
PCA aplicado ao espectrograma.....	7
Ponto ótimo de compactação.....	10
Valor de armazenamento poupado.....	12
Conclusão.....	12
Referências.....	13

Introdução

Arquivos de áudio configuram um dos principais veículos de dados amplamente produzidos na era digital. Desde a invenção do fonógrafo por Thomas Edison em 1877, o som tornou-se um meio para armazenamento e leitura de informações. Com o advento das tecnologias digitais, o processamento de áudio evoluiu significativamente, permitindo não apenas a reprodução e gravação, mas também a análise, transformação e manipulação dos sinais de áudio para diversos fins.



Thomas Edison e o fonógrafo

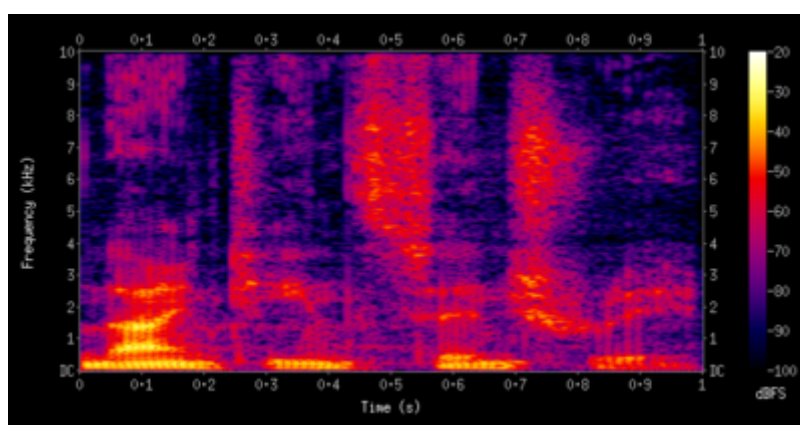
Um dos principais desafios relacionados à manipulação de dados é sua compactação, isto é, a redução da memória necessária para sua transmissão e armazenamento ao menor detrimento possível da informação transmitida. É relevante pontuar a diferença entre compressão de áudio e compressão de arquivo de áudio, sendo a primeira referente à normalização de volume e ou amplitude de um sinal sonoro, enquanto a segundo, objeto de nossa análise, compreende a compactação de informações previamente mencionadas.

O ponto de partida deste projeto é a compreensão de que a informação contida em arquivos de áudio pode ser modelada para tornar-se objeto de nosso estudo de matrizes. Visualizando o áudio como um espectrograma, pode ser convertê-lo em uma matriz para então realizar a análise de seus componentes principais. Com a comparação entre as diferenças obtidas entre a versão original e sua reconstrução compactada, podemos definir um Grau adequado de compressão de dados, pela redução de posto.

Representação

O formato de arquivo de áudio escolhido foi o Waveform (WAV), pela maior disponibilidade de funções e bibliotecas para sua manipulação. Compreendendo a necessidade de converter o arquivo de áudio em um objeto de análise matricial, a primeira, e mais simples alternativa, foi sua conversão direta em um vetor, prontamente disponível na biblioteca *numpy*, contudo representaria um impasse para a aplicação do PCA. Desse modo, uma implementação se mostrou uma alternativa viável para a representação e ainda com um bônus visual: o espectrograma.

Um espectrograma, de forma resumida, é um gráfico que analisa a densidade espectral da energia, num plano tempo x frequência. No contexto sonoro, ele mostra a distribuição de frequências como componentes do som analisado em relação ao tempo, seguindo uma série de Fourier. Esta representação não apenas nos fornece informações visuais sobre o comportamento de um sinal sonoro, mas também nos permite representá-lo como uma matriz Frequência x Tempo, onde cada elemento indica a magnitude, em decibéis, da banda em questão no intervalo de tempo especificado pela taxa de amostragem.



