

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Alvaro Domingues de Freitas

Identificação de Timbres utilizando o Envelope de Ondas Sonoras

Juiz de Fora

2026

Alvaro Domingues de Freitas

Identificação de Timbres utilizando o Envelope de Ondas Sonoras

Trabalho de conclusão de curso apresentado ao corpo docente do curso de Sistemas de Informação da Universidade Federal de Juiz de Fora como requisito parcial para a obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Marcos de Mendonça Passini

Juiz de Fora

2026

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Domingues de Freitas, Alvaro.

Identificação de Timbres utilizando o Envelope de Ondas Sonoras /
Alvaro Domingues de Freitas. – 2026.

29 f. : il.

Orientador: Marcos de Mendonça Passini

Trabalho de Conclusão de Curso (graduação) – Universidade Federal de
Juiz de Fora, Instituto de Ciências Exatas. Bacharelado em Sistemas de
Informação, 2026.

1. Processamento de Áudio. 2. Timbre. 3. ADSR. 4. Envelope Sonoro.
I. Passini, Marcos de Mendonça, orient. II. Título.

Alvaro Domingues de Freitas

Identificação de Timbres utilizando o Envelope de Ondas Sonoras

Trabalho de conclusão de curso apresentado ao corpo docente do curso de Sistemas de Informação da Universidade Federal de Juiz de Fora como requisito parcial para a obtenção do grau de Bacharel em Sistemas de Informação.

Aprovada em: 20 de Janeiro de 2026.

BANCA EXAMINADORA

Prof. Dr. Marcos de Mendonça Passini - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Rafael Alves Bonfim de Queiroz
Universidade Federal de Ouro Preto

Me. Alexandre Zuchi Carvalho
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Do momento da minha entrada na universidade até agora, passei por muitas coisas, boas e ruins. Pessoas chegaram, outras foram embora. Eu ri, e também chorei. Fracassei. Fracassei muito. Caí tantas vezes que perdi a conta. E, ainda assim, quando eu já não queria mais nada e não via saída, percebi que havia pessoas que nunca iriam desistir de mim, mesmo quando eu já tivesse desistido. Esse é o meu agradecimento formal a vocês.

À Graziella e ao Aurélio, por serem pais que eu nunca vou merecer o suficiente. Vou me lembrar de cada madrugada esperando meu pai chegar do trabalho, e de ele, mesmo depois de horas e horas de um serviço desgastante, ter tempo de sorrir, me abraçar, tomar café e assistir a Vôlei ou Fórmula 1 comigo. Vou me lembrar de cada conselho da minha mãe, exagerado ou não. Não vou esquecer, também, de quando você me pediu desculpas por não podermos lanchar depois do PISM porque tínhamos que pagar a gasolina, mesmo que tenha sido você quem sacrificou tanto para que eu pudesse chegar até a universidade naquele dia. Fui privilegiado, abençoado e amado indefinidamente, desde o momento em que nasci, justamente por ter vocês na vida.

À Daniella, minha irmã, que me aconselhou e me guiou, mesmo que em alguns casos me dando exemplo do que NÃO fazer, baderneira que sempre foi. Ver as pessoas falando que queriam ter com o irmão a relação que eu tenho com você me deixa muito feliz, e espero que seja assim para sempre.

Ao João Pedro, meu irmão. Sim, irmão, porque, se eu o vejo da forma que vejo a Dani, acho justo chamá-lo assim. Só fui entender que “sangue não é tudo” depois de te conhecer. Você me viu no meu ponto mais baixo e se recusou a soltar minha mão. Esqueça o sangue. Devo minha vida a você, hoje e sempre.

Às meninas que entraram na minha vida no momento em que a única coisa na qual se interessavam era rir de mim e colar nos meus trabalhos. As noites que passei discutindo sobre boas *novels* com a Brenda, e também quando me juntava com a Mariana para tirá-la do sério, e os dias quando chegávamos às 6:50 e a única pessoa tão enérgica quanto eu era a Beatriz, foram os melhores. Não posso deixar de citar talvez a mais importante: à Camila, por ter sido quem me fez crescer, me mostrar o que podia ser mesmo quando eu não acreditava. Depois de tudo, é estranho imaginar que o seu parágrafo sobre mim acabou sendo mais doce que o meu sobre você. Mas eu nunca vou me arrepender de ter você na minha vida. Obrigado por tudo. Ainda torço muito por você. E pelas meninas também, ok? Sem ciúmes!

Ao Arthur, Torres, Wesley, Wallyson, Yuri, Igor, Biel e Deivin. Um mais maluco que o outro, do jeito que eu sempre sonhei que meus amigos fossem. As nossas palhaçadas, chamadas de 10 horas no Discord e cachaçadas são os pontos altos da minha vida, onde eu mais me diverti, onde eu mais pude ser eu mesmo. Muitos dos meus piores dias foram

salvos por vocês. Como o Biel já diz: “De sempre, para sempre”. ÍDOLOS!

Finalmente, agradecer ao Marcos Passini, quem vai ler isso primeiro, e provavelmente por último. Ensinou-me muitas coisas, muitas mesmo, não só em sala de aula. Deu-me conselhos e contou-me experiências que abriram meus olhos. Teve uma paciência homérica em me explicar a mesma coisa 40 vezes, já que eu sou um cabeçudo. E também foi quem me lembrou do que eu gosto de fazer. Aqueles 40 minutos mostrando minha pesquisa na Sala 209 não dão nem para chamar de aula, mas foram o suficiente para me lembrar do quanto eu queria ser pesquisador, do quanto eu queria ensinar também. Se, quando eu chegar lá, for 50% do que você é, já serei um dos melhores.

“Não vou mais fugir dos meus problemas. A partir de amanhã, vou olhar todos nos olhos. A partir de amanhã, vou ouvir quando as pessoas falarem comigo. A partir de amanhã... vou fazer as coisas do jeito certo.”

(Shoya Ishida em A Voz do Silêncio)

RESUMO

O timbre é um atributo multidimensional cuja análise puramente espectral muitas vezes obscurece a dinâmica temporal dos transientes, decisiva para a identificação da fonte sonora. Neste contexto, este trabalho foca na análise do modelo ADSR (*Attack, Decay, Sustain, Release*) do envelope de amplitude, explorando sistematicamente formulações matemáticas para sua construção. A metodologia proposta envolve a extração do envelope via Transformada de Hilbert e sua parametrização através de uma competição de modelos ajustados por Mínimos Quadrados, permitindo a criação de perfis de amplitude temporal para cada som. Estes perfis foram utilizados para treinar um classificador *Random Forest*, que alcançou uma acurácia global de 81,0% na distinção entre famílias de instrumentos. Os resultados demonstram que o envelope de amplitude, especialmente através dos coeficientes de curvatura das fases, oferece informações distintivas robustas, validando sua eficácia para a diferenciação de fontes sonoras e contribuindo para pesquisas futuras em Recuperação de Informação Musical (MIR) e síntese sonora.

