

Modelagem de Efeitos de Distorção em Sinais de Guitarra Elétrica com Uso de Redes Perceptron Multicamadas

Murilo G. Correia¹, Samuel C. De Almeida¹, Alisson A. Cardoso¹

¹Escola de Engenharia Elétrica, Mecânica e de Computação –
Universidade Federal de Goiás (UFG) – Goiânia – GO – Brasil

{murilogcorreia, carvalho.carvalho}@discente.ufg.br, alsnac@ufg.br

Abstract. *This study investigates the modeling of distortion effects in electric guitar signals using multilayer perceptron (MLP) networks. It is demonstrated that the trained MLP can effectively reproduce these effects, validating this with metrics such as mean squared error (MSE) and Kolmogorov-Smirnov (KS). This approach has potential applications in the music industry and audio production, representing an advancement in simulating complex sound effects.*

Resumo. *Este trabalho investiga a modelagem dos efeitos de distorção de sinais de guitarra elétrica com redes perceptron multicamadas, traduzida do inglês multilayer perceptron (MLP). É demonstrado que a MLP treinada pode reproduzir efetivamente esses efeitos, validando isso com métricas como erro quadrático médio (EQM) e Kolmogorov-Smirnov (KS). Esta abordagem tem aplicações potenciais na indústria musical e na produção de áudio e representa um avanço na simulação de efeitos sonoros complexos.*

1. Introdução

As guitarras elétricas e os pedais fazem parte das características do rock como estilo musical. Esses pedais introduzem uma variedade de sons, principalmente sons agressivos e excêntricos, podendo também serem mais suaves e brandos. Essa diversidade ajuda tanto os músicos a criarem sua identidade pessoal quanto a da banda. Um exemplo disso é a utilização de pedais de distorção com sonoridade pesada usados pela Banda Nirvana, tipo o Fuzz BIG MUFF. Tais pedais contribuíram para a sonoridade mais característica da banda bem como do movimento *Grunge* do qual ela faz parte [Jones 2018].

Sabendo da importância desses circuitos eletrônicos, se faz necessário entendê-los. Porém, compreender o funcionamento de um pedal é complexo pela característica não linear presente nesses pedais. Essa característica se dá pelo uso de componentes não lineares na construção do circuito, sendo esses componentes grandes responsáveis pela sonoridade ofertada pela distorção. O problema é evidenciado com as tecnologias de distorções digitais e a necessidade delas de se aproximar do som ofertado pelos circuitos analógicos, uma vez que esses circuitos analógicos estão atrelados a sonoridade histórica de bandas e artistas.

Uma alternativa prática e também motivação desta pesquisa é utilizar-se das inteligências artificiais como meio de modelar um sistema não linear. A ideia central é fornecer a uma inteligência artificial o sinal de entrada de uma guitarra sem distorções bem como o sinal de saída, aplicando o sinal de entrada a uma distorção. A partir da aplicação do aprendizado de máquina, a inteligência artificial deve ser capaz de entregar

uma modelagem do efeito de distorção usado no sinal de guitarra elétrica. Sendo uma tarefa complexa, a rede neural se apresenta como possível solução.

A rede neural possui esse nome por buscar no cérebro humano, principalmente no comportamento dos neurónios no processo cognitivo, inspiração de como processar dados para executar uma tarefa. De forma semelhante ao processamento cognitivo humano, a unidade básica de uma rede neural, o neurônio artificial, se liga a diversos outros por meio dos terminais de entrada, analogamente aos dendritos. Os valores recebidos são ponderados e combinados por uma função matemática. A saída então se dá como a resposta do neurônio para a entrada e esse é interconetado a rede neural, dependendo das características da rede, até a camada de saída. Dessa forma, a função implementada por um neurônio é a combinação das funções realizadas pelos neurônios anteriores e, assim, a informação é processada [Faceli et al. 2021].

