

VIETNAMESE RVC

Công cụ huấn luyện, chuyển đổi nhạc cụ và giọng nói chất lượng và hiệu suất cao đơn giản.

Mô Tả Dự Án

Dự án này là một công cụ chuyển đổi giọng nói. Với mục tiêu tạo ra các sản phẩm chuyển đổi giọng nói chất lượng cao và hiệu suất tối ưu, dự án cho phép người dùng thay đổi giọng nói một cách mượt mà, tự nhiên.

Dự án này hướng tới sự thử nghiệm nghiên cứu của cá nhân hơn là về sự trải nghiệm, độ ổn định và có thể xảy ra lỗi trong quá trình sử dụng, nếu bạn muốn hưởng đến một dự án có sự ổn định, mượt mà nhất hãy dùng thử <https://github.com/IAHispano/Applio>, nếu bạn muốn hưởng tới thử nghiệm đây sẽ là dự án dành cho bạn.

Dự án này có thể sẽ không cung cấp bản đóng gói sẵn, chỉ cung cấp mã nguồn và một số hướng dẫn. Để sử dụng được dự án này bạn sẽ phải tự cài đặt thủ công, quá trình cài đặt có thể sẽ rất phức tạp nên nếu bạn vẫn muốn sử dụng có thể liên hệ tôi thông qua discord.

Hướng Dẫn Cài Đặt:

Trước tiên bạn cần tải mã nguồn về máy, bạn có thể thực hiện nó thông qua hai cách.

- **Cách 1: Sử dụng đối với Git**
 - git clone <https://github.com/PhamHuynhAnh16/Vietnamese-RVC.git>
 - cd Vietnamese-RVC
- **Cách 2: Tải trực tiếp trên GitHub**
 - Truy cập link: ` <https://github.com/PhamHuynhAnh16/Vietnamese-RVC/archive/refs/heads/main.zip> ` để tải trực tiếp về.
 - Giải nén tệp Vietnamese-RVC-main.zip.
 - Vào thư mục, gõ cmd trên thanh địa chỉ và nhấn Enter để mở Terminal.

Tiếp theo bạn sẽ thực hiện tiếp các bước cài đặt.

Đối Với Hệ Điều Hành Windows

- Cần cài đặt bộ [Visual C++ Redistributable Runtimes](#) trước khi tiến hành tiếp.

- Bạn có thể chạy trực tiếp tệp `run_install.bat` để cài đặt hoặc làm theo các bước bên dưới.
- Khi chạy `run_install.bat` và bạn sử dụng CUDA thì sẽ được hỏi có cài TensorRT hay không, nếu chọn có bạn sẽ cần cài thêm [TensorRT](#) từ Nvidia, giải nén và thêm đường dẫn thư mục bin của nó vào [PATH](#) hệ thống.
- Khi chạy `run_install.bat` và bạn sử dụng XPU thì sẽ được hỏi chọn phiên bản ONNXRUNTIME, tôi khuyên bạn chọn DIRECTML vì quá trình thử nghiệm của tôi OPENVINO có một số lỗi về kích thước động.
- Khi sử dụng ONNXRUNTIME OPENVINO, bạn cần cài đặt [OpenVino Toolkit](#) giải nén và thêm **Release** từ `\runtime\bin\intel64\Release` vào [PATH](#) hệ thống.
- Hãy chuẩn bị và cài đặt PYTHON phiên bản 3.10.x, 3.11.x hoặc 3.12.x và tạo môi trường ảo.
- Sau quá trình cài đặt, bạn có thể xóa thư mục `.uv` bên trong thư mục `assets` để giảm nhẹ dung lượng.

1. Tạo Môi Trường Ảo:

```

...

python -m venv env
mkdir -p assets/.uv
set UV_CACHE_DIR="assets/.uv"
env\Scripts\python.exe -m pip install uv
env\Scripts\python.exe -m uv pip install six packaging python-dateutil platformdirs
pywin32 wget
env\Scripts\python.exe -c "from main.app.install import
install_ffmpeg;install_ffmpeg()"

```

...