Exemplo de espectrograma

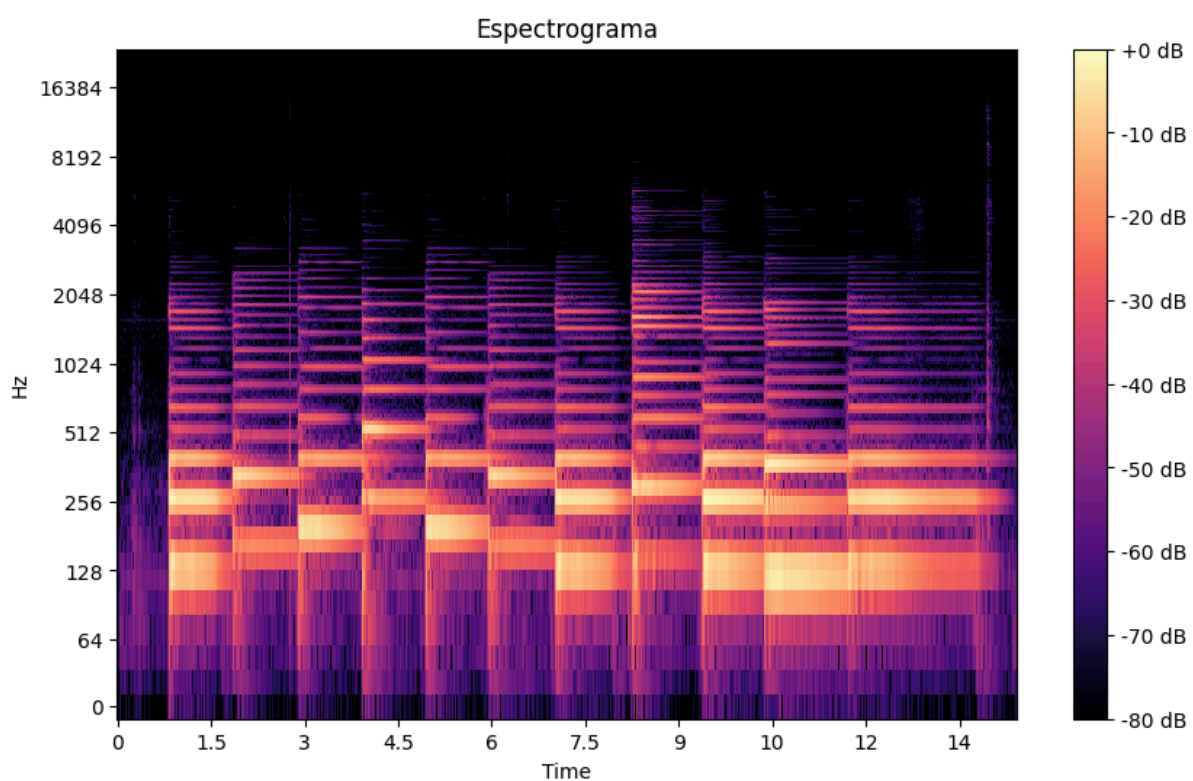
Para obter tal visualização, de forma resumida uma vez que não é o foco de nosso estudo, uma Transformada de Fourier de Curto Prazo (STFT) é aplicada ao sinal de áudio que é representado como um vetor de amostras no domínio do tempo. De maneira análoga, a Inversa é aplicada à matriz espectral para que seja obtido novamente um sinal de áudio, que pode ser novamente convertido para um arquivo.

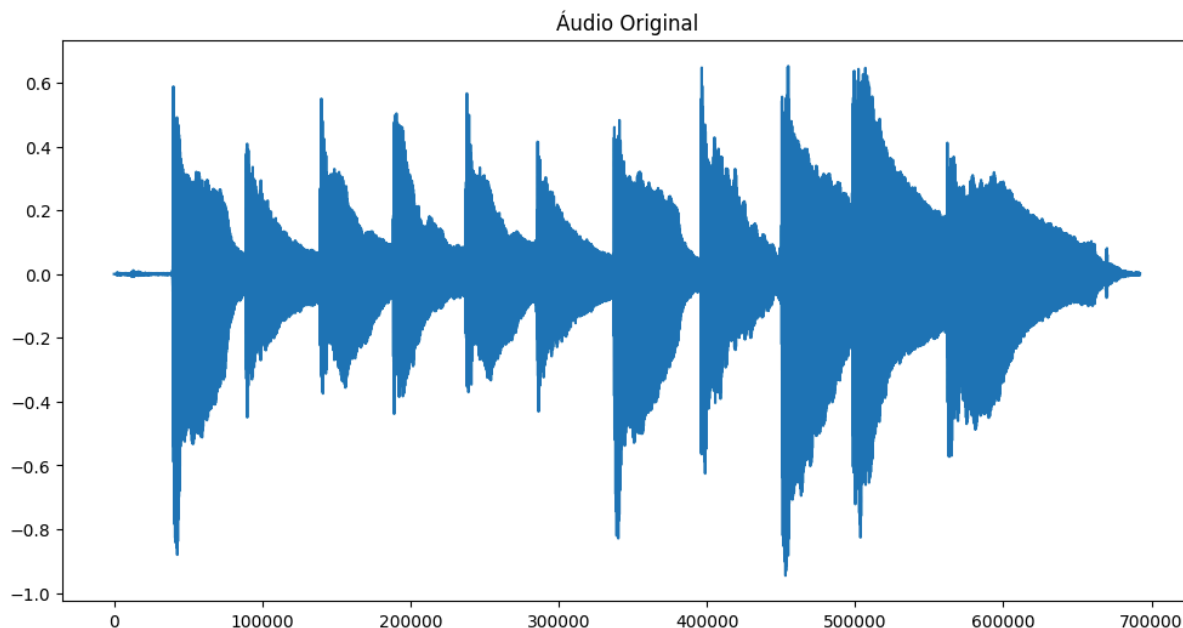
Para manter o escopo desse projeto sobre a manipulação matricial e análise de componentes principais do sinal auditivo, considera-se como objetivo final promover a compactação da matriz espectral, e não o arquivo de áudio em si, uma vez que o método de

