AI É NÓS!

MÓDULO 4 |
CRIAÇÃO DE ÁUDIO COM IAS GENERATIVAS E SOFTWARES LIVRES

FICHA TÉCNICA

Instrutor/conteudista:
Raphael Ferreira

Coordenação pedagógica:
Fernanda Carvalho

Design instrucional:
Raphael Ferreira

Projeto gráfico e diagramação:
Fernanda Carvalho e Julia Lechi



MÓDULO 4

1 - Introdução....4

2 - Um breve panorama das IAs no mundo dos sons....4

3 - Som e arte....6
3.1 Tipos de artes do som....7

4 - Pesquisa artística....8

5 - Som e áudio....9
5.1 Som....9
5.1.1 Características do som....10
5.2 Áudio....11
5.2.1 Analógico x digital....11

6 - Produção de áudio....12
6.1 Processos de Produção....12
6.1.1 Sampleamento (Amostragem)....13
6.1.2 Síntese sonora 13
6.2 Efeitos sonoros....13
6.3 Softwares de produção de áudio....14
6.4 Softwares para produção de áudio generativo....14

7 - Modelos de IA para áudio....16
7.1 Funções (tasks)....16
7.2 Parâmetros....17

8 - Engenharia de prompts....18
8.1 Prompts postivos....18
8.2 Prompts negativos....18
8.3 Gerador de prompts....19

9 - Pós-produção....19
9.1 Pós-produção em softwares de IA generativa....20
9.2 Formatos de áudio....21

REFERÊNCIAS....22

MÓDULO 4

Competência:

Utilizar de maneira contextualizada os fundamentos da criação de áudios generativos com IAs, contemplando ferramentas proprietárias, ferramentas abertas e softwares livres..

Resultados de aprendizagem:

Aplicar demonstrativamente os fundamentos da da criação de áudios com IAs generativas de modo crítico e propositivo.

AULA 1

VIDEOAULAS DISPONÍVEIS NA ABA "MÓDULO 4 - AULA 1"

1 - Introdução

Neste módulo, daremos um rolê pelo mundo dos sons. Aprenderemos a criar faixas de áudio completas, desde o seu planejamento inicial e pesquisa até a finalização, utilizando inteligências artificiais (IAs) generativas e softwares livres.

Passaremos pela história das IAs no universo sonoro e veremos os princípios básicos e termos fundamentais para a compreensão do trabalho com áudio. Apresentaremos as diversas possibilidades de utilização do áudio em obras de arte de diferentes linguagens.

Exploraremos os principais efeitos para uso criativo em suas faixas e visitaremos modelos de IA generativa Open Source, além de softwares de edição que integram algoritmos avançados ao seu fluxo de trabalho. O objetivo é que você entenda como essas ferramentas podem auxiliar de forma colaborativa no processo de criação em áudio.

2 - Um breve panorama das IAs no mundo dos sons

A ideia de a música ser feita por máquinas não é nova! Sempre criamos ferramentas para nos ajudarem a expandir nossa capacidade musical, desde o violino e o piano até o simples sino dos ventos (Figura 1), que faz música sozinho com a força do vento. No passado, até compositores como Mozart usavam um "jogo de dados" para criar novas melodias de forma aleatória, um método antigo de composição automatizada. Isso nos mostra que a automação, um fator importante na utilização das inteligências artificiais é algo que acompanha a humanidade ao longo de sua história, causando estranhamentos e adimiração, bem como a matemática e a física estão intrínsecamente

ligados ao mundo da música e do som, fator em comum com a computação.

Figura 1 – Sino dos ventos



Fonte: Imagem do autor (criação com IA generativa)

Depois de um tempo, a partir da década de 1940, artistas da Música Concreta expandiram as possibilidades da música, usando sons gravados do dia a dia (como o barulho de uma caneca ou de uma serra) como a principal matéria-prima para a criação. Essas técnicas, hoje são comuns em contexos diversos, como trilhas sonoras para obras audiovisuais, ou mesmo em videos virais de internet, no qual artistas tocam instrumentos inusitados para se expressar.

Na década de 1990, a popularização da computação digital e a música eletrônica, ampliaram esse universo, permitindo criar sons e ritmos totalmente novos diretamente no computador. Muitas polêmicas ainda dividem músicos puristas que prezam pelo som orgânico, derivado do virtuosismo do músico e os que abraçam as tecnologias digitais em suas produções e usam os computadores como colaboradores. Seja em performances de artistas da música eletrônica (Figura 2) com várias parafernalhas, como no show do tiozinho do churrasco (Figura 3) tocando seu teclado com um acompanhamento automático, diferentes tipos de IAs estão presentes nos processos.

Figura 2 – Performance de música eletrônica



Fonte: Imagem do autor (criação com IA generativa)

Figura 3 – Performance do “tio dos teclados”



Fonte: Imagem do autor (criação com IA generativa)

E não para por aí. A partir dos anos 2000, as inteligências artificiais além de possibilitar a criação de sons e músicas diretamente do computador ou de um celular, estão presentes nas recomendações de músicas de estilos semelhantes que recebemos ao utilizar plataformas de streaming, ou mesmo na identificação de voz presentes em assistentes virtuais, na tradução de discursos em tempo real, na criação de locuções para vídeos e muito mais.

As tecnologias que permitiram esses avanços no mundo dos sons são as **redes neurais, modelo computacional inspirado no cérebro humano, formado por camadas de “neurônios” interconectados que processam dados.** Ao receber uma grande quantidade de informações, a rede neural aprende a identificar padrões complexos, como harmonias, ritmos e tons.