2. Referencial Teórico

A distorção, comum na música, muda a forma do áudio, gerando um som saturado e expressivo na guitarra elétrica. Modelar e controlar essa distorção é vital para músicos e engenheiros de áudio. Redes Neurais Artificiais (RNA), especialmente Perceptron Multicamadas (MLPs), oferecem uma promissora abordagem para essa modelagem. Este referencial teórico explora distorção de áudio, aplicação de RNAs no processamento de áudio e a relevância das MLPs nesse contexto.

2.1. Distorção de áudio

A distorção de áudio pode ser classificada em dois tipos principais: harmônica e não harmônica. A distorção harmônica ocorre quando os harmônicos de um tom fundamental são produzidos de maneira controlada, resultando em um som mais rico e musical. Em contraste, a distorção não harmônica produz componentes de frequência não relacionados ao som original, muitas vezes resultando em uma cacofonia áspera [Reiss and McPherson 2014].

2.2. Pedal e simulação

2.2.1. Elementos para obtenção dos dados

A escolha arbitrária do pedal não afeta no resultado da pesquisa, bem como a forma que o sinal de saída do pedal é coletado. Como o sinal de entrada e o sinal de saída do pedal são fornecidos à rede neural, o que importa é se a rede neural consegue uma aproximação da função de transferência, mas qual é a função de transferência em si não importa para este escopo.

Dessa forma, também se justifica a utilização de um software de simulação de circuitos para simular o pedal pois não cabe a esta pesquisa avaliar a qualidade de simulação do software mas sim a capacidade da rede neural de relacionar dois sinais. Esse fato ressalva a aplicabilidade do processo em todos os sistemas dinâmicos não lineares, como os pedais de distorção. Assim, um pedal de distorção genérico foi escolhido pela sua relativa simplicidade e sua ampla utilização no gênero do rock. Semelhantemente, o software para simulação, o LTSpice ("Analog Devices", 2023), foi escolhido pela simplicidade e pelo fato de ser gratuito [Devices 2023].

2.2.2. Esquemático do pedal

O pedal foi dividido em 4 partes. A primeira dessas partes é o estágio de entrada de sinal (do inglês *input signal stage*). O sinal escolhido estava com uma tensão alta, mais alta do que costuma ser fornecido de entrada por uma guitarra. Por tal motivo, foi colocado uma fonte dependente com ganho de 0.2, valor esse que trouxe o sinal de entrada para um valor mais próximo da realidade e, na montagem do circuito analógico, o sinal da guitarra deve ser aplicado no nó sinalizado pela palavra *input*. A segunda parte do pedal é o estágio de alimentação de energia (do inglês *power supply stage*), o uso dos altos valores de resistência contribuem para uma alta impedância de entrada do circuito além de uma alta impedância para o terra do curto-circuito virtual. O capacitor tem o papel de eliminar a variação residual conhecida como ondulação (do inglês *ripple*), que nada mais é do que um sinal AC indesejado. A Fig. 1 mostra o diagrama elétrico do circuito do pedal.

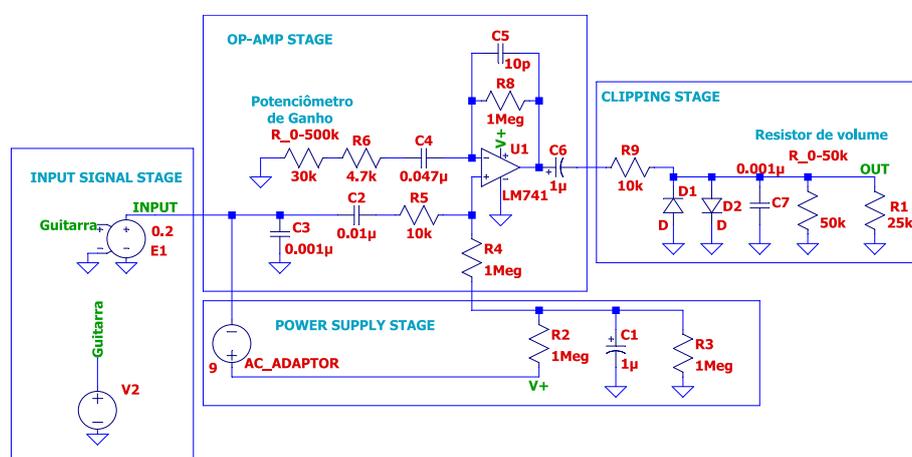


Figura 1. Diagrama elétrico do pedal.