Palavras-chave: timbre; envelope sonoro; ADSR; processamento de áudio.

ABSTRACT

Timbre is a multidimensional attribute whose purely spectral analysis often obscures the temporal dynamics of transients, which are decisive for sound source identification. In this context, this work focuses on the analysis of the ADSR (Attack, Decay, Sustain, Release) amplitude envelope model, systematically exploring mathematical formulations for its construction. The proposed methodology involves envelope extraction via the Hilbert Transform and its parameterization through a competition of models fitted by Least Squares, enabling the creation of temporal amplitude profiles for each sound. These profiles were used to train a Random Forest classifier, which achieved a global accuracy of 81.0% in distinguishing between instrument families. The results demonstrate that the amplitude envelope, especially through the phase curvature coefficients, offers robust distinctive information, validating its effectiveness for the differentiation of sound sources and contributing to future research in Music Information Retrieval (MIR) and sound synthesis.

Keywords: timbre; sound envelope; ADSR; audio processing.

LISTA DE FIGURAS

Figura 1 – Visualização de envelopes de amplitude: 1-Tabla 2-Trompa e 3-Flauta	15
Figura 2 – Representação gráfica do modelo ADSR clássico.	16
Figura 3 – Três tipos de envelopes de amplitude: (a) Baixo, (b) Teclado e (c) Pads de Percussão.	17
Figura 4 – Exemplo de reta ajustada através do Método dos Mínimos Quadrados (MMQ).	19
Figura 5 – Distribuição das amostras por família, evidenciando a predominância de percussão na base de dados.	23
Figura 6 – Matriz de Confusão do classificador Random Forest (Acurácia: 81,0%).	24
Figura 7 – Gráfico definindo a importância de cada <i>feature</i> no modelo de classificação.	25

LISTA DE ABREVIATURAS E SIGLAS

ADSR	<i>Attack, Decay, Sustain, Release</i>
ANSI	<i>American National Standards Institute</i>
MIR	<i>Music Information Retrieval</i>
MMQ	Método dos Mínimos Quadrados
PDS	Processamento Digital de Sinais
SNR	<i>Signal-to-Noise Ratio</i>
VSCO	<i>Versilian Studios Chamber Orchestra</i>
WAV	<i>Waveform Audio File Format</i>

SUMÁRIO

1	INTRODUÇÃO	11
1.1	CONTEXTUALIZAÇÃO	11
1.2	O PROBLEMA DO TIMBRE	11
1.3	OBJETIVOS	12
1.3.1	Objetivo Geral	12
1.3.2	Objetivos Específicos	12
1.4	JUSTIFICATIVA	13
1.5	ESTRUTURA DO TRABALHO	13
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	ACÚSTICA E PERCEPÇÃO SONORA	14
2.2	O ENVELOPE DE AMPLITUDE	15
2.3	O MODELO ADSR	15
2.4	MODELAGEM MATEMÁTICA DO ENVELOPE	17
2.5	PROCESSAMENTO DIGITAL DE SINAIS	17
2.5.1	Sinais Discretos e Amostragem	17
2.5.2	Transformada de Hilbert e Envelope Analítico	18
2.5.3	Ajuste de Curvas via Mínimos Quadrados (MMQ)	18
3	METODOLOGIA	20
3.1	BASE DE DADOS E PRÉ-PROCESSAMENTO	20
3.2	ALGORITMO DE EXTRAÇÃO DE ENVELOPE	20
3.3	SEGMENTAÇÃO AUTOMÁTICA VIA DERIVADAS	21
3.4	MODELAGEM HÍBRIDA E SELEÇÃO DE MODELOS	21
3.5	CLASSIFICAÇÃO TÍMBRICA	22
4	RESULTADOS E DISCUSSÃO	23
4.1	ANÁLISE DO CONJUNTO DE DADOS	23
4.2	DESEMPENHO DO CLASSIFICADOR	23
4.3	ANÁLISE DA MATRIZ DE CONFUSÃO	23
4.4	IMPORTÂNCIA DAS CARACTERÍSTICAS	25
5	CONSIDERAÇÕES FINAIS	27
5.1	LIMITAÇÕES	27
5.2	CONTRIBUIÇÕES	27
5.3	TRABALHOS FUTUROS	27
	REFERÊNCIAS	29

1 INTRODUÇÃO

O som é um elemento intrínseco à experiência humana, servindo como meio fundamental de comunicação, expressão artística e percepção do ambiente. Desde os primórdios, a humanidade busca não apenas produzir sons, mas também compreendê-los e manipulá-los. Com o avanço da tecnologia digital, o processamento de áudio tornou-se uma área vasta de pesquisa, permitindo que máquinas analisem e sintetizem sinais acústicos com precisão crescente (Müller, 2015).

No universo da música, a capacidade de discernir entre diferentes fontes sonoras é essencial. Um piano e um violino, por exemplo, podem executar a mesma nota (mesma frequência fundamental) na mesma intensidade (amplitude), e ainda assim serem percebidos como distintos. Essa distinção é possível graças ao timbre, uma qualidade multidimensional do som que confere sua textura característica, muitas vezes descrita como a "cor" do som (Grey, 1977).

Esta monografia insere-se no contexto da Computação Musical e do Processamento Digital de Sinais, propondo uma abordagem analítica para a identificação de timbres através da modelagem de seus envelopes temporais.

1.1 CONTEXTUALIZAÇÃO

A análise de sinais de áudio envolve a extração de características que permitam descrever o sinal de forma compacta e informativa. Enquanto a altura e a intensidade são grandezas físicas bem definidas — relacionadas respectivamente à frequência e à amplitude da onda —, o timbre é um atributo psicofísico mais complexo.