Com o avanço das redes neurais especialmente a partir de 2010, se popularizam as inteligências artificiais generativas, que permitem que hoje estejam disponíveis para qualquer usuário, Interfaces como Udio, Suno AI e Eleven Labs, programas que conseguem aprender o estilo de milhares de músicas e gerar faixas completas, com letra e voz, de forma incrivelmente realista a partir de poucos cliques e prompts de texto.

Apesar de a IA assustar alguns artistas e makers, devemos pensar que as tecnologias estão envolvidas em todas as atividades do trabalho humano, desde ferramentas rudimentares até as IAs generativas que abordaremos nesse módulo.

3 - Som e arte

O som é algo que faz parte de nossas vidas desde o útero materno. Mesmo pessoas com deficiências auditivas podem sentir as vibrações sonoras em ressonância com seu corpo. Além de um fenômeno físico e biológico o som é cultural e perpassa diversas linguagens artísticas. Nossa maneira de falar, os movimentos do corpo, o meio ambiente, emfim tudo produz algum tipo de som.

E esse contexto universal serve de matéria prima para as diversas potencialidades artísticas das sonoridades.

Quando falamos em arte e som, de primeira já nos surge a música, mas ela é só uma de suas inúmeras linguagens. Vamos conhecer algumas delas e ampliar o nosso repertório!

3.1 Tipos de artes do som



O som está relacionado à diversas linguagens, seja na **música,** nas **artes sonoras**, ou mesmo no **design sonoro** de obras audiovisuais. Todas elas se encontram nas fronteiras da **arte, ciência e tecnologia.** Vamos conhecer alguns dos tipos de artes do som:

Música

Música é a arte de organizar sons e silêncios ao longo do tempo, utilizando-os para criar uma forma de expressão estética, comunicativa ou lúdica. Dentro da música temos várias vertentes que contemplam desde a musica clássica à popular. Na música, alguns gêneros são mais dedicados ao entretenimento, já outros estão mais conectados com a pesquisa em arte ciência e tecnologia. Confira alguns gêneros mais experimentais:

Música Eletroacústica - Mistura sons eletrónicos (sintetizadores) com sons acústicos gravados. **Ex:** Morton Subotnick - *Silver Apples of the Moon* (primeira composição eletrônica de longa duração a ser lançada por uma gravadora clássica)
Musica concreta - Usa sons do dia a dia (ruídos de trem, conversas) e os manipula para criar uma composição. **Ex:** Pierre Schaeffer - *Étude aux chemins de fer* (composta por uma montagem e manipulação de gravações de ruídos de locomotivas e trens.)

Arte Sonora

A arte sonora surge da interdisciplinaridade entre a música e demais linguagens artísticas, como as artes visuais. O que conhecemos hoje como arte contemporânea é fruto de de diversos movimentos de vanguarda que expandiram as fronteiras do que se entende por arte. Grupos como o Fluxus, os Futuristas e muitos outros tornaram possível com suas ações o surgimento do que hoje chamamos de artes sonoras, isto é, uma arte derivada dos sons, mas sem a necessidade de melodia ou harmonia. Confira alguns tipos de arte sonora:

- **Perfomance -** Explora as possibilidades do corpo e podem se basear nas potencialidades sonoras do mesmo. **Ex:** Inês Terra - *Língua Fora* **(**A artista usa o próprio corpo como uma fonte sonora, explorando o som produzido pela língua em relação a outros ruídos e ao silêncio. A obra é uma performance que transforma o som mais íntimo em uma experiência artística e corporal.)
- **Instalação Sonora -** Obra de arte criada para um espaço específico, onde o som é o elemento central da experiência. **Ex:** Doug Aitken - *Sonic Pavilion* (A obra é um pavilhão circular de vidro no Inhotim que, através de microfones instalados em um poço de mais de 200 metros de profundidade, capta e amplifica em tempo real o som da Terra.)
- Paisagens Sonoras (Soundscapes)
- **Paisagens Sonoras (Soundscapes)** - Registros de ambientes sonoros para capturar a "identidade" de um local. **Ex:** R. Murray Schafer - *The Vancouver Soundscape* (gravação documental de 1973, dirigida por R. Murray Schafer, que catalogou os sons naturais, urbanos e as assinaturas sonoras da cidade de Vancouver)
- **Poesia sonora -** Poesia que foca na sonoridade das palavras em vez do significado **Ex:** Paulo Bruscky - Poazia (o artista explora as potencialidades sonoras de um sal de frutas efervescente)

4 - Pesquisa artística

Se você acha que pesquisa é coisa de cientista de jaleco, pense de novo. **Pesquisar em artes é uma aventura de criação e descoberta, e o laboratório pode ser seu ateliê, seu computador, a rua ou sua própria mente**. Vamos direto ao ponto para você entender como funciona.

A pesquisa artística, como já discutimos, é um trânsito entre o fazer e o pensar. Na arte sonora, seu "fazer" é a escuta, a gravação, a manipulação de frequências, timbres e espacialidades. Seu "pensar" é sobre o que esse som significa, como ele afeta nossa percepção do espaço, do tempo e da emoção.

Agora, insira a IA generativa nesse fluxo. Ela não é apenas uma ferramenta, como um microfone ou um software de edição. Ela é um agente intermediário, uma parceira de diálogo com sua própria lógica, seus próprios vieses e sua própria "escuta" algorítmica.

Pesquisar em arte é como ser um detetive da sua própria criatividade. Em vez de apenas criar algo bonito, você investiga uma ideia ou pergunta através do seu processo artístico. Seus experimentos, seus rabiscos, suas gravações e até seus erros não são falhas, mas sim as pistas e descobertas da sua investigação. Não existe uma fórmula pronta: você inventa seu próprio jeito de explorar, seja com tinta, com sons de objetos do dia a dia ou com qualquer outro material. A arte se torna uma aventura para entender o mundo, onde o caminho que você percorre ao criar é tão importante quanto a obra de arte final.