armazenamento de dados dos formatos de áudio testados não empregam um modelo matricial, resultando em arquivos igualmente densos ao final da conversão.

A entrada

A entrada escolhida foi um arquivo WAV de uma gravação de 9 segundos feita por mim, na qual uma sequência de 11 notas são tocadas no piano. Este modelo foi escolhido por proporcionar uma visualização facilitada da série harmônica de cada nota no espectrograma, bem como a distinção clara entre elas na visualização do gráfico da onda, ambos dispostos abaixo.





A partir da matriz espectral gerada, é aplicado o PCA.

Análise de componentes principais e redução de posto

A análise de componentes principais é um método de simplificação de datasets com a preservação de padrões e tendências por meio da redução dimensional dos mesmos. Suas aplicações são vastas, desde a compressão de dados ao processamento dos mesmos em formatos diversos. É particularmente útil em algoritmos de aprendizado e identificação de padrões. O PCA não apenas reduz o tempo de computação, mas também ajuda a mitigar problemas de sobreajuste ao eliminar variáveis irrelevantes ou redundantes, como ruídos.

Sua implementação é dada nos seguintes passos:

1. Padronizar a escala da matriz original: Isso é feito subtraindo a média e dividindo pelo desvio padrão de cada variável, resultando em variáveis com média zero e variância um.
2. Calcular a matriz de covariância: Obtida como produto da matriz original transposta com si própria, ou vice versa.

3. Encontrar os componentes principais: obtidos pelo cálculo dos autovalores da matriz de covariância, cujos n maiores associam-se aos n componentes principais.
4. Construção da matriz V dos componentes principais: A matriz V é formada pelos autovetores associados aos maiores autovalores da matriz de covariância. Cada coluna de V representa um componente principal.
5. Construção da matriz dos coeficientes: A matriz dos coeficientes é obtida projetando os dados padronizados no subespaço gerado pelos componentes principais. Isso é feito multiplicando a matriz dos dados padronizados pela matriz V .
6. Construção da matriz reduzida: A matriz reduzida é obtida multiplicando a matriz dos coeficientes pela matriz V transposta. Esta matriz representa os dados originais projetados no espaço dos componentes principais, mantendo apenas as dimensões mais importantes.