Tradicionalmente, a síntese sonora utiliza modelos matemáticos para criar sons artificiais. Um dos modelos mais clássicos e difundidos na síntese é o envelope ADSR (Attack, Decay, Sustain, Release), que descreve a evolução da amplitude do som ao longo do tempo em quatro estágios distintos (Roads, 1996). Embora o ADSR tenha sido concebido para a *síntese* (criação) de sons, sua aplicação na *análise* (reconhecimento) de instrumentos acústicos reais apresenta um campo fértil para investigação. A hipótese central é que o perfil temporal de amplitude — capturado pelo envelope — carrega informações suficientes para segregar categorias de instrumentos, desde que extraído e parametrizado corretamente através de métodos numéricos robustos.

1.2 O PROBLEMA DO TIMBRE

A American National Standards Institute (1994) define formalmente o timbre como o atributo da sensação auditiva que permite a um ouvinte julgar que dois sons, apresentados de maneira similar e com mesma intensidade e altura, são diferentes.

A identificação automática de timbres é, portanto, um desafio de reconhecimento de padrões. O problema reside na alta variabilidade dos sinais acústicos naturais e na necessidade de reduzir a dimensionalidade dos dados brutos de áudio para um conjunto de parâmetros tratáveis (Peeters *et al.*, 2011).

É importante ressaltar que a análise por envelope sonoro (Domínio do Tempo) e a análise espectral via Transformada de Fourier (Domínio da Frequência) oferecem perspectivas ortogonais e complementares sobre o timbre (Oppenheim e Schafer, 2009). Enquanto a Transformada de Fourier decompõe o sinal em seus componentes harmônicos constitutivos, revelando a cor ou textura estática do som (estrutura espectral), ela frequentemente obscurece a informação sobre a evolução temporal dos transientes.

Em contrapartida, a modelagem do envelope captura a dinâmica da excitação e a dissipação de energia do sistema físico (o gesto sonoro), mas ignora o conteúdo freqüencial. Instrumentos como o violino e a flauta podem apresentar ataques suaves similares (confusão no domínio do tempo), mas espectros distintos; inversamente, um piano e uma guitarra podem compartilhar conteúdo harmônico, mas diferem drasticamente em seus envelopes de decaimento.

Neste trabalho, o problema específico abordado é a eficácia da representação simplificada do envelope sonoro pelo modelo ADSR na tarefa de distinção tímbrica. Testa-se, portanto, se a segmentação de um envelope complexo (extraído, por exemplo, via Transformada de Hilbert) em um modelo de funções, ajustadas via Mínimos Quadrados, preserva a identidade do instrumento original.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Diante da relevância do timbre na diferenciação de sons, o objetivo principal deste trabalho é investigar a eficácia da utilização do envelope de amplitude de ondas sonoras, parametrizado pelo modelo ADSR, como característica distintiva para a identificação e classificação de timbres instrumentais.

1.3.2 Objetivos Específicos

Para alcançar o objetivo geral, foram definidos os seguintes objetivos específicos:

- Revisar a literatura sobre a física do som, processamento digital de sinais e o modelo ADSR;
- Formalizar matematicamente o envelope ADSR para determinar a melhor aderência aos sinais acústicos reais;

- Desenvolver e implementar uma metodologia computacional para a extração do envelope analítico do sinal (utilizando a Transformada de Hilbert) e a subsequente segmentação automática nos estágios ADSR (utilizando métodos de otimização e Mínimos Quadrados);
- Avaliar a capacidade de discriminação dos parâmetros extraídos ao aplicá-los em um conjunto de amostras de diferentes instrumentos musicais.

1.4 JUSTIFICATIVA

A relevância deste estudo é dupla: acadêmica e prática. Do ponto de vista acadêmico, contribui para a área de Recuperação de Informação Musical (Music Information Retrieval - MIR) ao explorar métodos de segmentação de sinal e ajuste de curvas em dados reais, testando os limites de modelos teóricos de síntese quando aplicados à análise.

Na prática, algoritmos eficientes de identificação de timbre têm aplicações diretas em transcrição automática de música, indexação de bancos de dados de áudio, melhoria de sistemas de compressão de áudio e desenvolvimento de novos sintetizadores que aprendem com amostras reais. Além disso, o estudo aprofundado das formulações matemáticas (linear versus exponencial) e dos algoritmos de ajuste (como Mínimos Quadrados) fornece um entendimento crítico sobre como simplificações matemáticas afetam a representação da realidade física.

1.5 ESTRUTURA DO TRABALHO

Este trabalho está organizado em cinco capítulos, conforme descrito a seguir:

- **Capítulo 1 - Introdução:** Apresenta o contexto, a problematização, os objetivos e a justificativa da pesquisa.
- **Capítulo 2 - Fundamentação Teórica:** Aborda os conceitos fundamentais de acústica, o modelo ADSR, e as ferramentas matemáticas utilizadas, como a Transformada de Hilbert e o Método dos Mínimos Quadrados.
- **Capítulo 3 - Metodologia:** Detalha o processo de desenvolvimento do algoritmo de extração e segmentação do envelope, explicando as decisões de implementação e as ferramentas de software utilizadas.
- **Capítulo 4 - Resultados e Discussão:** Apresenta os dados obtidos a partir da aplicação da metodologia em amostras de áudio, discutindo a eficácia da parametrização ADSR na diferenciação dos timbres.
- **Capítulo 5 - Conclusão:** Sintetiza as descobertas do trabalho, aponta limitações encontradas e sugere trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

A análise de sinais de áudio visando a identificação e classificação de instrumentos musicais demanda uma abordagem multidisciplinar, exigindo a convergência entre a acústica física, a psicoacústica e o processamento digital de sinais. O timbre, objeto central deste estudo, não é uma grandeza escalar simples, mas sim um atributo multidimensional e complexo. Ao contrário da frequência ou da amplitude, que podem ser mapeadas diretamente para as sensações de altura e intensidade, o timbre é determinado por uma intrincada combinação de fatores espectrais (conteúdo harmônico) e temporais (dinâmica da amplitude).

Historicamente, a teoria clássica, baseada nos trabalhos de Helmholtz, sugeria que o espectro estacionário — a composição de harmônicos e suas amplitudes relativas — era o determinante primário da cor do som Benson (2006). Entretanto, avanços na Psicoacústica durante o século XX demonstraram que essa visão era incompleta. A evolução temporal do som, especificamente a forma como a amplitude se comporta desde o início da vibração até sua extinção, conhecida como envelope de amplitude, desempenha um papel determinante na identificação da fonte sonora.

Este capítulo estabelece o arcabouço teórico para a modelagem desse envelope, transitando dos princípios da acústica até os algoritmos de otimização matemática utilizados para a extração de parâmetros.