Hoje, você pode levar essa aventura para o mundo digital usando Inteligências Artificiais generativas como parceiras criativas. Para pesquisar com sons, por exemplo, você não precisa de um estúdio caro; seu laboratório se torna uma conversa com a IA. Você a provoca com textos (prompts) para que ela imagine e crie paisagens sonoras, ritmos estranhos ou timbres que ninguém nunca ouviu. A grande sacada é que os **"erros" e as respostas inesperadas da IA são suas maiores descobertas, revelando a lógica de uma "mente" não-humana.** Assim, sua pesquisa se transforma em uma exploração de novos universos sonoros.

5 - Som e áudio

Estamos nos referindo a som e áudio, mas afinal, não é tudo a mesma coisa? Nada disso, apesar de usarmos essas palavras com frequência, cada uma tem o seu lugar. Vamos entender quais são as diferenças e principais características de cada uma delas.

5.1 Som

Som é uma onda de energia mecânica, longitudinal e tridimensional que se propaga através de um meio físico (figura 4).

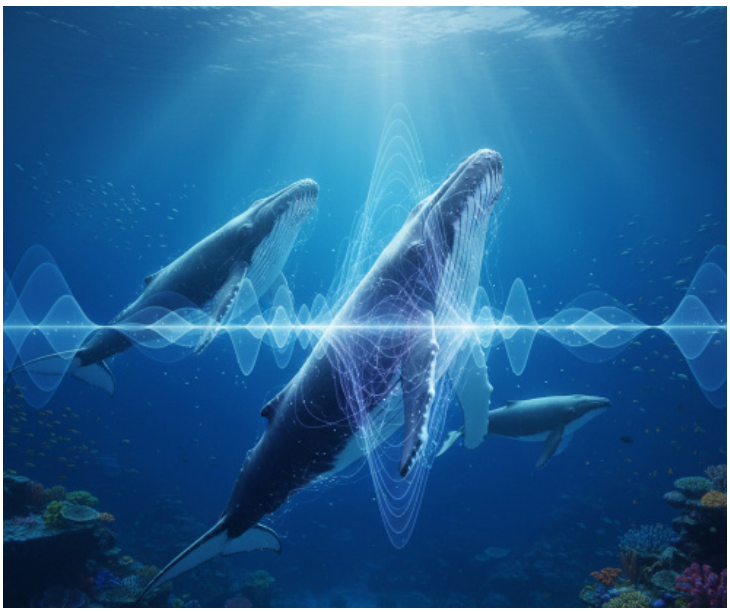
Figura 4 – Ondas sonoras



Fonte: Imagem do autor (criação com IA generativa)

Como exemplo de meio físico temos o ar, a água, ou um material sólido. Essa onda é gerada por uma vibração que empurra as moléculas de ar próximas a ela, criando diferenças de pressão que resultam no som. Um exemplo do som se propagando na água é o canto das baleias (Figura 5).

Figura 5 – Representação visual do canto das baleias



Fonte: Imagem do autor (criação com IA generativa)

Resumindo, ele é o resultado de uma vibração que se propaga e é percebida no nosso corpo, especialmente nos ouvidos. Mas também sentimos as vibrações do som em outros tecidos, como quando sentimos o pancadão bater no estômago.

5.1.1 Características do som

E quais são as principais características do som? Vamos ver agora, cada uma delas!

- **Intensidade -** A **intensidade** do som se refere à **força das vibrações**, ou seja, **amplitude** da onda sonora. Quando falamos para o DJ aumentar o som, estamos nos referindo à intensidade, que cresce ao girarmos o botão de volume.
- **Altura (Frequência) -** A frequência é o tamanho da onda. Frequências **baixas** (aqui nos referimos ao tamanho da onda e não à intensidade) caracterizam **sons graves**, isto é a voz de um locutor de rádio, um rugido de um leão, um trovão, um bumbo batendo. Já os **sons altos**, ou **agudos** são os sons da voz de uma criança, de uma mulher, de um apito, de pássaros cantando, de um violino. As frequências médias, ficam entre as altas e baixas, como um violão, um violoncelo ou a voz de um adolescente.
- **Timbre -** Timbre é a "**cara**", a "**identidade"** do som. É a característica que nos permite distinguir um som do outro. Quando tocamos uma nota musical, por exemplo um "lá" que vibra em 440hz, esses mesmos 440hz soarão de forma diferente em um piano, um saxofone, uma guitarra ou uma voz cantando. Conseguimos identificar os instrumentos por conta de seu timbre.

5.2 Áudio

O áudio é a materialização do som em um dispositivo (Figura 6). É o som que foi capturado, processado, armazenado ou transmitido por meios eletrônicos. Enquanto o som é a vibração física que se propaga, o áudio é a representação elétrica ou digital dessa vibração.

Figura 6 – Gravação de uma faixa de áudio em estúdio



Fonte: Imagem do autor (criação com IA generativa)

5.2.1 Analógico x digital

O áudio pode ser analógico ou digital e por mais que para os nativos digitais essa conversa não faça muito sentido, para músicos puristas, hipsters e admiradores do disco de vinil são fieis seguidores da sonoridade analógica.

Mas afinal, qual é a diferença entre o analógico e digital? O áudio **analógico** é a representação do som como um **sinal elétrico contínuo**, que varia de forma **idêntica à onda sonora original.** Presente em mídias como fitas cassete e vinis, ele é conhecido por sua sonoridade **“quente” e natural**, mas também é mais vulnerável a ruídos e à degradação ao longo do tempo. É considerado um som mais “vivo”, mais orgânico, apesar desta ser uma concepção bem subjetiva.