2. Cài Đặt Theo Từng Phần Cứng:

Tiếp tục chạy các lệnh này để tiếp tục cài đặt theo từng phần cứng.

Đối với CPU (Sử dụng CPU)

```

...

env\Scripts\python.exe -c "from main.app.install import
remove_onnxruntime;remove_onnxruntime()"

```

```
env\Scripts\python.exe -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/cpu
```

```
env\Scripts\python.exe -m uv pip install onnxruntime
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

```
...
```

Đối với CUDA (Sử dụng GPU NVIDIA)

- Đối với Cuda 11.8 (Dành cho GPU 10-Series trở lên)

```
...
```

```
env\Scripts\python.exe -c "from main.app.install import
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install numpy==1.26.4 numba==0.61.0
```

```
env\Scripts\python.exe -m uv pip install torch==2.7.0 torchaudio==2.7.0 torchvision --
index-url https://download.pytorch.org/whl/cu118
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-gpu==1.20.1
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

```
...
```

- Đối với Cuda 12.1 (Dành cho GPU 20-Series đến 30-Series trở lên)

```
...
```

```
env\Scripts\python.exe -c "from main.app.install import
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install numpy==1.26.4 numba==0.61.0
```

```
env\Scripts\python.exe -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/cu121
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-gpu==1.20.1
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

```
...
```

- Đối với Cuda 12.8 (Dành cho GPU 30-Series đến 40-Series trở lên)

```
...
```

```
env\Scripts\python.exe -c "from main.app.install import
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/cu128
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-gpu
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

```
env\Scripts\python.exe -m uv pip install faiss-cpu==1.13.2
```

```
...
```

- Đối với Cuda 13.0 (Dành cho GPU 50-Series trở lên)

```
...
```

```
env\Scripts\python.exe -c "from main.app.install import  
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install torch torchaudio torchvision --index-url  
https://download.pytorch.org/whl/cu130
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-gpu
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

```
env\Scripts\python.exe -m uv pip install faiss-cpu==1.13.2
```

```
...
```

- Nếu bạn muốn sử dụng TensorRT cho ONNXRUNTIME:

- Cài đặt Runtime [TensorRT](#) từ Nvidia, giải nén và thêm đường dẫn thư mục bin của nó vào [PATH](#) hệ thống.

- Tiếp theo cài đặt:

```
...
```

```
env\Scripts\python.exe -m uv pip install tensorrt
```

```
...
```

- Nếu bạn muốn sử dụng Compile để biên dịch mô hình:

- Cài đặt thư viện:

```
...
```

```
env\Scripts\python.exe -m uv pip install triton-windows
```

```
...
```

- Lưu Ý:

Hãy lựa chọn đúng phiên bản đối với GPU của bạn, nếu không nó có thể gây ra lỗi liên quan đến suy luận hoặc huấn luyện.

TensorRT khá là không ổn định nên là không khuyến khích cài đặt và sử dụng.

Compile có thể không thực sự cần thiết, nó chỉ thực sự hữu dụng khi bạn cần suy luận theo lô hoặc suy luận với các đầu vào lớn.

Đối với OPENCL (Sử dụng đối với GPU hỗ trợ OPENCL, có thể là IGPU, AMD, INTEL, NVIDIA)

```
...
```

```
env\Scripts\python.exe -c "from main.app.install import  
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install numpy==1.26.4 numba==0.61.0
```

```
env\Scripts\python.exe -m uv pip install torch==2.6.0 torchaudio==2.6.0 torchvision
```

```
env\Scripts\python.exe -m uv pip install https://github.com/artiom-  
beilis/pytorch_dlprim/releases/download/0.2.0/pytorch_ocl-0.2.0+torch2.6-cp311-  
none-win_amd64.whl
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-directml
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

...

- Lưu ý:

Có vẻ như OPENCL đã không còn được hỗ trợ tiếp.