2.1 ACÚSTICA E PERCEPÇÃO SONORA

Do ponto de vista físico, o som é uma perturbação mecânica que se propaga em meios elásticos, como o ar ou a água, através de ondas longitudinais de pressão. Para que um som musical seja produzido, é necessário um sistema vibratório (corda, coluna de ar, membrana) e um mecanismo de excitação.

As três grandezas físicas fundamentais de uma onda sonora correlacionam-se com atributos perceptivos específicos:

- **Frequência (f):** Grandeza física medida em Hertz (Hz). Sua contraparte perceptiva é a **Altura**, que define a sensação de notas graves ou agudas. É determinada fundamentalmente pela frequência de vibração primária do corpo sonoro (f_0).
- **Amplitude (A):** Relacionada à energia ou pressão da onda. Sua contraparte perceptiva é a **Intensidade**, popularmente referida como volume. A relação entre amplitude física e intensidade percebida é logarítmica, descrita pela Lei de Weber-Fechner Silverthorn (2017).

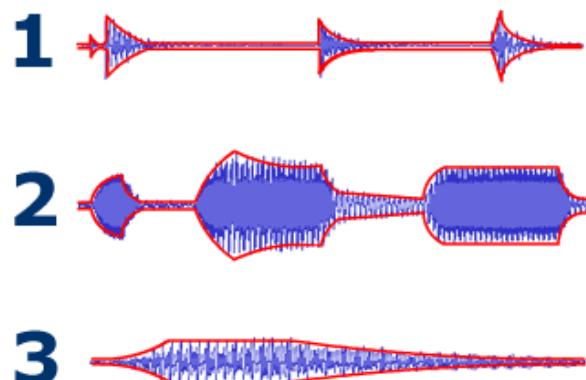
- **Forma de Onda / Espectro:** A complexidade da vibração. Sua contraparte é o **Timbre**, a qualidade que confere identidade à fonte sonora.

A definição de timbre pela *American National Standards Institute* (ANSI) é feita por exclusão: é o atributo da sensação auditiva que permite a um ouvinte julgar que dois sons dissimilares, possuindo a mesma intensidade e altura, são diferentes. Essa definição, embora correta, é vaga quanto à causa. A pesquisa moderna entende o timbre como uma assinatura espectro-temporal, onde os transientes de ataque (o início do som) carregam uma quantidade desproporcional de informação sobre a identidade do instrumento.

2.2 O ENVELOPE DE AMPLITUDE

O envelope de amplitude é a curva que descreve a variação temporal da intensidade de um sinal sonoro. Enquanto a forma de onda contém as oscilações rápidas que determinam a frequência, o envelope é a envoltória que contorna esses picos, modulando o som ao longo do tempo.

Figura 1 – Visualização de envelopes de amplitude: 1-Tabla 2-Trompa e 3-Flauta



Fonte: Adaptado de Wikipedia (2005).

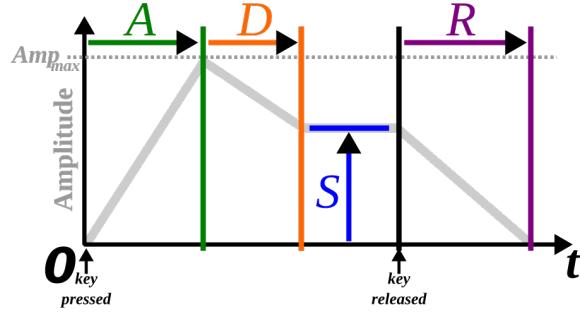
A importância do envelope é evidenciada em experimentos de síntese e edição: ao remover o ataque inicial (os primeiros milissegundos) de uma gravação de piano, o som resultante torna-se difícil de distinguir de uma flauta ou órgão. Isso ocorre porque o envelope temporal dita a dinâmica da energia: instrumentos percussivos liberam energia de forma explosiva, enquanto instrumentos de sopro ou arco podem sustentar ou aumentar a energia gradualmente.

2.3 O MODELO ADSR

Para fins de síntese e análise computacional, o envelope natural — que pode ser complexo e irregular — é frequentemente simplificado pelo modelo **ADSR** (*Attack, Decay*,

Sustain, Release). Este modelo, padronizado inicialmente pela indústria de sintetizadores na década de 1970, divide a vida de uma nota musical em quatro estágios distintos (Beauchamp, 2007).

Figura 2 – Representação gráfica do modelo ADSR clássico.



Fonte: Adaptado de Wikimidia (2007).

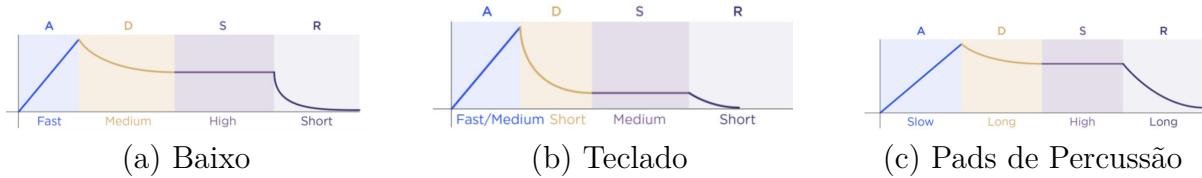
Cada fase representa um comportamento físico e perceptivo:

1. **Ataque (Attack - T_A)**: O intervalo de tempo necessário para que a amplitude suba do zero até o seu pico máximo.
 - *Física*: Representa a inércia do sistema sendo vencida pela força de excitação. Em um piano (percussivo), o ataque é quase instantâneo (< 10 ms). Em um violoncelo (friccionado), o ataque é lento (> 100 ms), pois a corda precisa entrar em ressonância gradualmente.
2. **Decaimento (Decay - T_D)**: O tempo que a amplitude leva para cair do pico máximo até o nível de sustentação.
 - *Física*: Após o choque inicial, a energia excedente se dissipada rapidamente ou se estabiliza em um modo de vibração sustentável.
3. **Sustentação (Sustain - L_S)**: Diferentemente das outras fases que são medidas em tempo, o Sustain é um **nível** de amplitude (geralmente expresso como uma porcentagem do pico).
 - *Física*: Corresponde ao equilíbrio entre a energia fornecida pelo músico e a energia dissipada pelo instrumento. Instrumentos de excitação contínua (sopros, arco, órgãos) possuem $L_S > 0$. Instrumentos de excitação impulsiva (piano, violão, percussão) teoricamente possuem $L_S \approx 0$ ou um decaimento contínuo, pois não há injeção de energia nova após o ataque.
4. **Liberação (Release - T_R)**: O tempo para a amplitude cair do nível de sustentação (ou do nível atual) até o silêncio, após o músico encerrar a nota.