Já o áudio **digital** é a **representação do som usando código binário** (zeros e uns). O som é amostrado e transformado em dados numéricos, sendo o formato de áudio presente em **CDs, MP3s e serviços de streaming**. Suas principais vantagens são a reprodução precisa, a **resistência a ruídos** e a f**acilidade para edição e armazenamento**. Sim, o digital revolucionou a produção de áudio e se hoje estamos aqui estudando audio e IAs é graças ao áudio digital.

Antes do áudio digital, gravar algum registro fonográfico era para poucos. Com a popularização do áudio digital, qualquer um com um computador e uma ideia na cabeça passou a ter a chance de produzir suas próprias faixas. Hoje, com as IAs generativas, a produção de áudio se tornou ainda mais simples. Com um punhado de palavras e uma ideia na cabeça, nem o computador é mais necessário, qualquer um pode criar suas canções com seus celulares ou qualquer outro dispositivo digital que se conecte à internet.

6 - Produção de áudio

A **produção de áudio** é a arte e a t**écnica de capturar, criar, editar e manipular sons para gerar um produto final**, como uma música, um podcast, uma trilha sonora para cinema ou um efeito sonoro. É o processo que transforma ideias sonoras em realidade, seja por meio de gravações ou da criação de sons do zero.

O áudio produzido pode ter diversas aplicações que integram o campo das artes.

- **Trilhas Sonoras -** Criadas para filmes, séries, videogames ou peças de teatro, elas ajudam a contar a história e a evocar emoções no público. As trilhas sonoras de filmes, séries e jogos são formadas por três elementos principais que são a trilha musical, que inclui a música, diálogos e efeitos Sonoros.

DICA

Você pode usar as IAS generativas para criar efeitos diversos. O Stable Audio é uma boa ferramenta open source para tal finalidade.

- Produção Musical - O foco é a criação de canções para artistas, bandas ou playlists. Envolve desde a gravação dos instrumentos e vocais até a mixagem e masterização final. A produção musical é um processo de cinco etapas que são a composição e arranjo (Criação da melodia, letra e estrutura da música); gravação (captura dos sons de instrumentos e vocais); edição (limpeza das gravações, corrigindo erros e organizando as faixas); mixagem (Equilíbrio de todas as faixas e aplicação de efeitos para criar um som coeso) e masterização (polimento final para otimizar o som da música em qualquer sistema de áudio).

DICA

As IAs generativas são usadas na produção musical para compor novas melodias, criar arranjos complexos gerar batidas originais e sintetizar timbres únicos com modelos abertos como Magenta e proprietários como o Suno são bons nessas tarefas.

6.1 Processos de Produção

A forma como o som é criado é uma das distinções mais importantes na produção de áudio. As duas principais abordagens são o sampleamento e a síntese sonora e cada um dos processos tem suas peculiaridades, apesar de muitas vezes se misturarem.

6.1.1 Sampleamento (Amostragem)

O **sampleamento** é o processo de **gravar um som já existente para usá-lo na produção**. Um gravador digital captura um trecho de áudio (o “sample”), que pode ser uma nota de piano, uma batida de bateria, ou até um vocal. Depois, esse arquivo de áudio pode ser reproduzido, manipulado e usado como um instrumento em uma composição.

6.1.2 Síntese sonora

A **síntese sonora** é o processo de **criar um som do zero usando circuitos eletrônicos ou software**. Em vez de gravar, você constrói a onda sonora a partir de formas de onda básicas (como ondas senoidais, quadradas ou dente de serra) e manipula parâmetros como a frequência, a modulação e o timbre.

6.2 Efeitos sonoros

O som pode ser transformado para soar como se estivesse debaixo d’água, em um templo ou como um remix de DJ. Essas transformações são feitas por efeitos sonoros (FX), que se baseiam em três matrizes principais: Filtragem de Pente (Comb Filtering), Reverberação (Reverb) e Saturação (Saturation).

Filtragem de Pente (Comb Filtering)

A Filtragem de Pente ocorre quando dois sons idênticos são tocados com um pequeno atraso no tempo. Mesmo um atraso mínimo entre o som original e sua cópia já “muda a cara” do áudio. Os principais efeitos gerados por essa matriz são:

- **Delay -** Cria repetições do som, variando de um leve efeito de profundidade a um eco claramente audível, muito usado para dar ritmo extra.
- **Chorus -** Simula um coro (múltiplas fontes sonoras) cantando a mesma nota, adicionando corpo e profundidade ao som, comum em guitarras new wave (anos 80) e vocais.
- **Flanger -** Gera um som metalizado e cíclico, parecido com um avião passando. É dramático e cria texturas psicodélicas.
- **Phaser -** Parecido com o Flanger, cria um som robotizado e psicodélico ao manipular a fase (picos e vales) da onda sonora.

Reverberação e Saturação

Essas matrizes definem a espacialidade e a intensidade do som.

- **Reverberação (Reverb) -** É a ambientação do som, mostrando como ele reflete no ambiente. O reverb permite que o ouvinte perceba se o som está em um quarto, em um hall ou em uma igreja, dando dimensão ao áudio. O eco é um tipo de reverb longo.
- **Saturação -** É um aumento de sinal que adiciona calor, brilho e presença ao som, sendo uma forma sutil de distorção. Quando levada ao extremo, causa a distorção que faz o som parecer “rasgado”, como as guitarras de rock ou o som de um megafone.

Os efeitos sonoros são cruciais para a IA generativa. Eles podem ser incluídos nos prompts (comandos de texto) para que os modelos criem músicas com ambiências específicas (ex: “melodia com reverb de catedral”) ou aplicados na pós-produção para finalizar e editar o áudio gerado pela IA.