Ademais, a terceira parte do pedal é o estágio de amplificação do Amp Op (do inglês *Op Amp Amplifier Stage*, ou somente *Op Amp Stage* como referido no esquemático). O amplificador se encontra na configuração não-inversora, o que gera um sinal de saída como relação dos resistores R8, R6 e R0, sendo esse último o potenciômetro que permite ao usuário controlar o ganho. Além disso, nesse estágio, há o capacitor C3 cuja função é evitar ruído de rádio frequência bem como ajudar em descargas eletrostáticas e oscilações. O capacitor C6 remove voltagem DC ao próximo estágio. Os outros capacitores contribuem para a resposta em frequência com pico em torno de 1,5 KHz, característica comum em outros pedais e que ajuda a guitarra a ter destaque em relação a outros instrumentos na música ao ter maior ganho na faixa de frequência audível para os seres humanos.

O último estágio é o de recorte (do inglês *Clipping Stage*). Nele, o resistor R9 limita a corrente que chega aos diodos, protegendo tais. Os diodos possuem uma tensão de saturação e, pela disposição deles, saturam com tensões menores que o negativo da tensão de saturação e valores maiores que o positivo dessa tensão. Com isso, o sinal é cortado além desses limites, dando a sonoridade da distorção. Quanto mais abrupto o corte, mais distorcido o som fica. Finalmente, há um potenciômetro para regulagem do volume bem como a impedância de saída do pedal, essa última que não está presente na

construção do pedal mas é importante para a simulação por simular a carga com máxima transferência.

2.3. Redes Neurais Artificiais em Processamento de Áudio

Redes neurais perceptron multicamadas (MLPs) são um tipo de RNA comumente usado em problemas de aprendizagem supervisionada. Eles consistem em camadas de neurônios (perceptrons) organizadas em uma estrutura de camadas, incluindo uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. A MLP é conhecido por sua capacidade de modelar funções não lineares complexas [Zhang et al. 2023].

A capacidade de modelar funções não lineares complexas da MLP é explicada ao entender que uma rede com uma camada intermediária pode implementar qualquer função contínua. A utilização de duas camadas intermediárias permite a aproximação de qualquer função [Faceli et al. 2021]. Sendo a função que se deseja modelar não linear, usar o método com múltiplas camadas permite uma aproximação dessa função e, por isso, a MLP foi escolhida como método de aprendizado de máquina aplicado.

2.4. Métodos de Avaliação dos resultados

A utilização das técnicas de teste de erro quadrático médio (EQM) e Kolmogorov-Smirnov (KS) é muito importante na avaliação de resultados de projetos de processamento de áudio, como a modelagem dos efeitos de distorção de um sinal de guitarra elétrica utilizando uma rede MLP.

2.4.1. Erro Quadrático Médio (EQM)

O erro quadrático médio é uma métrica comumente usada nos campos de processamento de sinais e aprendizado de máquina para quantificar a diferença entre um sinal estimado e um sinal de referência [Bishop and Nasrabadi 2006]. Este projeto usa EQM para medir a fidelidade de uma rede neural perceptron multicamadas (MLP) a um sinal de áudio de uma guitarra elétrica com som distorcido como referência. Quanto menor o valor do EQM, mais próxima a saída MLP está do sinal de referência, indicando melhor capacidade do modelo em reproduzir o efeito de distorção desejado.

O Erro Quadrático Médio (EQM) é calculado como:

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

onde:

- n é o número de amostras,
- y_i é o valor real da amostra i ,
- \hat{y}_i é o valor previsto da amostra i .