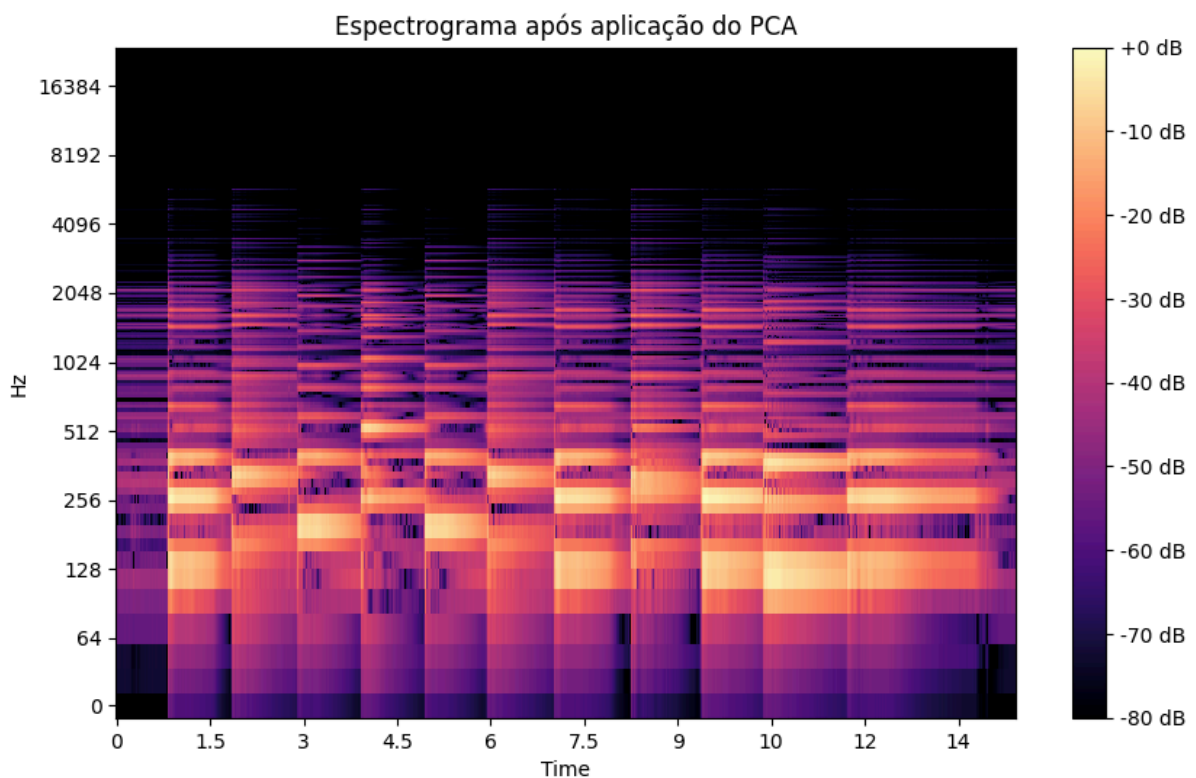
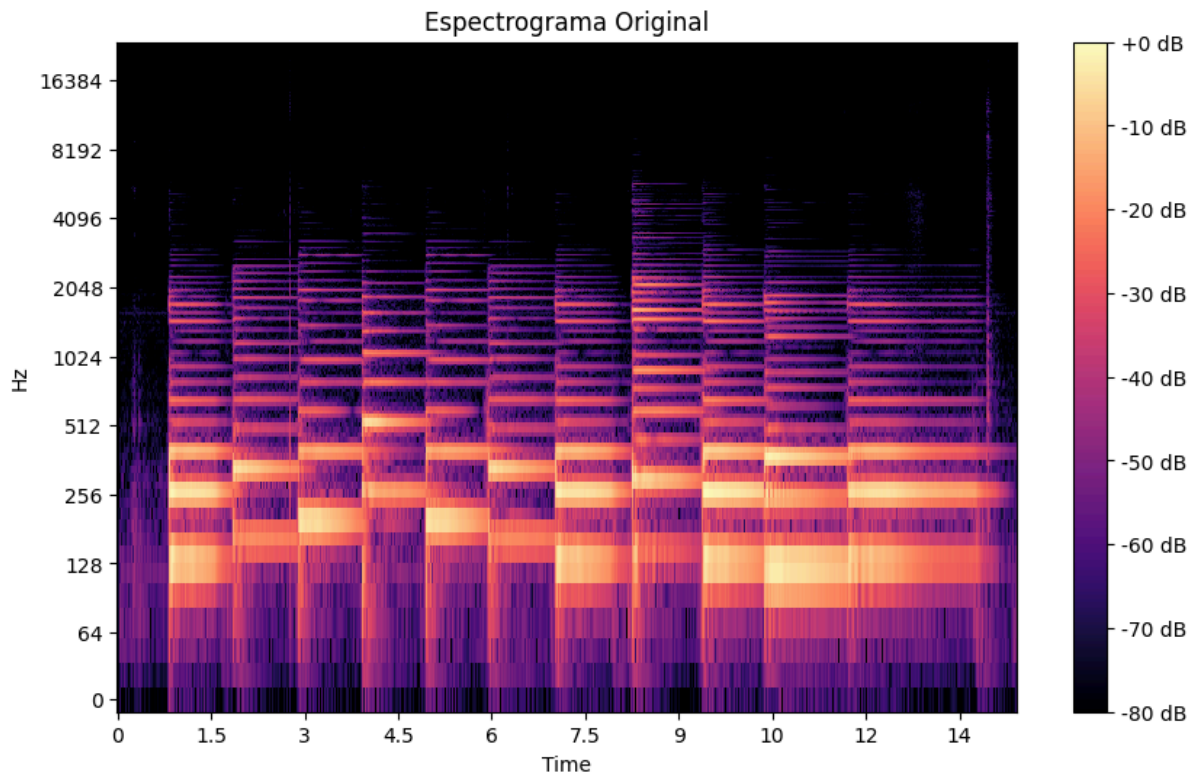
Assim, é possível obter uma matriz de posto reduzido em relação à original, mantendo as principais características dos dados dispostos.