6.3 Softwares de produção de áudio

Para criar e produzir áudio, usamos softwares chamados DAWs (Digital Audio Workstations). Elas são como um estúdio de gravação virtual, permitindo gravar, editar e finalizar áudios de forma profissional.

Exemplos de DAWs

- **Proprietárias (Pagas) -** Pro Tools (padrão da indústria), Ableton Live (popular na música eletrônica) e FL Studio (muito usado no Hip Hop).
- **Livres (Gratuitas) -** Audacity (https://www.audacityteam.org/)

O Poder dos Plug-ins

As DAWs se tornam ainda mais poderosas com a instalação de aplicativos extras de terceiros, chamados Plug-ins.Eles funcionam como “acessórios” para expandir as funcionalidades da DAW.
Os Tipos de Plug-ins Podem ser **efeitos clássicos** (como reverb e delay) ou **instrumentos virtuais** (simuladores de piano, amplificadores de guitarra ou baterias eletrônicas).

Há muitas empresas que produzem plug-ins de alta qualidade, como Native Instruments e Universal Audio (pagos), mas também há ótimas opções gratuitas, como as da Analog Obsession (https://analogobsession.com).

6.4 Softwares para produção de áudio generativo



O mercado da produção musical e audiovisual está sendo remodelado por ferramentas de inteligência artificial generativa. Confira algumas das plataformas que oferecem diversos serviços ao usuário.

Ferramentas proprietárias

Os modelos mais conhecidos, como **Suno AI** e **Udio**, são proprietários (pagos e de código fechado) e se destacam por criar músicas completas, com letra, voz e instrumentação, a partir de um simples comando de texto. A conveniência é alta, mas o usuário não tem controle sobre o processo interno de criação da IA.

A palataforma **ElevenLabs** se especializa em síntese e clonagem de voz de alta fidelidade, um recurso vital para a

produção de audiobooks e vocais gerados por IA. Quem não se lembra daquela clássica voz irônica dos memes de vídeo que circulam nas redes sociais? É o Adam, um avatar de voz amplamente utilizado na plataforma ElevenLabs. Todas as plataformas oferecem planos gratuitos com algumas limitações, além de assinaturas com pacotes específicos de benefícios.

Ferramentas open source

Para ter controle total sobre a criação, a interface **ComfyUI** (já citada no módulo 3 deste curso) é uma ferramenta **open source (**código aberto) que usa um sistema de blocos interconectados (nós) para personalizar cada etapa do fluxo de trabalho da IA generativa de áudio, imagem e vídeo.Nele você poderá rodar modelos abertos diretamente em seu computador. Para os que desejam utilizá-lo diretamente no navegador, uma possibilidade é a computação em nuvem, como na plataforma runpod.

Para quem busca experimentação gratuita, o mercado também oferece boas opções, mesmo que não sejam de código aberto. O aplicativo **Speechma** é uma alternativa para a síntese de voz gratuita, permitindo a criação de áudio vocal sem custos. As plataformas **Labs** e **AI Studio** da Google também oferecem a possibilidade de criar locuções e faixas musicais diretamente no navegador.

Por fim, a comunidade de IA se apoia em plataformas de teste acessíveis. O **Hugging Face Spaces** e o **Google Colab** são ambientes online que permitem que qualquer pessoa use e experimente diversos modelos de IA de áudio (muitos deles open source) diretamente no navegador. Isso elimina a necessidade de instalar programas complexos ou ter um computador potente, tornando a exploração da tecnologia acessível a todos.

LINKS ÚTEIS

Suno AI - https://www.suno.com
Udio - https://www.udio.com
ElevenLabs - https://www.elevenlabs.io
Speechma - https://www.speechma.com
Google Labs - https://labs.google/fx/pt/tools/music-fx
Google AIStudio - https://aistudio.google.com/generate-speech
ComfyUI - https://www.comfy.org
Rundpod - https://www.runpod.io/

VIDEOAULAS DISPONÍVEIS
NA ABA ”MÓDULO 4 - AULA 2”

7 - Modelos de IA para áudio

A mágica da geração musical por IA se baseia em duas arquiteturas principais: os modelos T**ransformer** e os modelos de **difusão** (Diffuser). Os Transformers são essenciais para entender e gerar a estrutura sequencial e a linguagem musical (como notas e harmonias coesas). Já os modelos de Difusão (como o Diffrhytm) se concentram em criar o áudio final com alta fidelidade e realismo, eliminando ruídos e produzindo detalhes sonoros profissionais. É a combinação dessas tecnologias que permite que a IA não apenas componha, mas produza faixas prontas.

Para quem busca mais liberdade e controle, o movimento open source (código aberto) oferece alternativas. Projetos como o **Magenta (do Google), Stable Audio, AceStep, Diffrythm, Yue AI, Kokoro TTS** são alguns exemplos de modelos que podem ser utilizados gratuitamente e localmente por meio de interfaces como o ComfyUI. Os modelos também podem ser baixados e testados na plataforma Huggin Face.

7.1 Funções (tasks)

Confira as quatro principais funções (tasks) dos modelos de IA para áudio generativo:

1. Geradores de voz (Text-to-speech - TTS) - Comumente chamada de Text-to-Speech (TTS), esta IA transforma qualquer texto que você digita em fala de computador, com uma voz que parece bem humana. Ela é a base para assistentes virtuais como Siri e Alexa, narrações de vídeos e também é usada em ferramentas de acessibilidade para ajudar pessoas com deficiência visual a “ler” textos.