2.4.2. Teste de Kolmogorov-Smirnov (KS)

O teste Kolmogorov-Smirnov é um método estatístico usado para comparar duas distribuições de probabilidade e avaliar se elas vêm da mesma população [Stephens 1974].

Neste projeto, o KS foi aplicado para verificar se a saída MLP segue a mesma distribuição estatística do sinal distorcido da guitarra elétrica. Um valor alto para a estatística KS ou um valor p menor que o valor significativo de 0.05 indica que as duas distribuições são diferentes, sugerindo que a saída da MLP não reproduz adequadamente o efeito de viés desejado.

O Teste de Kolmogorov-Smirnov (KS) é usado para verificar a semelhança entre duas distribuições de dados. A estatística de teste KS é calculada como:

$$KS = \sup_x |F_1(x) - F_2(x)| \quad (2)$$

onde:

- $F_1(x)$ é a função de distribuição acumulada da primeira distribuição,
- $F_2(x)$ é a função de distribuição acumulada da segunda distribuição,
- \sup_x denota o supremo sobre todos os valores possíveis de x .

O valor p associado ao KS é usado para determinar a significância estatística da diferença entre as duas distribuições.

2.5. Trabalhos Relacionados

Alguns trabalhos relacionados nesta área oferecem perspectivas importantes:

Título	Referência	Correlação	Diferencial
"Deep Learning for Audio Signal Processing"	[Purwins et al. 2019]	Explora o uso de redes neurais em processamento de áudio	Destaca-se pela revisão abrangente de técnicas e aplicações em processamento de áudio
"Neural Audio Synthesis of Musical Notes with WaveNet Autoencoders"	[Engel et al. 2017]	Modelagem de síntese neural de áudio, incluindo efeitos	Aplica autoencoders WaveNet na síntese de áudio musical, destacando-se pela aplicação específica em notas musicais

Tabela 1. Trabalhos Relacionados

Diferencial deste trabalho: Este trabalho diferencia-se ao fornecer uma análise específica sobre distorção em sinais de guitarra elétrica.

3. Metodologia

3.1. Obtenção e Preparação dos Dados

Após definir os elementos necessários para coletar os dados, incluindo o software de simulação e o circuito eletrônico simulado, um áudio genérico de uma guitarra sem efeitos foi selecionado para ser usado como sinal de entrada. O programa permitiu incorporar esse áudio como entrada do circuito de simulação, utilizando a funcionalidade *waveform*. Com os parâmetros adequados definidos, o programa gerou o sinal de saída em formato de áudio, fornecendo os dados essenciais para as etapas seguintes da pesquisa.

Os dados foram cuidadosamente preparados para garantir a qualidade das entradas da MLP. Os áudios da guitarra elétrica com e sem distorção foram gravados a 48 kHz por 30 segundos, resultando em 1.440.000, capturando detalhes sutis. Foi introduzida a distorção aplicando ganho específico, criando pares de áudios correspondentes: um com e outro sem distorção. Isso permitiu uma comparação direta dos efeitos da distorção na saída da MLP em relação ao áudio não distorcido. A Figura 2 mostra um curto trecho (0,01s) comparando o áudio com distorção e o áudio sem distorção.