OPENCL không hỗ trợ quá tốt tác vụ đa luồng nên khi chạy trích xuất thường sẽ bị khóa ở 1 luồng.

Chỉ nên cài đặt trên python 3.11 do không có bản biên dịch cho python 3.10 với torch 2.6.0.

Demucs có thể gây quá tải và tràn bộ nhớ đối với GPU (nếu cần sử dụng demucs hãy mở tệp config.json trong main\configs sửa đổi số demucs_cpu_mode thành true).

DDP không hỗ trợ huấn luyện đa GPU đối với OPENCL.

Một số thuật toán khác phải chạy trên cpu nên có thể hiệu suất của GPU có thể không sử dụng hết.

Đối với DIRECTML (Sử dụng đối với GPU hỗ trợ DIRECTML, có thể là IGPU, AMD, INTEL, NVIDIA)

...

```
env\Scripts\python.exe -c "from main.app.install import  
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install numpy==1.26.4 numba==0.61.0
```

```
env\Scripts\python.exe -m uv pip install torch==2.4.1 torchaudio==2.4.1 torchvision
```

```
env\Scripts\python.exe -m uv pip install torch-directml==0.2.5.dev240914
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-directml
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

...

- Lưu ý:

DIRECTML đã ngừng phát triển một khoảng thời gian dài.

DIRECTML không hỗ trợ quá tốt tác vụ đa luồng nên khi chạy trích xuất thường sẽ bị khóa ở 1 luồng.

DIRECTML có hỗ trợ 1 phần fp16 nhưng không được khuyến khích sử dụng vì có thể chỉ nhận được hiệu năng tương đương fp32.

DIRECTML không có hàm để dọn dẹp bộ nhớ, tôi đã tạo 1 hàm đơn giản để dọn dẹp bộ nhớ nhưng có thể sẽ không quá hiệu quả.

DIRECTML được thiết kế để suy luận chứ không phải dùng để huấn luyện mặc dù có thể hoàn toàn chạy được huấn luyện nhưng sẽ không được khuyến khích.

Đối với XPU (Sử dụng đối với GPU INTEL)

- Trước tiên nếu như bạn muốn sử dụng ONNXRUNTIME OPENVINO hãy cài đặt [OpenVino Toolkit](#) giải nén và thêm Release từ `\runtime\bin\intel64\Release` vào [PATH](#) hệ thống.

- Đối với sử dụng ONNXRUNTIME OPENVINO

...

```
env\Scripts\python.exe -c "from main.app.install import  
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install torch torchaudio torchvision --index-url  
https://download.pytorch.org/whl/xpu
```

```
env\Scripts\python.exe -m uv pip install openvino==2025.4.1
```

```
env\Scripts\python.exe -m uv pip install triton-windows
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-openvino==1.24.1
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

```
env\Scripts\python.exe -m uv pip install faiss-cpu==1.13.2
```

...

- Đối với sử dụng ONNXRUNTIME DIRECTML

...

```
env\Scripts\python.exe -c "from main.app.install import  
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install torch torchaudio torchvision --index-url  
https://download.pytorch.org/whl/xpu
```

```
env\Scripts\python.exe -m uv pip install triton-windows
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-directml
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

```
env\Scripts\python.exe -m uv pip install faiss-cpu==1.13.2
```

...

- Lưu ý:

- ONNXRUNTIME OPENVINO gặp lỗi với kích thước động nên là khuyên không sử dụng OPENVINO.

- ONNXRUNTIME DIRECTML có thể sẽ chậm hơn cả sử dụng mô hình PYTORCH thuần nên là cũng không quá khuyên sử dụng các mô hình ONNX.

- XPU không hỗ trợ DDP, nên là bạn sẽ không thể sử dụng huấn luyện đa GPU.

- XPU không hỗ trợ kiểu dữ liệu FP64 nên lớp GradScaler không hoạt động, lớp này đã được điều chỉnh và ép kiểu về FP32 có thể mất một chút chính xác.

Đối với ZLUDA (Sử dụng đối với GPU AMD hỗ trợ ROCm)

Kiểm tra GPU của bạn có được hỗ trợ hay không: [ROCM-Requirements](#).

Tải và cài đặt: [VC++ Runtime](#) và [HIP-SDK](#).

Thêm thư mục bin từ HIP-SDK vào [PATH](#) hệ thống.

... ..