2. Geradores de música (Text-to-music) - Esses modelos são conhecidos como Text-to-Music (quando criam música a partir de uma descrição em texto) ou MIDI-to-Audio (quando convertem dados musicais em áudio). Sua função é compor músicas, melodias, batidas e até arranjos completos, tudo do zero. Servem para ajudar DJs e produtores a criar novas faixas, gerar trilhas sonoras para jogos e vídeos, e são perfeitos para quem quer experimentar novos estilos musicais.

3. Geradores de efeitos sonoros (Text-to-Audio) - Chamados de Text-to-audio (um termo mais geral para qualquer áudio gerado por texto), esses modelos criam sons específicos que você descreve, como o barulho de chuva, um trovão ou o som de uma porta fechando. São ideais para criar efeitos sonoros em jogos, filmes, animações ou para adicionar sons interessantes em podcasts e outras produções de áudio.

4. Transferência de estilo de áudio (Audio-to-Audio) - Essa IA pega o “jeito” de um som (como a voz de um cantor famoso ou o timbre de um violino) e aplica em outra música ou melodia. Músicos podem usá-la para dar uma “nova voz” às suas composições, ou criar versões de músicas com sons super diferentes e criativos, abrindo um leque de experimentação.

7.2 Parâmetros

Os parâmetros são o treinamento da IA (o que ela aprendeu) e as opções de ajuste que você tem para guiá-la. Eles são a chave para conseguir um áudio de qualidade e de acordo com os objetivos planejados.

A IA de áudio é como um estúdio de música superpotente, e os parâmetros são os botões que você usa para dizer exatamente como quer o som. Eles controlam a “receita” final. Ajustar os parâmetros é como ser o produtor musical. Você guia a IA (o estúdio) para criar exatamente o que você tem em mente.

Principais parâmetros de IAs de áudio:

- **Prompt -** É o comando principal. Sua descrição detalhada do conteúdo a ser gerado (gênero, instrumentos, emoção, atmosfera).
- **Duração -** É o tamanho do áudio. Define o tempo exato (em segundos) que a faixa terá.
- **CFG Scale (Guidance Scale) -** É o nível de obediência da IA.Controla o quanto a IA deve seguir seu prompt. Valor alto para rigidez; valor baixo para mais criatividade.
- **Taxa de Amostragem (Sample Rate) -** É a Qualidade HD do Som. Determina a fidelidade técnica. Taxas mais altas geram áudio mais detalhado e profissional, mas exigem mais tempo de processamento.
- **Seed -** É o Código de reprodução. Um número que, se repetido com os mesmos comandos, garante que o resultado será exatamente o mesmo. Essencial para refinar uma música que você gostou.
- **Passos de Geração (Generation Steps) -** É o esforço de refinamento da IA. Controla o número de etapas para aprimorar o áudio. Mais passos significam um som mais polido e detalhado, embora demore mais.

LINKS ÚTEIS

Stable Audio - https://huggingface.co/stabilityai/stable-audio-open-1.0
AceStep - https://huggingface.co/spaces/ACE-Step/ACE-Step
Diffrythm - https://diffrhythm.com/ai-music-generator
Kokoro TTS - https://huggingface.co/spaces/ACE-Step/ACE-Step

8 - Engenharia de prompts

A engenharia de prompts é a **arte de dar as instruções certas para a inteligência artificial**, garantindo que o áudio gerado seja exatamente o que você tem em mente. É como se você fosse um diretor de som, e a IA, a sua equipe de produção. Em vez de simplesmente dizer “faça um som”, você precisa ser o mais detalhado possível, seja para criar uma música, um efeito sonoro, uma locução, uma trilha sonora, foley ou uma arte sonora.

Os elementos de um bom prompt para áudio generativo são como a receita para uma criação sonora precisa. Comece com o **conteúdo principal** (música, locução, foley), **defina a atmosfera** ou **emoção** (alegre, tensa), e, dependendo do tipo de áudio, **inclua detalhes específicos**. Para músicas, especifique gênero, instrumentação e estrutura. Para locuções, defina as características da voz, a intenção e as pausas. E para foley ou paisagens sonoras, descreva o contexto, as ações e os materiais. Combinar esses elementos resulta em um prompt detalhado que guia a IA para gerar exatamente o som que você imaginou.

8.1 Prompts postivos

- **Músicas e Efeitos Sonoros -** Comunique o tipo de áudio, gênero, emoção e qualidade de produção para criar faixas ou efeitos sonoros precisos. **Ex**: “Música de rock clássico, instrumental, com guitarra elétrica distorcida, bateria com groove e baixo marcante. O som deve ter uma atmosfera nostálgica e poderosa, soando como uma gravação de fita analógica dos anos 80.”
- **Locuções e Narrações (Voz) -** Descreva a identidade completa da voz (idade, tom, sotaque) e a intenção da fala, usando pontuação para guiar o ritmo. **Ex:** Voz feminina, jovem (20 e poucos anos), com tom animado e entusiasmado. A voz deve estar narrando uma propaganda de bebida energética. O texto é: ‘Prepare-se! (pausa dramática) O limite é o céu!’”
- **Trilhas Sonoras (BGM) e Efeitos (Foley) -** Para trilhas, defina a cena e o clímax; para foley, descreva a ação, o material e o contexto do som que precisa ser criado. **Ex:** “Crie uma trilha sonora de suspense crescente para uma cena noturna na floresta. A música deve usar cordas lentas e notas de piano isoladas. Adicione o foley de galhos secos quebrando sob passos leves a cada 5 segundos.”
- **Artes Sonoras e Paisagens Sonoras -** Crie um ambiente sonoro descrevendo o local, os elementos específicos e a atmosfera emocional que devem compor a paisagem. **Ex:** “Crie uma paisagem sonora urbana e caótica, com o som de buzinas misturadas com sussurros de pessoas em uma língua desconhecida. O ambiente deve evocar um sentimento de solidão em meio à multidão.”