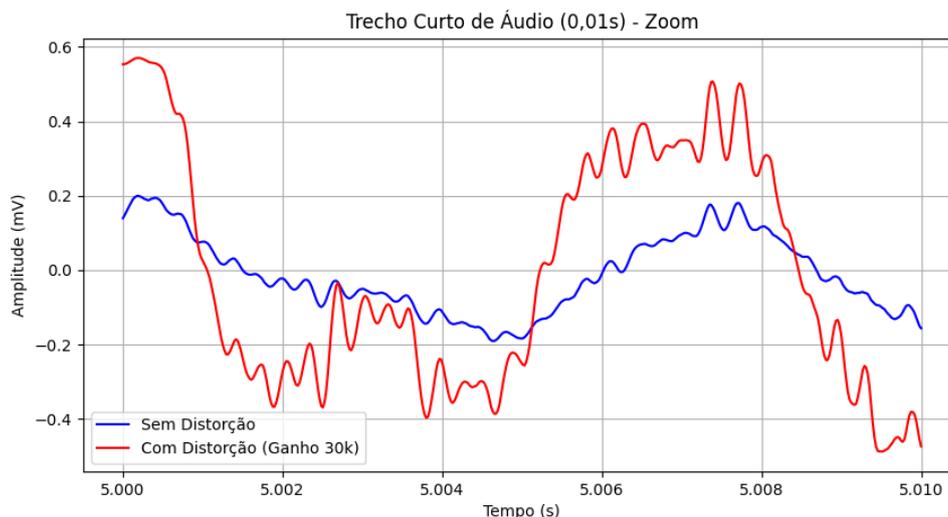


Figura 2. Comparação entre o áudio com distorção e o áudio sem distorção.

O áudio foi então carregado em um tensor para preservar sua estrutura temporal e permitir a manipulação eficiente dos dados durante o treinamento e avaliação da MLP. Cada par de vozes correspondentes foi usado como entrada e rótulo durante o treinamento da rede. Tal abordagem à preparação de dados garantiu que a MLP fosse treinada e avaliada em condições realistas, refletindo com precisão os desafios colocados pelas distorções nos sinais da guitarra elétrica.

3.2. Arquitetura da Perceptron Multicamadas (MLP)

A arquitetura da MLP desempenha um papel fundamental na modelagem de efeitos de distorção em sinais de guitarra elétrica. A perceptron multicamadas foi projetada para aprender representações complexas de sinais de áudio, permitindo a reprodução de efeitos de distorção realistas. Esta seção detalha a arquitetura MLP usada neste estudo.

A MLP utilizada neste projeto consiste em três camadas principais: uma camada de entrada, duas camadas ocultas e uma camada de saída. A camada de entrada é responsável por receber o sinal de áudio bruto, as camadas ocultas aprendem uma representação intermediária do sinal. A camada de saída produz uma representação estimada do sinal de áudio distorcido.

Aqui estão os detalhes das camadas da MLP:

- **Camada de Entrada:** A camada de entrada possui 1.440.000 neurônios, correspondendo ao número de amostras dos áudios em cada entrada. Isso ocorre devido à frequência de amostragem de 48 kHz e ao uso dos áudios de entrada de 30 segundos. A frequência de amostragem determina o número de amostras por segundo,

resultando em $48.000 \text{ amostras/segundos} \times 30 \text{ segundos} = 1.440.000$ amostras no total por áudio com cada neurônio recebendo uma amostra de cada áudio, ou seja, duas amostras por neurônio.

- **Camadas Ocultas:** A MLP possui duas camadas ocultas, cada uma com 64 neurônios. As camadas ocultas são responsáveis por aprender representações intermediárias dos sinais de áudio. A função de ativação da unidade de retificação unificada, traduzida da palavra *rectified linear unit* (ReLU), é aplicada após cada camada oculta para introduzir não linearidades na rede.
- **Camada de Saída:** A camada de saída possui 1.440.000 neurônios, correspondendo ao número de amostras do áudio na saída. Ela produz a representação estimada do sinal de áudio distorcido.

A função de ativação ReLU é amplamente utilizada nas camadas ocultas da MLP. A função ReLU é definida matematicamente como $f(x) = \max(0, x)$.

A função ReLU mapeia qualquer valor negativo para zero, mantendo os valores positivos inalterados. Isso introduz não linearidades nas camadas ocultas, permitindo que a MLP aprenda relações não lineares complexas nos dados de entrada. A arquitetura da MLP, com suas camadas de entrada, camadas ocultas e camada de saída, juntamente com a função de ativação ReLU, executa um papel crucial na capacidade da rede de aprender e reproduzir efeitos de distorção em sinais de guitarra elétrica.