```
env\Scripts\python.exe -c "from main.app.install import  
remove_onnxruntime;remove_onnxruntime()"
```

```
env\Scripts\python.exe -m uv pip install numpy==1.26.4 numba==0.61.0
```

```
env\Scripts\python.exe -m uv pip install torch==2.7.0 torchaudio==2.7.0 torchvision --  
index-url https://download.pytorch.org/whl/cu118
```

```
env\Scripts\python.exe -m uv pip install onnxruntime-gpu==1.20.1
```

```
env\Scripts\python.exe -m uv pip install -r requirements.txt
```

... ..

Sao chép path-zluda-hipxx.bat (Thay xx theo phiên bản HIP SDK của bạn) và run_app.bat ra thư mục chính thay thế các tệp hiện tại.

Chạy tệp path-zluda-hipxx.bat.

-Lưu ý:

Zluda hoạt động bằng cách biên dịch mã cuda sang hip sdk và quá trình này diễn ra cực kỳ chậm và trong lúc đó gpu của bạn sẽ không được sử dụng.

Nếu GPU của bạn không nằm trong danh sách hỗ trợ (ví dụ như: gfx803) bạn có thể thử dùng HIP SDK 5.7.0 và ghi đè thư mục library trong ROCm/5.7.0/bin/roclblas từ [ROCMLibs](<https://github.com/likelovewant/ROCMLibs-for-gfx1103-AMD780M-APU/releases>) (Không Khuyến Khích).

Nếu Onnxruntime không hoạt động, bạn có thể thử đổi sang phiên bản onnxruntime-directml.

Zluda không được thiết kế cho các hệ thống thời gian thực nên chức năng thời gian thực của ứng dụng sẽ bị vô hiệu hóa.

Bạn có thể thử Onnxruntime ROCm nhưng nó chỉ hoạt động trên python 3.10.x hoặc 3.12.x.

```

```
env\Scripts\python.exe -m pip uninstall onnxruntime-gpu
env\Scripts\python.exe -m uv pip install onnxruntime-rocm
```

```

Đối Với Hệ Điều Hành Linux

1. Thiết Lập Môi Trường Ảo:

Hãy thay phiên bản PYTHON 3.11.9 thành 3.12.0 nếu như bạn dùng ROCm.

```

```
sudo apt update -y

sudo apt install -y curl git build-essential libssl-dev zlib1g-dev libbz2-dev libreadline-
dev libsqlite3-dev wget llvm libncurses5-dev xz-utils tk-dev libxml2-dev libxmlsec1-dev
libffi-dev liblzma-dev

curl https://pyenv.run | bash

export PATH="$HOME/.pyenv/bin:$PATH"

eval "$(pyenv init -)"

eval "$(pyenv virtualenv-init -)"

source ~/.bashrc

pyenv local --unset

pyenv shell --unset

pyenv install 3.11.9

pyenv global 3.11.9

python -m pip install uv

python -m uv pip install six packaging python-dateutil platformdirs wget
```

```

2. Cài Đặt Theo Từng Phần Cứng:

Tiếp tục chạy các lệnh này để tiếp tục cài đặt theo từng phần cứng.

Đối với CPU (Sử dụng CPU)

```

```
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"

python -m uv pip install torch==2.7.0 torchaudio==2.7.0 torchvision --index-url
https://download.pytorch.org/whl/cpu
```

```
python -m uv pip install onnxruntime
python -m uv pip install -r requirements.txt
...
```