- *Física*: Representa o amortecimento natural do sistema. O som não para instantaneamente; a caixa de ressonância e o ambiente reverberante continuam vibrando por um breve período.

A comparação visual entre envelopes de diferentes famílias de instrumentos ilustra a capacidade discriminatória deste modelo.

Figura 3 – Três tipos de envelopes de amplitude: (a) Baixo, (b) Teclado e (c) Pads de Percussão.



Fonte: Adaptado de Drew Swicher (2019).

2.4 MODELAGEM MATEMÁTICA DO ENVELOPE

Enquanto sintetizadores digitais antigos utilizavam segmentos de reta para gerar o ADSR (Modelo Linear Piecewise), a acústica de instrumentos reais obedece a leis de decaimento exponencial, governadas pela dissipação de energia devido ao atrito e resistência do ar.

Um modelo matematicamente mais preciso para as fases de decaimento e liberação é dado por funções exponenciais da forma:

$$y(t) = A_0 \cdot e^{-\frac{t}{\tau}} \quad (2.1)$$

Onde τ é a constante de tempo do sistema. Embora a aproximação linear por partes seja frequentemente utilizada como uma simplificação robusta que reduz o custo computacional (Beauchamp, 2007), um conjunto de retas faz com que o envelope se torne não diferenciável nos pontos entre as etapas da curva ADSR, o que faz com que a percepção do som soe anti-natural ao ouvido humano, cuja audição é logarítmica, portanto, seria melhor utilizar um modelo que mantivesse a suavidade nesses pontos de transição.

2.5 PROCESSAMENTO DIGITAL DE SINAIS

A aplicação prática do modelo teórico a gravações de áudio requer técnicas avançadas de Processamento Digital de Sinais (PDS).

2.5.1 Sinais Discretos e Amostragem

O sinal de áudio contínuo $x(t)$ é convertido em uma sequência discreta $x[n]$ através da amostragem em intervalos T_s . A precisão da análise temporal depende diretamente

da taxa de amostragem ($F_s = 1/T_s$), que deve satisfazer o critério de Nyquist para evitar distorções espectrais (aliasing) (Oppenheim e Schafer, 2009).

2.5.2 Transformada de Hilbert e Envelope Analítico

Extrair o envelope de um sinal oscilatório cru não é trivial, pois o valor médio do sinal tende a zero. Métodos simples como a retificação (valor absoluto) seguida de filtragem passa-baixa introduzem atrasos de fase (atraso de grupo) que podem distorcer a medição precisa do tempo de ataque.

A abordagem do **Sinal Analítico** via Transformada de Hilbert é mais adequada para esta aplicação. A Transformada de Hilbert, denotada por $\mathcal{H}\{x[n]\}$, gera um sinal que é uma versão de $x[n]$ defasada em 90° . O sinal analítico $z[n]$ é construído como um número complexo:

$$z[n] = x[n] + i\mathcal{H}\{x[n]\} \quad (2.2)$$

O envelope instantâneo $E[n]$ é, portanto, a magnitude (módulo) deste vetor complexo girante:

$$E[n] = |z[n]| = \sqrt{x[n]^2 + (\mathcal{H}\{x[n]\})^2} \quad (2.3)$$

Este método permite obter um envelope suave, que tangencia perfeitamente os picos do sinal sem introduzir atrasos temporais significativos.

2.5.3 Ajuste de Curvas via Mínimos Quadrados (MMQ)

Uma vez obtido o envelope bruto $E[n]$ via Transformada de Hilbert, o desafio principal deste trabalho é ajustar o modelo ADSR assim idealizado aos dados reais, que contêm ruído e micro-variações. Este é um problema clássico de regressão.

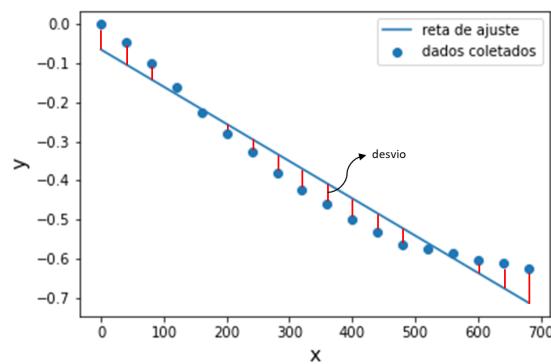
O Método dos Mínimos Quadrados (MMQ) busca encontrar os parâmetros da curva que minimizam a soma dos quadrados dos resíduos. O resíduo é a diferença vertical entre o envelope real medido e o modelo proposto.

Matematicamente, definindo o erro para cada amostra, onde y_i é o valor definido do envelope real e x_i é o valor no modelo ADSR calculado, buscamos minimizar a função de erro S :

$$S(\mathbf{c}) = \sum_{k=0}^N \left(\sum_{j=0}^P c_j \phi_j(x_k) - y_k \right)^2 \quad (2.4)$$

A aplicação do MMQ permite que o algoritmo ignore pequenas oscilações irrelevantes (ruído ou vibrato) e encontre a tendência macroscópica que melhor define o ataque ou o decaimento, resultando em parâmetros T_A ou T_D estatisticamente robustos.

Figura 4 – Exemplo de reta ajustada através do Método dos Mínimos Quadrados (MMQ).



Fonte: Adaptado de Guilherme Castilho (2023).

3 METODOLOGIA

Este capítulo detalha o fluxo de processamento desenvolvido para a extração de características tímbricas e a subsequente classificação dos instrumentos. A metodologia foi implementada utilizando a linguagem Python, com suporte das bibliotecas **Librosa** para manipulação de áudio, **SciPy** para computação científica e **Scikit-Learn** para aprendizado de máquina (Pedregosa *et al.*, 2011).

3.1 BASE DE DADOS E PRÉ-PROCESSAMENTO

Para a validação do estudo, foi utilizado um conjunto de dados composto por amostras de instrumentos orquestrais provenientes da biblioteca *VSCO 2: Community Edition* (Versilian Studios (2024)), desenvolvida pela Versilian Studios. Esta biblioteca fornece amostras de alta qualidade gravadas individualmente, sendo uma referência em bancos de dados de áudio abertos.