3.3. Treinamento da MLP

O treinamento da MLP (Multilayer Perceptron) desempenhou um papel fundamental neste projeto, permitindo que a rede neural modelasse os efeitos da distorção em um sinal de guitarra elétrica. Nesta seção apresentam-se os principais aspectos deste processo de treinamento. O otimizador Adam [Kingma and Ba 2014] foi escolhido para otimizar os parâmetros MLP durante o treinamento. O otimizador Adam é amplamente utilizado no treinamento de redes neurais devido à sua eficiência e capacidade de lidar com problemas complexos de otimização.

Além disso, a função de perda escolhida foi a função de perda EQM. A função de perda EQM é adequada para problemas de regressão, como o objetivo deste projeto, em que se deseja minimizar a diferença entre a saída MLP e o áudio de referência distorcido de um pedal de guitarra. O treinamento MLP foi conduzido em 100 épocas. Este número de épocas foi determinado após observação do comportamento da rede durante o treinamento e convergência do modelo.

3.4. Avaliação dos Resultados

A avaliação dos resultados obtidos neste projeto foi realizada por meio de métricas quantitativas, visando mensurar a qualidade das previsões da MLP e sua similaridade com a fala de referência, apresentadas na seção 2.4.

4. Simulações e Resultados

Esta seção apresenta os resultados de simulações realizadas para avaliar a capacidade do modelo MLP em modelar a distorção de áudio de um sinal de guitarra elétrica. Foi imprescindível utilizar-se de rigorosas técnicas de treinamento e avaliação para garantir a qualidade dos resultados.

4.1. Treinamento do Modelo

O modelo MLP foi treinado por 100 épocas usando o otimizador Adam [Kingma and Ba 2014], com função de perda EQM (Erro Quadrático Médio). Foram utilizadas técnicas de cálculo de precisão para avaliar o desempenho, medindo a proximidade das previsões do modelo aos valores reais com uma tolerância específica, dessa forma estabelecendo o cálculo da acurácia. Durante o treinamento, foi implementado um agendador que ajustou a taxa de aprendizado entre épocas para melhorar a convergência, conforme ilustrado na Figura 3, que mostra a perda e a precisão ao longo das épocas.

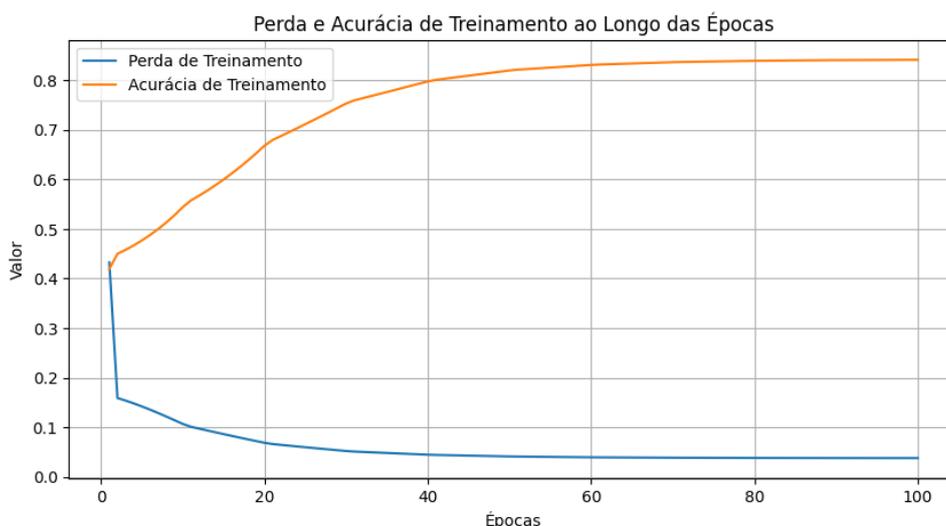


Figura 3. Perda e Acurácia de Treinamento ao Longo das Épocas.