Para fins de compactação de dados, vale observar que armazenar a matriz de componentes principais juntamente à matriz dos coeficientes demandará uma quantidade menor de memória em comparação à matriz original. Isto ocorre porque o PCA reduz a dimensionalidade dos dados ao projetá-los em um subespaço de menor dimensão, formado pelos componentes principais. Em vez de armazenar todos os dados originais em suas dimensões completas, apenas os componentes mais significativos são mantidos.

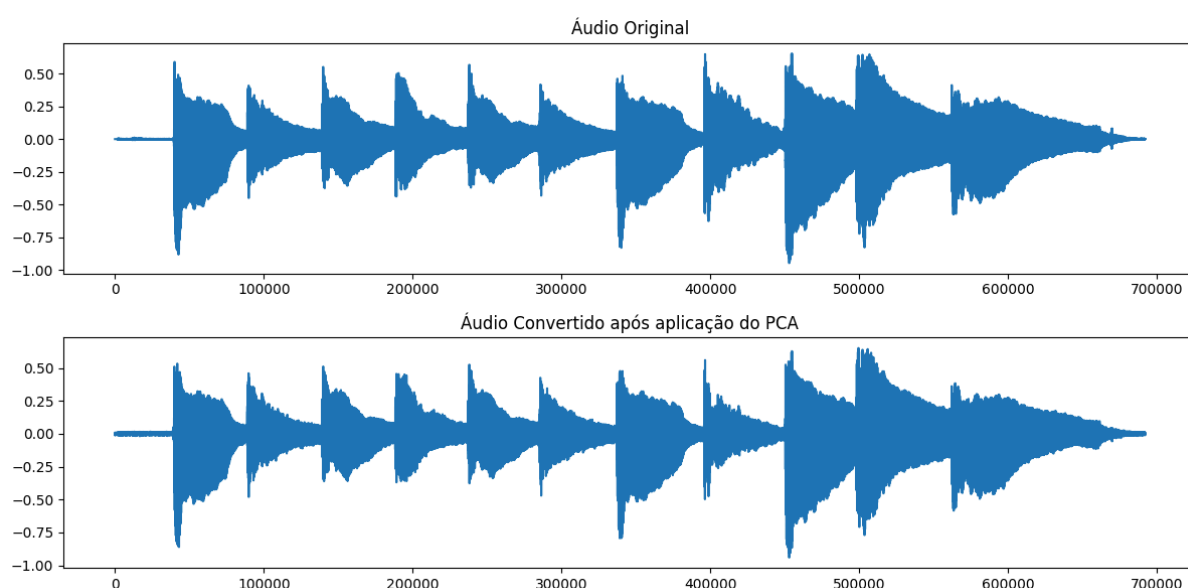
PCA aplicado ao espectrograma

Uma função que implementa a análise de componentes principais é aplicada ao espectrograma original. Ela recebe também um parâmetro índice, que indica o percentual de componentes principais mantidos para a nova matriz. Toma-se como chute inicial 90%, destoando pouco da original. A nova matriz é gerada, bem como uma nova visualização espectral e uma nova onda para a faixa gerada

Observa-se que o decaimento natural da magnitude das frequências após cada ataque de nota, natural por tratar-se de um piano, é homogeneizado com a aplicação do PCA, que identifica os componentes principais dentre vetores de frequência correspondentes à distribuição harmônica de cada nota tocada.