8.2 Prompts negativos

Assim como na criação de imagens, os **prompts negativos** são essenciais para refinar a geração de áudio. Eles servem para dizer ao modelo de IA o **que não incluir na saída final,** ajudando a remover ruídos indesejados, instrumentos que não se encaixam ou atmosferas que não combinam. Um prompt negativo é uma **lista de palavras-chave ou frases que atuam como “filtros”** para a inteligência artificial. Enquanto o prompt principal diz o que você quer, o prompt negativo diz o que você quer evitar.

Alguns modelos de áudio generativo não tem um campo de prompts negativos em separado. Caso isso ocorra, você pode incorporá-los no prompt principal usando palavras como “sem” ou “exceto”.

Exemplo: “Música de rock clássico, com riff de guitarra pesado, bateria energética, sem solo de piano, sem vocais masculinos.”

8.3 Gerador de prompts

Para facilitar os processo de criação de prompts, podemos utilizar geradores de prompts, aplicativos criados para produzir comandos textuais para modelos de IAs generativas. Temos alguns modelos proprietários como o Suno Prompt, o gerador do ChatGPT bem como podemos criar geradores de prompts em LLMs como o Gemini, e Qwen.

Gerador de prompt online

https://sunoprompt.com/
https://chatgpt.com/g/g-y2J4bAO8A-ai-music-prompt-generator
https://poe.com/StableAudioGenerator

DICA

Você também pode usar o gerador de prompts que criamos.
Acesse pelo link:
https://chat.qwen.ai/s/deploy/869cbb6e-ce73-4078-87e5-c82c4de53130

9 - Pós-produção

A **pós-produção** de áudio é a etapa final e crucial na criação de qualquer material sonoro, seja para filmes, vídeos, podcasts ou músicas. É o processo de **refinar, aprimorar e montar todos os elementos de áudio para que o resultado final** seja coeso, claro e impactante.



Isso envolve uma série de tarefas, como a **edição**, onde se corta e organiza os sons; a **mixagem,** que balanceia os volumes e frequências para garantir que todos os elementos (diálogos, músicas, efeitos) sejam ouvidos com clareza; e a **masterização**, que otimiza o áudio final para diferentes plataformas, garantindo que ele soe consistente em qualquer dispositivo. Em essência, a pós-produção transforma a gravação bruta em um produto polido e profissional, pronto para o consumo público.

Um detalhe importante é que a pós-produção de áudio é algo trabalhoso e desafiador para iniciantes, pois além de um conhecimento específico na área de engenharia de áudio é necessário treinar os ouvidos e ter disponíveis equipamentos de alta fidelidade sonora, além de um ambiente propício como um home-studio ou um estúdio profissional. Portanto, neste módulo focaremos na pós-produção realizada com ferramentas de IAs específicas.

9.1 Pós-produção em softwares de IA generativa

A pós-produção em softwares com foco nas inteligências artificiais são possibilidades que democratizam a produção de áudio para o público geral. Modelos que utilizam de diversas tasks, como a análise de áudio, transferência de estilo entre outras funções trazem aos usuários a chance de fazer suas gravações e gerações soarem de forma profissional, ou o mais próximo desse ponto.

Aplicativos proprietários bem como os open source oferecem diferentes soluções e algoritmos que entregam serviços de mixagem, masterização, separação de instrumentos, entre diversas funções que você poderá descobrir navegando em cada uma das diversas plataformas que estão em constante atualização.

As plataformas de referência na pós-produção de áudio são **proprietárias** como a **Suno**, **Landr**, **Moises** e **BandLab** que funcionam por meio de assinaturas mensais ou anuais. Algumas delas oferecem soluções gratuitas para determinadas funções.

- **Para limpeza de voz -** Se você precisa limpar narrações ou vocais, o ElevenLabs Voice Isolator (https://elevenlabs.io/voice-isolator) é excelente. Ele remove ruídos de fundo e ecos, deixando o áudio da voz cristalino. O Elevenlabs pode ser baixado como app diretamente em dispositivos móveis nas lojas officiais de aplicativos.
- **Para mixagem -** A plataforma Moises (https://studio.moises.ai/library/) permite que você separe faixas individualmente de forma gratuita no plano free. Uma boa opção para utilizar em mixagens, alterando elementos de cada faixa individualmente.O Moises pode ser baixado como app diretamente em dispositivos móveis, nas lojas officiais de aplicativos.
- **Para masterização -** O BandLab Mastering (https://www.bandlab.com/mastering) é uma ferramenta fantástica e gratuita que faz a masterização do seu áudio. Ele ajusta o volume e o equilíbrio do som para que sua música ou gravação soe bem em qualquer lugar. O bandnlab pode ser baixado como app diretamente em dispositivos móveis nas lojas officiais de aplicativos.