4.2. Avaliação do Modelo

Para avaliar o desempenho do modelo treinado, a função de perda EQM foi calculada entre a saída MLP e o áudio de entrada distorcido. A EQM foi calculada conforme descrito na Equação 1.

Além disso, foi realizado o Teste Kolmogorov-Smirnov (KS) para comparar a distribuição da saída MLP e a distribuição do áudio distorcido. O teste KS é uma medida estatística que quantifica a semelhança entre duas distribuições de dados. A estatística KS e o valor p foram calculados e os resultados foram os seguintes:

- Teste de Kolmogorov-Smirnov (KS) - Estatística: 0,0279
- Teste de Kolmogorov-Smirnov (KS) - Valor p: 0,0

Pelo fato do valor de p ser menor que o valor significativo de 0.05, o que sugere uma diferença entre as duas distribuições, este escopo buscou uma possível explicação. Ao gerar uma função de distribuição acumulada, notou-se uma diferenciação para os maiores valores de amplitude em módulo. A Figura 4 mostra a curva função de distribuição acumulada.

4.3. Resultados do Erro Quadrático Médio (EQM)

O Erro Quadrático Médio (EQM) foi calculado como uma medida da diferença entre a saída MLP e o áudio distorcido. Aqui estão os resultados obtidos:

- Erro Quadrático Médio (EQM): 0,0002564572

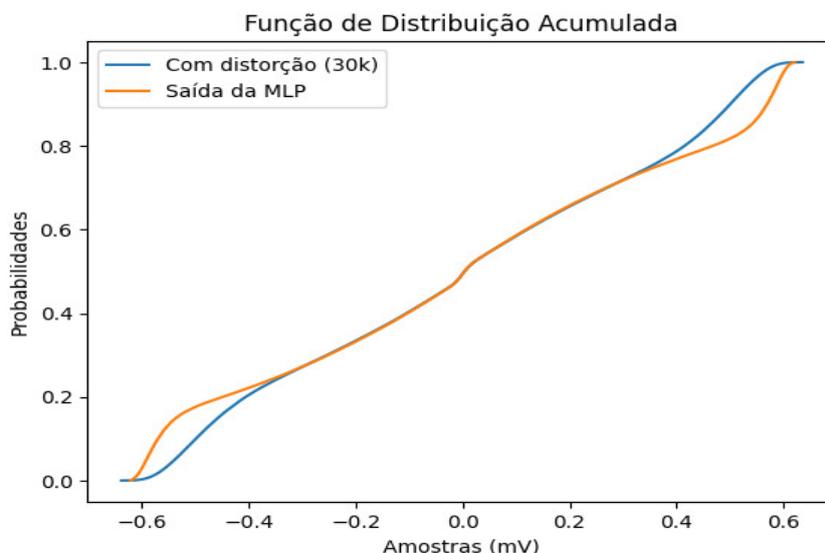


Figura 4. Gráfico da Função de Distribuição Acumulada.

4.4. Comparação entre Saída da MLP e Áudios com Distorção

Observando a Figura 5, pode-se notar que a saída da MLP (mostrada em verde) é muito semelhante ao áudio distorcido (mostrado em vermelho). Isso indica que o modelo foi capaz de modelar efetivamente as distorções presentes e no fornecimento de uma saída que se assemelhe ao sinal com distorção de referência.