Esse conjunto contém amostras de variadas durações, a depender do tipo de instrumento tocado, e também são tocadas notas variadas.

A base de dados foi organizada em quatro famílias principais: Cordas, Madeiras, Metais e Percussão. Como as amostras originais já são disponibilizadas no formato WAV, não houve necessidade de conversão de *codecs* com perdas, garantindo a integridade espectral e temporal do sinal desde a aquisição.

O pré-processamento das amostras envolveu as seguintes etapas:

- **Normalização:** Todas as amostras foram normalizadas em amplitude para o intervalo $[0, 1]$, eliminando variações de volume de gravação (ganho) que poderiam enviesar o classificador, garantindo que a análise foque na forma do envelope e não na intensidade absoluta.
- **Filtragem:** Foi aplicado um filtro de suavização (média móvel) leve sobre o envelope extraído para reduzir ruídos de alta frequência e micro-variações que poderiam gerar falsos positivos na detecção automática das transições entre as fases ADSR.

3.2 ALGORITMO DE EXTRAÇÃO DE ENVELOPE

A extração do envelope de amplitude $E[n]$ foi realizada através do método do **Sinal Analítico** via Transformada de Hilbert, conforme fundamentado na Seção 2.5.2. Este método foi escolhido por sua capacidade de seguir os picos da forma de onda sem introduzir o atraso de fase característico de filtros passa-baixa tradicionais.

3.3 SEGMENTAÇÃO AUTOMÁTICA VIA DERIVADAS

Diferentemente da abordagem simplista de limiares fixos, este trabalho implementou um algoritmo de segmentação baseado na análise da taxa de variação do sinal. Como o envelope extraído $E[n]$ é uma sequência discreta, o cálculo da derivada contínua $\frac{dE}{dt}$ é aproximado numericamente pelo gradiente discreto (diferenças finitas).

O vetor de derivadas $\Delta E[n]$ foi calculado computacionalmente através da diferença central entre amostras adjacentes, conforme a aproximação:

$$\Delta E[n] \approx \frac{E[n+1] - E[n-1]}{2} \quad (3.1)$$

Essa operação transforma o perfil de amplitude em um perfil de velocidade de variação. Para mitigar a amplificação de ruído inerente ao processo de diferenciação, o envelope $E[n]$ foi previamente submetido a uma suavização por média móvel. Com base no vetor ΔE , o algoritmo segmenta as fases seguindo a lógica de transição de sinais:

1. **Ataque (T_A)**: Compreende o intervalo de crescimento rápido. O algoritmo identifica o início quando $E[n]$ supera um limiar de ruído (1% da amplitude máxima) e define o final no índice do **pico global**, ponto onde a derivada cruza o eixo zero ($\Delta E[n] = 0$) vindo de valores positivos.
2. **Decaimento (T_D)**: Inicia-se imediatamente após o pico global. Nesta fase, a derivada assume valores negativos ($\Delta E[n] < 0$). O fim do decaimento é determinado dinamicamente monitorando a magnitude da derivada; quando $|\Delta E[n]|$ cai abaixo de um limiar de tolerância ϵ (aproximando-se de zero), o sistema entende que a curva estabilizou, marcando o início da sustentação.
3. **Sustentação (L_S)**: Região de estabilidade relativa onde a variação da amplitude é mínima. Matematicamente, é o intervalo onde a derivada oscila em torno de zero, mantendo o nível de energia do instrumento.
4. **Liberação (T_R)**: Identificada quando a derivada rompe a estabilidade e assume novamente valores negativos acentuados ($\Delta E[n] \ll 0$), indicando a extinção da nota, até que a amplitude retorne ao nível de silêncio.

3.4 MODELAGEM HÍBRIDA E SELEÇÃO DE MODELOS

Para capturar a complexidade do timbre, não foi imposto um único modelo matemático. Implementou-se uma rotina de Comparação de Modelos utilizando o Método dos Mínimos Quadrados (MMQ). Para cada fase (Ataque, Decaimento, Liberação), o algoritmo ajustou diferentes funções candidatas:

- Modelo Linear ($y = at + b$);
- Modelo Exponencial ($y = Ae^{-Bt} + C$);
- Modelo Sigmoidal (Logístico).

O modelo selecionado para compor o vetor de características foi aquele que apresentou o menor Erro Quadrático Médio (MSE) em relação ao envelope real.

3.5 CLASSIFICAÇÃO TÍMBRICA

A etapa final consistiu na classificação supervisionada dos instrumentos.

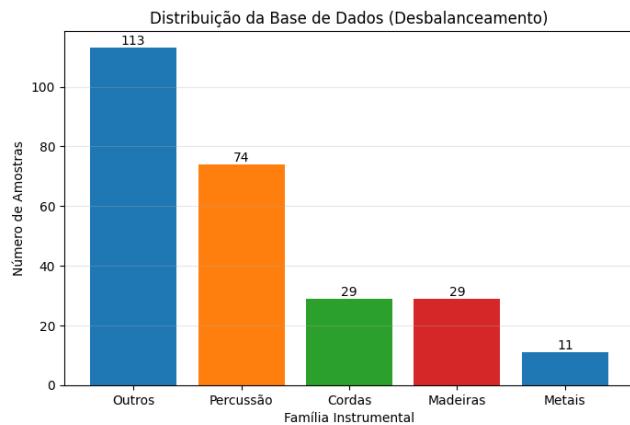
- **Vetor de Características:** Cada amostra foi representada por um vetor expandido contendo não apenas os tempos (T_A, T_D, T_R), mas também os conjuntos de coeficientes obtidos via MMQ (C_A, C_D, C_S, C_R) e a amplitude na fase de sustentação. Isso permitiu ao modelo aprender a curvatura do som, não apenas sua duração.
- **Modelo:** Utilizou-se o algoritmo *Random Forest* (Floresta Aleatória) com 100 árvores de decisão.
- **Validação:** O conjunto de dados foi dividido aleatoriamente em 70% para treinamento e 30% para teste. A avaliação de desempenho foi feita através da Acurácia Global e da Matriz de Confusão.

4 RESULTADOS E DISCUSSÃO

4.1 ANÁLISE DO CONJUNTO DE DADOS

O processo de extração de características foi aplicado a uma base de dados contendo amostras de Cordas, Madeiras, Metais e Percussão. Observou-se, porém, que parte das amostras não foram analizadas devido a dificuldade do algoritmo de identificar através da nomenclatura das amostras a qual família elas pertenciam, gerando um resultado que ignorou parte delas.