### **Đối với CUDA (Sử dụng GPU NVIDIA)**

#### **- Cài đặt Cuda Toolkit.**

```
...
sudo apt update
sudo apt install nvidia-cuda-toolkit
...
```

#### **- Đối với Cuda 11.8 (Dành cho GPU 10-Series trở lên)**

```
...
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
python -m uv pip install numpy==1.26.4 numba==0.61.0
python -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/cu118
python -m uv pip install onnxruntime-gpu==1.20.1
python -m uv pip install -r requirements.txt
...
```

#### **- Đối với Cuda 12.1 (Dành cho GPU 20-Series đến 30-Series trở lên)**

```
...
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
python -m uv pip install numpy==1.26.4 numba==0.61.0
python -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/cu121
python -m uv pip install onnxruntime-gpu==1.20.1
python -m uv pip install -r requirements.txt
...
```

#### **- Đối với Cuda 12.8 (Dành cho GPU 30-Series đến 40-Series trở lên)**

```
...
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
python -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/cu128
python -m uv pip install onnxruntime-gpu
```

```
python -m uv pip install -r requirements.txt
```

```
python -m uv pip install faiss-cpu==1.13.2
```

```
```
```

- Đối với Cuda 13.0 (Dành cho GPU 50-Series trở lên)

```
```
```

```
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
```

```
python -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/cu130
```

```
python -m uv pip install onnxruntime-gpu
```

```
python -m uv pip install -r requirements.txt
```

```
python -m uv pip install faiss-cpu==1.13.2
```

```
```
```

- Nếu bạn muốn sử dụng TensorRT cho ONNXRUNTIME:

- Cài đặt Runtime [TensorRT](#) từ Nvidia. Giải nén và đặt biến môi trường:

```
```
```

```
tar -xvf TensorRT-*.tar.gz
```

```
cd TensorRT-*
```

```
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$(pwd)/lib
```

```
export PATH=$PATH:$(pwd)/bin
```

```
```
```

- Tiếp theo cài đặt:

```
```
```

```
python -m uv pip install tensorrt
```

```
```
```

- Nếu bạn muốn sử dụng Compile để biên dịch mô hình:

- Cài đặt thư viện:

```
```
```

```
python -m uv pip install triton
```

```
```
```

- Lưu Ý:

Hãy lựa chọn đúng phiên bản đối với GPU của bạn, nếu không nó có thể gây ra lỗi liên quan đến suy luận hoặc huấn luyện.

TensorRT khá là không ổn định nên là không khuyến khích cài đặt và sử dụng.

Compile có thể không thực sự cần thiết, nó chỉ thực sự hữu dụng khi bạn cần suy luận theo lô hoặc suy luận với các đầu vào lớn.

Đối với ROCm (Sử dụng đối với GPU AMD hỗ trợ RDNA)

- Đối với ROCm 5.7

```

```
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
```

```
python -m uv pip install numpy==1.26.4 numba==0.61.0
```

```
python -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/rocm5.7
```

```
python -m uv pip install onnxruntime-rocm
```

```
python -m uv pip install -r requirements.txt
```

```

- Đối với ROCm 6.4

```

```
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
```

```
python -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/rocm6.4
```

```
python -m uv pip install onnxruntime-rocm
```

```
python -m uv pip install -r requirements.txt
```

```
python -m uv pip install faiss-cpu==1.13.2
```

```

- Đối với ROCm 7.2

```

```
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
```

```
python -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/rocm7.2
```

```
python -m uv pip install onnxruntime-rocm
```

```
python -m uv pip install -r requirements.txt
```

```
python -m uv pip install faiss-cpu==1.13.2
```

```

- Lưu Ý:

Sử dụng ROCm trên dự án này không thực sự được khuyến nghị vì tôi chưa thử nghiệm được ROCm.

ROCm cũng khá là kén GPU AMD, nên là hãy chọn cho đúng phiên bản cho GPU của bạn.

Hãy thử nghiệm bằng `python -c "import torch; print(torch.cuda.is_available())"` nếu như trả về False nghĩa là cài đặt thất bại.

Đối với XPU (Sử dụng đối với GPU INTEL)

- Trước tiên nếu như bạn muốn sử dụng ONNXRUNTIME OPENVINO hãy cài đặt [OpenVino Toolkit](#) giải nén và đặt biến môi trường hoặc chạy `setupvars.sh` trong cùng một phiên.

- Cài đặt môi trường cho GPU Intel:

```

```
sudo apt-get update
```

```
sudo apt-get install -y software-properties-common
```

```
sudo add-apt-repository -y ppa:kobuk-team/intel-graphics
```

```
sudo apt-get install -y libze-intel-gpu1 libze1 intel-metrics-discovery intel-opencl-icd
clinfo intel-gsc
```

```
sudo apt-get install -y intel-media-va-driver-non-free libmfx-gen1 libvpl2 libvpl-tools
libva-glx2 va-driver-all vainfo
```

```
sudo apt-get install -y libze-dev intel-ocloc
```

```

- Đối với sử dụng ONNXRUNTIME OPENVINO

```

```
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
```

```
python -m uv pip install torch torchaudio torchvision --index-url
https://download.pytorch.org/whl/xpu
```

```
python -m uv pip install openvino==2025.4.1
```

```
python -m uv pip install triton
```

```
python -m uv pip install onnxruntime-openvino==1.24.1
```

```
python -m uv pip install -r requirements.txt
```

```
python -m uv pip install faiss-cpu==1.13.2
```

```

- Đối với sử dụng ONNXRUNTIME CPU

...

```
python -c "from main.app.install import remove_onnxruntime;remove_onnxruntime()"
```

```
python -m uv pip install torch torchaudio torchvision --index-url  
https://download.pytorch.org/whl/xpu
```