No modelo **open source**, ou com licenças mais permissivas temos a **Songmastr**, o **Audacity** (que apesar de não ser uma ferramenta dedicada ao uso de IAs generativas conta com plugins que cumprem esta função como o **OpenVINO** entre vários outras empresas do mercado, além de modelos de linguagem que cumprem funções úteis em pós-produção como o **Music Mixing Style Transfer,** e o **SonicMaster.**

- **Para mixagens e restauração de áudio -** O SonicMaster é uma opção de modelo de linguagem que cumpre diversas funções. Você faz o upload da faixa e acrescenta seus obejtivos no prompt, como por exemplo “abaixe os vocais”, “retire os ecos”, “equilibre a gravação”, “aumente os agudos”, “reduza os graves”. Os comandos devem ser digitados em inglês. Você poderá usar

o google tradutor ou uma LLM para te auxiliar. O modelo pode ser rodado localmente ou utilizado no navegador através da plataforma Hugging Face Spaces (https://huggingface.co/spaces/amaai-lab/SonicMaster)

- **Para masterização -** O Songmastr é uma opção Open Source que disponibiliza a masterização em cotas gratuitas ou por meio da compra de créditos de IA. Acesse pelo navegador, faça o upload do áudio a ser masterizado e o upload de um áudio de referência para que a mágica aconteça. Acesse pelo navegador através do link https://www.songmastr.com/.
- **Para rodar localmente -** O OpenVINO é um pacote de plug-ins que oferece diversas funções de podução e pós-produção de áudio. Com ele você pode fazer transcrições de áudio para texto, retirar ruídos de gravações, gerar músicas e tudo isso diretamente de dentro do Audacity, uma DAW open source multiplataformas. O OpenVino está disponível para Windows e MACOS e pode ser instalado por meio do link https://www.audacityteam.org/download/openvino/. (É necessário ter o Audacity instalado em sua máquina)

9.2 Formatos de áudio

Após os processos de edição, mixagem e masterização do áudio é hora de exportá-lo para a execução em dispositivos diversos. Para isso, contamos com diversos formatos de arquivo. Os formatos de áudio são a forma como o som é armazenado digitalmente, e cada um tem características distintas que afetam a qualidade, o tamanho do arquivo e a compatibilidade. A escolha do formato ideal depende do objetivo, seja para streaming, arquivamento ou edição profissional.

- **MP3 -** O MP3 é o formato mais popular e universalmente compatível. Sua principal característica é a compressão com perda de dados, que reduz drasticamente o tamanho do arquivo ao remover informações sonoras que o ouvido humano raramente detecta. Ideal para streaming e armazenamento em dispositivos móveis.
- **FLAC -** O FLAC é um formato de áudio sem perda de dados. Ele comprime o arquivo sem descartar nenhuma informação da gravação original, garantindo a qualidade perfeita do som. Ideal para arquivamento de coleções de música, para audiófilos e para distribuição de áudio de alta fidelidade.
- **WAV -** O WAV é um formato de áudio sem compressão, ou seja, ele não perde nenhuma informação, mantendo a qualidade máxima da gravação. É o padrão em estúdios de gravação, mas é inviável para streaming e uso cotidiano devido ao seu tamanho.

E chegamos ao fim deste módulo!

Bora produzir!

Assista aos vídeos das aulas para conferir na prática como gerar e finalizar suas obras.

REFERÊNCIAS

ALMEIDA, José Roberto P. de. Inteligência Artificial em Música (Parte 1). Musicologia, 2023. Disponível em: https://www.blogs.unicamp.br/musicologia/2023/07/17/inteligencia-artificial-em-musica-parte-1/. Acesso em: 5 out. 2025.

ALMEIDA, José Roberto P. de. John Chowning, o músico que "tropeçou" na síntese FM. Musicologia, 2019. Disponível em: https://www.blogs.unicamp.br/musicologia/2019/11/20/35/. Acesso em: 5 out. 2025.

AUDACITY TEAM. Audacity with OpenVINO AI Effects now on Mac. Audacity Blog, 2024. Disponível em: https://www.audacityteam.org/blog/openvino-mac/. Acesso em: 5 out. 2025.

DE NEGRI, Fernanda. Inteligência artificial e redes neurais. IPEA, 2019. Disponível em: https://www.ipea.gov.br/cts/pt/central-de-conteudo/artigos/artigos/106-inteligencia-artificial-e-redes-neurais. Acesso em: 5 out. 2025.

LUNA, Ianni Barros; CORREA, Antenor Ferreira. Arte sonora: atravessamentos entre som e imagem. In: SIMPÓSIO INTERNACIONAL DE INOVAÇÃO EM MÍDIAS INTERATIVAS, 5., 2018, Goiânia. Anais [...]. Goiânia: UFG, 2018. Disponível em: https://files.cercomp.ufg.br/weby/up/777/o/6_-_Ianni_Barros.pdf. Acesso em: 4 out. 2025.

GIBSON, David. The art of mixing: a visual guide to recording, engineering, and production. routledge, 2019.

RUSSELL, Aden. FM Synthesis: A Simple, No-BS Guide. EDMProd, 2024. Disponível em: https://www.edmprod.com/fm-synthesis/. Acesso em: 5 out. 2025.

RUSSELL, Aden. The Ultimate Guide to Subtractive Synthesis. EDMProd, 2024. Disponível em: https://www.edmprod.com/subtractive-synthesis/. Acesso em: 5 out. 2025.

TEIXEIRA, Marcos Almada; PALAZZO, L. A. M. O uso de redes neurais artificiais na composição de trilhas sonoras para jogos. In: CONGRESSO DA ASSOCIAÇÃO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM MÚSICA, 33., 2023, Rio de Janeiro. Anais [...]. Rio de Janeiro: UFRJ, 2023. Disponível em: https://anppom.org.br/anais/anaiscongresso_anppom_2023/papers/1628/public/1628-7842-1-PB.pdf. Acesso em: 5 out. 2025.

APOIO:

ASSIMPO

DISTÓPICA

REALIZAÇÃO:

PROJETO REALIZADO PELA SOCIEDADE CIVIL COM RECURSOS ORIUNDOS DA POLÍTICA DE FOMENTO À CULTURA MUNICIPAL

INCENTIVO:

LMIC
LEI MUNICIPAL DE INCENTIVO À CULTURA

BELO HORIZONTE
PREFEITURA

PROJETO 0272/2024 - Este projeto é realizado com recursos da Lei Municipal de Incentivo à Cultura de Belo Horizonte