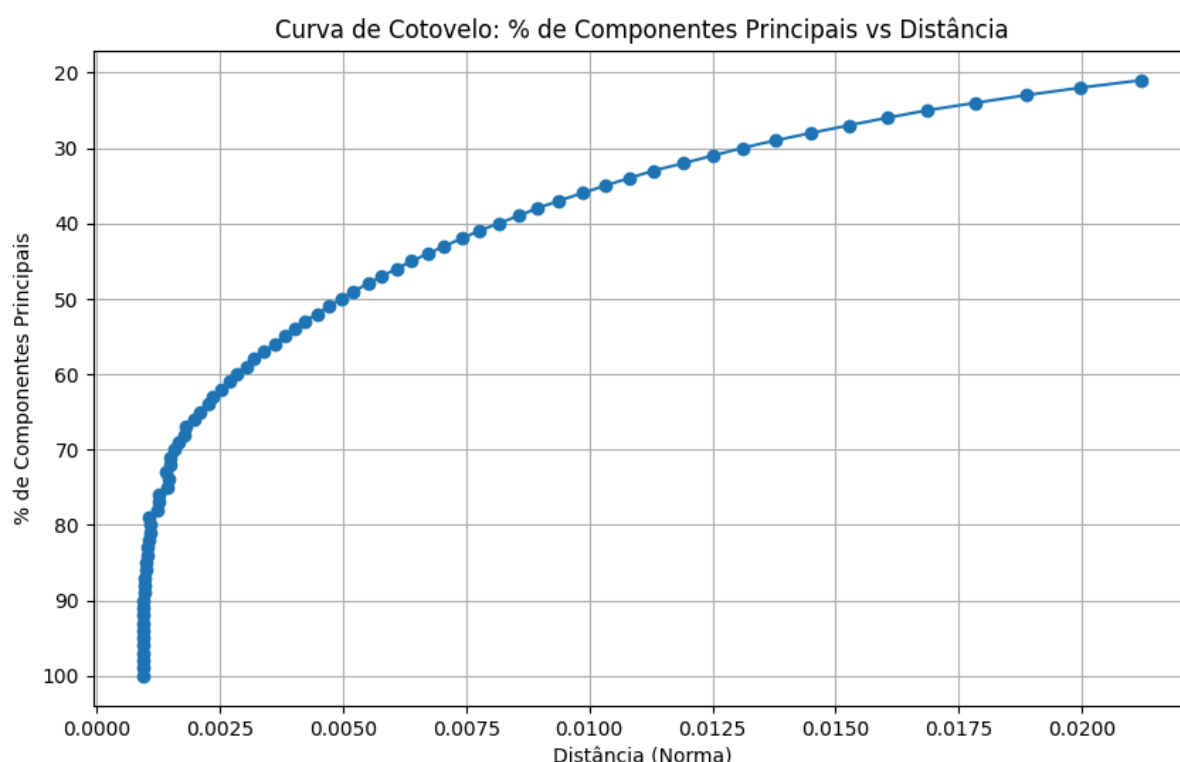
Observa-se que o decaimento natural da magnitude das frequências após cada ataque de nota, natural por tratar-se de um piano, é homogeneizado com a aplicação do PCA, que identifica os componentes principais dentre vetores de frequência correspondentes à distribuição harmônica de cada nota tocada.