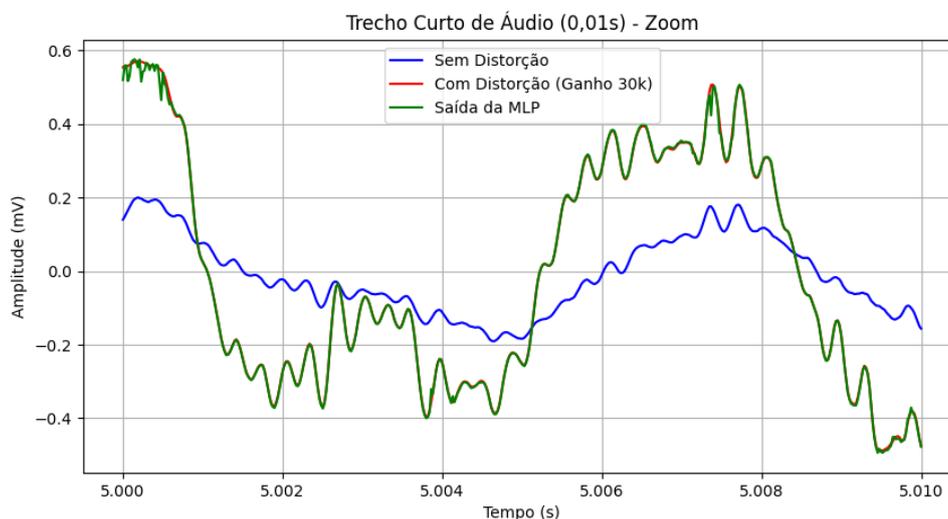


Figura 5. Comparação entre a saída da MLP, o áudio com distorção e o áudio sem distorção.

5. Conclusão

Este estudo investigou a modelagem de efeitos de distorção em sinais de guitarra elétrica utilizando-se de uma rede MLP (*Multilayer Perceptron*). Ao longo do projeto, foi abordado diversos aspectos que vão desde a teoria da distorção de áudio até a implementação da arquitetura MLP e avaliação dos resultados. Os experimentos mostraram que a MLP

pode reproduzir o efeito de distorção dos sinais da guitarra elétrica com resultado bastante aproximado do esperado baseando-se na comparação do resultado gerado pela MLP com o de referência dado na Figura 5 e através da análise do erro quadrático médio (EQM) cujo valor bem próximo de zero demonstra a fidelidade do sinal gerado pela MLP com o de referência.

Como foi utilizado um método de cálculo de precisão que leva em consideração a proximidade da saída MLP prevista e o valor real com uma tolerância definida, isso permitiu avaliar de forma mais realista a capacidade da rede de reproduzir os efeitos desejados. Ao treinar MLP, recomenda-se a utilização de técnicas como o otimizador Adam e a função de perda EQM. Um total de 100 épocas foram usadas para treinamento, e o progresso foi monitorado usando gráficos de perda e precisão entre épocas.

Este trabalho representa um avanço significativo na modelagem de efeitos de distorção em áudio e tem aplicações potenciais na indústria musical e na produção de áudio. Pesquisas futuras podem investigar mais as capacidades das diversas inteligências artificiais na criação e simulação de efeitos de áudio complexos bem como a avaliação da metodologia em outros sistemas dinâmicos não lineares. Em resumo, a abordagem utilizando redes MLP demonstra o potencial para reproduzir efeitos de distorção em guitarras elétricas e destaca as aplicações potenciais destas técnicas em aplicações criativas e industriais.

Referências

- Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.
- Devices, A. (2023). *LTspice*. <https://www.analog.com/en/design-center/design-tools-and-calculators/lts Spice-simulator.html>. Acesso em 7 de outubro de 2023.
- Engel, J., Resnick, C., Roberts, A., Dieleman, S., Norouzi, M., Eck, D., and Simonyan, K. (2017). Neural audio synthesis of musical notes with wavenet autoencoders. In *International Conference on Machine Learning*, pages 1068–1077. PMLR.
- Faceli, K., Lorena, A. C., Gama, J., Almeida, T. A. d., and Carvalho, A. C. P. d. L. F. d. (2021). *Inteligência artificial: uma abordagem de aprendizado de máquina*.
- Jones, C. (2018). Music and culture professor c. armstrong 19 december 2018.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S.-Y., and Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 13(2):206–219.
- Reiss, J. D. and McPherson, A. (2014). *Audio effects: theory, implementation and application*. CRC Press.
- Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American statistical Association*, 69(347):730–737.
- Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2023). *Dive into Deep Learning*. Cambridge University Press. <https://D2L.ai>.