Figura 5 – Distribuição das amostras por família, evidenciando a predominância de percussão na base de dados.



Fonte: Elaborado pelo autor (2026).

4.2 DESEMPENHO DO CLASSIFICADOR

O modelo de classificação *Random Forest*, treinado com os vetores de características ADSR expandidos (tempos + coeficientes de curvatura), alcançou uma Acurácia Global de 81.0% no conjunto de teste. Este resultado valida a hipótese central do trabalho de que o envelope temporal carrega informações distintivas suficientes para a categorização instrumental.

4.3 ANÁLISE DA MATRIZ DE CONFUSÃO

A avaliação detalhada do desempenho do classificador foi realizada através da **Matriz de Confusão**. Trata-se de uma estrutura tabular que confronta as categorizações preditas pelo modelo com as classes reais (rótulos verdadeiros) das amostras de teste.

Formalmente, para um problema de classificação multiclasse com K classes distintas (neste estudo, $K = 4$: Cordas, Madeiras, Metais e Percussão), a matriz de confusão C é definida como uma matriz quadrada de dimensão $K \times K$. Cada elemento $C_{i,j}$ da matriz

representa a contagem de observações que pertencem à classe verdadeira i e que foram classificadas pelo algoritmo como pertencentes à classe j .

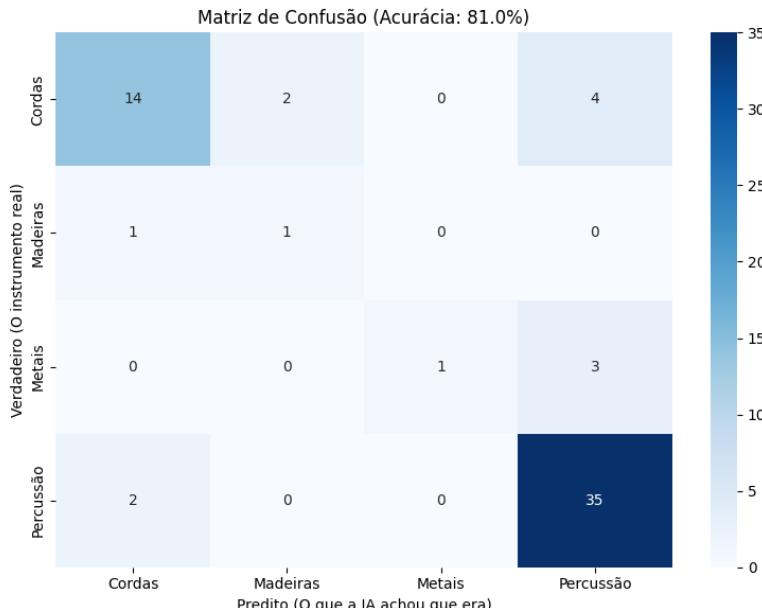
Nesta representação:

- Os elementos da **diagonal principal** ($i = j$) correspondem aos acertos do modelo (Verdadeiros Positivos para a classe i);
- Os elementos fora da diagonal ($i \neq j$) representam os erros de classificação, explicitando as confusões entre pares específicos de instrumentos (Falsos Positivos e Falsos Negativos).

A análise da matriz de confusão é superior à simples métrica de acurácia global, pois permite identificar vieses do modelo e classes específicas onde o classificador apresenta maior dificuldade de generalização (Pedregosa *et al.*, 2011).

A Figura 6 apresenta a Matriz de Confusão obtida, permitindo uma análise detalhada dos acertos e erros por classe.

Figura 6 – Matriz de Confusão do classificador Random Forest (Acurácia: 81,0%).



Fonte: Elaborado pelo autor (2026).

A análise da matriz revela comportamentos distintos entre as famílias:

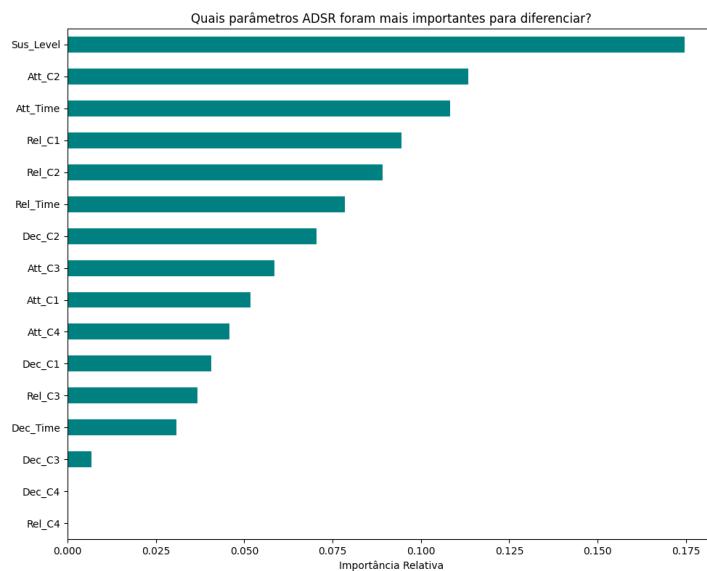
- **Percussão:** O modelo obteve desempenho excelente, identificando corretamente 35 das 37 amostras de teste. Isso se deve às características únicas do envelope percussivo: ataque extremamente rápido (quase instantâneo) e ausência de sustentação ($L_S \approx 0$), perfis que o modelo matemático capturou com facilidade.

- **Cordas:** Houve uma taxa de acerto consistente (14 amostras), com algumas confusões com Madeiras e Percussão. As confusões com percussão (4 casos) podem ser atribuídas a técnicas de execução como o *pizzicato*, que geram envelopes similares aos de instrumentos percussivos.
- **Metais e Madeiras:** Estas classes apresentaram maior dificuldade de classificação. Devido ao menor número de amostras disponíveis para treinamento, o modelo teve dificuldade em generalizar os padrões dessas famílias, confundindo-as ocasionalmente com Cordas e Percussão.

4.4 IMPORTÂNCIA DAS CARACTERÍSTICAS

A análise de importância das variáveis é uma técnica fundamental para a interpretabilidade de modelos de aprendizado de máquina, permitindo quantificar a contribuição relativa de cada atributo para o poder preditivo do modelo. No contexto do algoritmo *Random Forest*, essa métrica é calculada através da Diminuição Média da Impureza. Durante o processo de treinamento, cada vez que uma árvore de decisão seleciona uma variável para realizar uma divisão em um nó, calcula-se o quanto essa divisão reduziu a impureza das classes resultantes (geralmente medida pelo Índice Gini ou Entropia). A importância final de uma característica é determinada pela média ponderada dessas reduções de impureza acumuladas em todas as árvores que compõem a floresta.