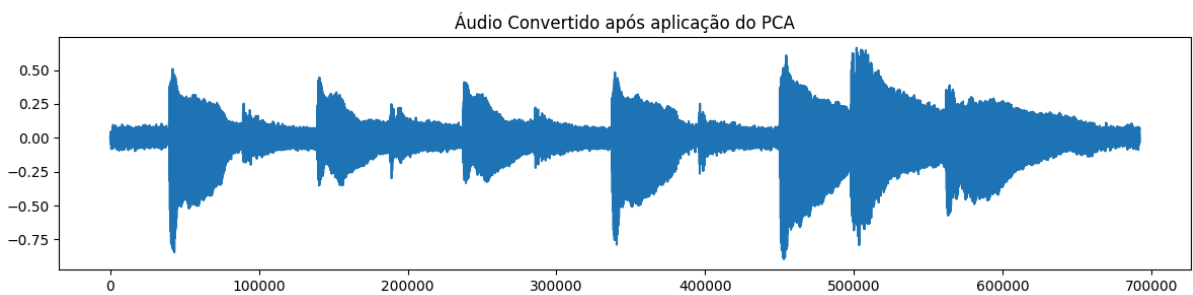
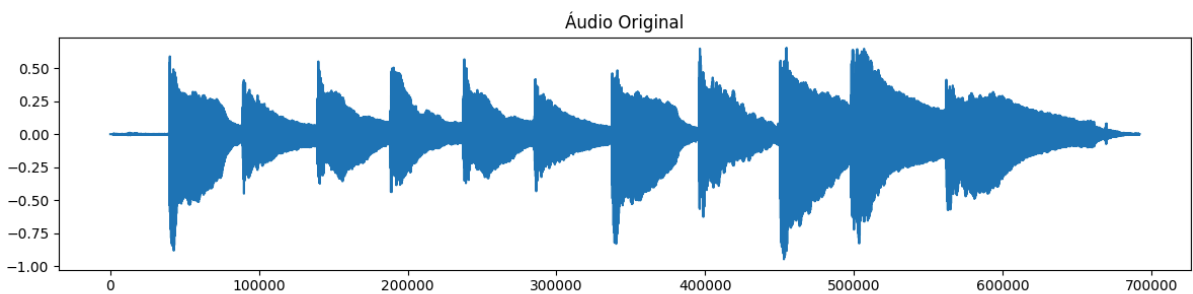
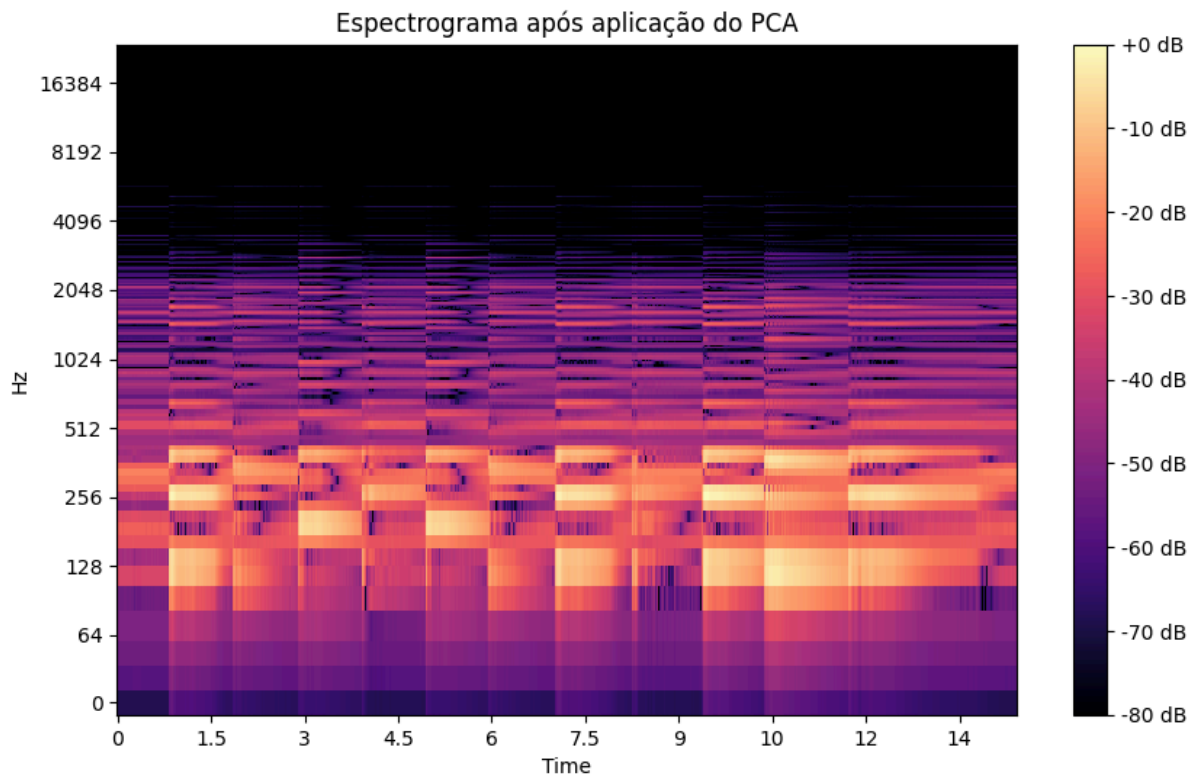
É notável pela visualização das ondas antes e pós PCA que seu formato se manteve em grande parte inalterado, com total inteligibilidade das notas tocadas. Foi gerado um novo arquivo de áudio e sua audição comprova as expectativas geradas pelas imagens acima: houve mínima, ainda que audível, distorção de áudio.

Ponto ótimo de compactação

Um gráfico foi produzido para dispor pontos que relacionam o percentual de componentes principais mantidos após a implementação do PCA com o erro gerado pelo mesmo. O erro foi interpretado como a norma do vetor distância entre o vetor da onda original e o vetor da onda de posto reduzido. O objetivo deste gráfico era permitir a identificação do “cotovelo da curva”, isto é, o ponto máximo de redução de posto no qual as características originais do modelo são, em grande parte, preservadas. Assim, partindo de 100% dos componentes principais, 1% for reduzido a cada iteração, produzindo a seguinte curva:



Como visto, não é possível determinar o “cotovelo” ou um ponto exato de quebra na qual componentes principais passam a adicionar ruído, mas é válido assumir que para valores até 80% encontramos mínima perda de informações em relação ao arquivo original. Assim, o valor ótimo escolhido para o índice PCA estabelecido para compactação é 0.80. Assim, geram-se novas visualizações para o valor escolhido



Agora, entretanto, notam-se mais severas distorções sobre a onda do sinal de áudio de posto reduzido. Ouvindo o arquivo isto é notável, mas ainda é possível discernir com facilidade as notas tocadas.