Figura 7 – Gráfico definindo a importância de cada *feature* no modelo de classificação.



Fonte: Elaborado pelo autor (2026).

A análise resultante, apresentada na Figura 7, indicou que os coeficientes de curvatura (parâmetros da função exponencial/linear) tiveram peso significativo na decisão do classificador, frequentemente superando a importância apenas da duração absoluta

(T_A ou T_D). Isso confirma que a **forma** como o som cresce ou decai (se linearmente ou exponencialmente) é um discriminador tímbrico tão importante quanto o tempo que ele leva para fazê-lo.

5 CONSIDERAÇÕES FINAIS

Esta monografia investigou a eficácia do envelope sonoro, parametrizado via modelo ADSR, como descritor para a identificação automática de timbres. Através de uma abordagem que combinou processamento digital de sinais (Transformada de Hilbert) e métodos numéricos de otimização (Mínimos Quadrados), foi possível extrair vetores de características detalhados de amostras reais.

Os resultados obtidos, com uma acurácia de 81.1%, demonstram que o envelope temporal é, de fato, uma impressão digital robusta do instrumento. A implementação de uma comparação de modelos matemáticos provou-se acertada, permitindo ao algoritmo adaptar-se à física específica de cada instrumento, capturando desde o decaimento exponencial natural de um prato de bateria até a sustentação linear de um violino.

5.1 LIMITAÇÕES

A principal limitação encontrada residiu no balanceamento da base de dados. A menor disponibilidade de amostras de sopros (Metais e Madeiras) na biblioteca utilizada impactou a capacidade do modelo de generalizar padrões para essas classes específicas, gerando confusões com a família das Cordas.

5.2 CONTRIBUIÇÕES

O trabalho contribui ao apresentar uma metodologia completa que não apenas extrai os tempos do ADSR, mas também quantifica a curvatura das transições através dos coeficientes das funções ajustadas. A criação de um vetor de características expandido, que inclui esses coeficientes, mostrou-se vital para aumentar a precisão da classificação, fornecendo uma base sólida para futuros experimentos em Recuperação de Informação Musical (MIR).

5.3 TRABALHOS FUTUROS

Como desdobramento natural desta pesquisa, vislumbra-se a expansão significativa da base de dados, incorporando o conjunto completo da *Versilian Studios* e outras bibliotecas de alta confiabilidade para garantir o equilíbrio estatístico entre as classes. Nesse contexto, planeja-se a validação cruzada utilizando o renomado acervo da *Philharmonia Orchestra* de Londres. Olá, Rafael, Embora este banco demande um pré-processamento adicional devido ao formato nativo comprimido (MP3), sua excelência na rotulagem e a diversidade de técnicas de execução justificam sua inclusão para testar a robustez do algoritmo em cenários de Big Data musical.

Concomitantemente à expansão dos dados, pretende-se evoluir a granularidade da classificação. O objetivo é transitar da atual categorização por famílias para a identificação de instrumentos singulares, capacitando o sistema a distinguir sutilezas tímbricas entre instrumentos correlatos, como o violoncelo e a viola, ou o piano e o teclado elétrico. Para suportar essa precisão, será fundamental testar um espectro mais amplo de funções candidatas na etapa de modelagem matemática, incluindo funções de surto (*surge functions*) e modelos não-lineares complexos que descrevam com maior fidelidade os transientes de ataque.

Por fim, reconhecendo que o timbre é um fenômeno multidimensional, a evolução definitiva deste trabalho reside na hibridização de descritores. Propõe-se a fusão dos parâmetros temporais ADSR, aqui validados, com descritores espetrais consolidados, como os Coeficientes Cepstrais de Frequência Mel (MFCCs) ou o Centroide Espectral (Felix, 2019). Acredita-se que a união entre a precisão temporal do envelope e a riqueza frequencial do espetro permitirá aumentar o nível de precisão, aproximando o sistema da capacidade discriminatória do ouvido humano.

REFERÊNCIAS

American National Standards Institute, 1994. ANSI S1.1-1994: Acoustical Terminology.

Beauchamp, J. W., 2007. Analysis, Synthesis, and Perception of Musical Sounds: The Sound of Music. Springer, London.

Benson, D. J., 2006. Music: A Mathematical Offering. Cambridge University Press, Cambridge.

Drew Swicher, 2019. ADSR: The Best Kept Secret of Pro Music Producers! Página sobre Música. Disponível em: <https://mastering.com/adsr/>

Felix, G. A., 2019. Classificação de timbres de instrumentos musicais usando aprendizagem de máquina supervisionado. Monografia de Bacharelado. Universidade Federal de Juiz de Fora.

Grey, J. M., 1977. Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America* 61 (5), 1270–1277.

Guilherme Castilho, 2023. Aplicando Regressão Linear com o Método dos Mínimos Quadrados Através de Diferentes Ferramentas Computacionais. Página sobre Música. Disponível em: <https://medium.com/@guikstilho/aplicando-regress%C3%A3o-linear-com-o-m%C3%A9todo-dos-m%C3%A9dios-quadrados-atrav%C3%A9s-de-diferentes-ferramentas-d54b35d03f3c>

Müller, M., 2015. Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications. Springer, Cham.

Oppenheim, A. V., Schafer, R. W., 2009. Discrete-Time Signal Processing, terceira Edição. Pearson, New Jersey.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., McAdams, S., 2011. The Timbre Toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America* 130 (5), 2902–2916.

Roads, C., 1996. The Computer Music Tutorial. MIT press, Cambridge, MA.

Silverthorn, D. U., 2017. Fisiologia Humana: Uma Abordagem Integrada, 7ª Edição. Artmed, Porto Alegre.

Versilian Studios, 2024. VSCO 2: Community Edition (CE). Biblioteca de Instrumentos Virtuais. Disponível em: <https://vis.versilstudios.com/vsco-community.html>

Wikimidia, 2007. ADSR. Página Informativa. Disponível em: https://commons.wikimedia.org/wiki/File:ADSR_parameter.svg

Wikipedia, 2005. ADSR. Página Informativa. Disponível em: <https://pt.wikipedia.org/wiki/ADSR>