Valor de armazenamento poupado

Por fim, uma última função compara a quantidade de bytes necessários para armazenar a matriz original com a quantidade de bytes necessários para armazenar a matriz de componentes principais junto a matriz de coeficientes. A razão entre ambos nos dá a economia de armazenamento ótima obtida, igual a 30%.

Bytes necessários para armazenar o espectrograma original: 11094600

Bytes necessários para armazenar os componentes principais: 4437840

Bytes necessários para armazenar os coeficientes: 3362000

Total de bytes necessários para armazenar os componentes da transformada PCA: 7799840

Percentual de redução: 29.70%

Conclusão

A aplicação da Análise de Componentes Principais (PCA) ao espectrograma de um arquivo de áudio demonstrou ser uma técnica eficaz para a compactação de dados, permitindo uma significativa redução do tamanho do arquivo enquanto preserva a maior parte das informações essenciais. Ao padronizar as variáveis contínuas, calcular a matriz de covariância, identificar os componentes principais e reconstruir os dados no subespaço gerado pelos componentes principais, foi possível obter uma matriz de posto reduzido. Esse processo resultou em uma economia de armazenamento de aproximadamente 30%, o que é altamente benéfico para o armazenamento e transmissão de grandes volumes de dados de áudio. Além disso, a análise gráfica da relação entre o índice do PCA e a distância entre os vetores de áudio original e transformado permitiu identificar o ponto ótimo de compactação, garantindo que a qualidade do áudio fosse mantida dentro de limites aceitáveis. Esses resultados destacam a eficiência do PCA não apenas na redução dimensional, mas também na

manutenção da integridade dos dados, tornando-o uma ferramenta valiosa para diversas aplicações em processamento de áudio e compressão de dados.

Referências

SPECTROGRAM. Wikipedia, The Free Encyclopedia. Disponível em: <https://en.wikipedia.org/wiki/Spectrogram>. Acesso em: 18 jul. 2024.

YANG, Thao. Step-by-step explanation of principal component analysis. Built In. Disponível em: <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>. Acesso em: 18 jul. 2024.

GRAVAÇÃO. Wikipédia, a enciclopédia livre. Disponível em: <https://pt.wikipedia.org/wiki/Grava%C3%A7%C3%A3o#:~:text=A%20primeira%20grava%C3%A7%C3%A3o%20de%20som%20registrada%20%C3%A9%20da%20can%C3%A7%C3%A3o%20folcl%C3%B3rica,em%20um%20arquivo%20em%20Paris>. Acesso em: 18 jul. 2024.

AUDIO COMPRESSION EXPLAINED. SoundGuys. Disponível em: <https://www.soundguys.com/audio-compression-explained-29148/>. Acesso em: 18 jul. 2024.

TIMELINE OF AUDIO FORMATS. Wikipedia, The Free Encyclopedia. Disponível em: https://en.wikipedia.org/wiki/Timeline_of_audio_formats. Acesso em: 18 jul. 2024.

SCIPY.IO.WAVFILE.READ. SciPy v0.14.0 Reference Guide. Disponível em: <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.io.wavfile.read.html>. Acesso em: 18 jul. 2024.

HOW TO READ A MP3 AUDIO FILE INTO A NUMPY ARRAY? SAVE A NUMPY ARRAY TO MP3. Stack Overflow. Disponível em: <https://stackoverflow.com/questions/53633177/how-to-read-a-mp3-audio-file-into-a-numpy-array-save-a-numpy-array-to-mp3>. Acesso em: 18 jul. 2024.

SCIPY.IO.WAVFILE.WRITE. SciPy v0.14.0 Reference Guide. Disponível em: <https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.io.wavfile.write.html>. Acesso em: 18 jul. 2024.

CONVERT AUDIO FILE IN A FREQUENCY MATRIX. MATLAB Central. Disponível em: <https://www.mathworks.com/matlabcentral/answers/1634230-convert-audio-file-in-a-frequency-matrix>. Acesso em: 18 jul. 2024.

SHORT-TIME FOURIER TRANSFORM. Wikipedia, The Free Encyclopedia. Disponível em: https://en.wikipedia.org/wiki/Short-time_Fourier_transform. Acesso em: 18 jul. 2024.

NUMPY SPECTROGRAM. Lo.calho.st. Disponível em: <https://lo.calho.st/posts/numpy-spectrogram/>. Acesso em: 18 jul. 2024.

PCA IN PYTHON. Built In. Disponível em: <https://builtin.com/machine-learning/pca-in-python>. Acesso em: 18 jul. 2024.