```
python -m uv pip install triton
```

```
python -m uv pip install onnxruntime
```

```
python -m uv pip install -r requirements.txt
```

```
python -m uv pip install faiss-cpu==1.13.2
```

...

- Lưu ý:

ONNXRUNTIME OPENVINO gặp lỗi với kích thước động nên là khuyên không sử dụng OPENVINO.

ONNXRUNTIME CPU khá chậm, cẩn thận với việc sử dụng mô hình ONNX.

XPU không hỗ trợ DDP, nên là bạn sẽ không thể sử dụng huấn luyện đa GPU.

XPU không hỗ trợ kiểu dữ liệu FP64 nên lớp GradScaler không hoạt động, lớp này đã được điều chỉnh và ép kiểu về FP32 có thể mất một chút chính xác.

Hướng Dẫn Cơ Bản:

► TÍNH NĂNG TÁCH NHẠC:

- **Bước 1:** Hãy nhập đường dẫn liên kết vào ``Đường dẫn liên kết đến âm thanh`` và bấm Tải Xuống hoặc tải lên tệp âm thanh vào mục ``Thả âm thanh vào đây``.
- **Bước 2:** Hãy chọn ``Tách giọng bè``, ``Tách vang`` nếu như bạn muốn thêm bước tách bè và vang. (``Tách vang`` được đề xuất cho việc chuyển đổi giọng nói).
- **Bước 3:** Chọn các mô hình tách nhạc cho quá trình tách (Nếu mô hình cho bật ``khử tách nhạc`` hãy bật nó):
 - Demucs có tiêu thụ rất lớn về GPU nhưng chất lượng tốt.
 - VR-ARCH tạo ra giọng sạch, ít nhiễu và chạy nhanh.
 - MDX-NET cân bằng giữa chất lượng và tốc độ.
- **Bước 4:** Cài đặt các tham số, đối với người mới hãy để các tham số ở mức mặc định và bấm nút ``Tách Nhạc`` để bắt đầu tách.
- **Bước 5:** Sau khi tách nhạc hoàn tất, bạn có thể nghe thử tại các phần nghe thử trên giao diện.

► TÍNH NĂNG CHUYỂN ĐỔI ÂM THANH:

- **Bước 1:** Bấm `Tải lại` trong phần `mô hình và chỉ mục`, nếu như khi bấm vào phần `Tệp mô hình` không hiển thị. Sau khi có mô hình hãy chọn mô hình.
- **Bước 2:** Chọn đầu vào âm thanh:
 - o Để dùng lại các tệp âm thanh đã xử lý trước đó, hãy chọn "**Sử dụng âm thanh vừa tách**" và cấu hình các tùy chọn sau:
 - **Chuyển đổi giọng gốc:** Bật khi chỉ muốn xử lý giọng chính (không có bè).
 - **Chuyển đổi giọng bè:** Bật nếu bạn đã tách bè và muốn chuyển đổi luôn cả phần giọng bè đó.
 - **Không kết hợp giọng bè:** Bật nếu bạn muốn giữ riêng phần bè, không tự động ghép vào bản cuối.
 - **Kết hợp nhạc nền:** Bật nếu muốn hệ thống tự động ghép lại với nhạc sau khi chuyển đổi.
 - o Tải lên tệp âm thanh cần chuyển đổi giọng vào mục `Thả âm thanh vào đây`.
 - o Nếu cần chuyển đổi nhiều tệp cùng lúc, hãy tạo một thư mục chứa các tệp cần chuyển đổi và dán đường dẫn thư mục đó vào phần `Đường dẫn đầu vào âm thanh`.
- **Bước 3:** Cấu hình lại các tham số như `Dịch chuyển cao độ` và một số tham số khác.
 - o `Dịch chuyển cao độ`: Nếu đầu vào là giọng Nam và mô hình là giọng Nữ hãy tăng lên 12 và đầu vào là giọng Nữ và mô hình là giọng Nam thì hãy giảm xuống -12.
 - o `Bảo vệ phụ âm`: Bảo vệ các phụ âm gió (s, x, t...) để tránh tiếng bị méo thành tiếng điện tử. Giá trị đề xuất: **0.33**.
- **Bước 4:** Bấm nút `Chuyển Đổi Âm Thanh` để bắt đầu quá trình chuyển đổi âm thanh. Sau khi quá trình chuyển đổi hoàn tất, bạn có thể nghe thử tại các phần nghe thử trên giao diện.

► TÍNH NĂNG CHUYỂN ĐỔI VĂN BẢN:

- **Bước 1:** Bấm `Tải lại` trong phần `mô hình và chỉ mục`, nếu như khi bấm vào phần `Tệp mô hình` không hiển thị. Sau khi có mô hình hãy chọn mô hình.
- **Bước 2:** Nhập văn bản đầu vào, bấm `1. Chuyển Đổi Văn Bản` và đợi hoàn tất.
- **Bước 3:** Cấu hình lại các tham số như `Dịch chuyển cao độ` và một số tham số khác.
 - o `Dịch chuyển cao độ`: Nếu đầu vào là giọng Nam và mô hình là giọng Nữ hãy tăng lên 12 và đầu vào là giọng Nữ và mô hình là giọng Nam thì hãy giảm xuống -12.
 - o `Bảo vệ phụ âm`: Bảo vệ các phụ âm gió (s, x, t...) để tránh tiếng bị méo thành tiếng điện tử. Giá trị đề xuất: **0.33**.
- **Bước 4:** Bấm nút `2. Chuyển Đổi Giọng Nói` để bắt đầu quá trình chuyển đổi âm thanh. Sau khi quá trình chuyển đổi hoàn tất, bạn có thể nghe thử tại các phần nghe thử trên giao diện.

► TÍNH NĂNG CHUYỂN ĐỔI GIỌNG NÓI THỜI GIAN THỰC:

- **Bước 1: Thiết Lập Thiết Bị Đầu Vào & Đầu Ra**
 - o Tại mục Thiết bị âm thanh, chọn Micro (Input) và Loa/Tai nghe (Output).
 - o Lưu ý: Ưu tiên chọn các thiết bị có nhãn WASAPI để đạt tốc độ phản hồi nhanh nhất và độ trễ thấp nhất.
- **Bước 2: Chọn Mô Hình & Chỉ Mục**
 - o Bấm Tải lại nếu danh sách mô hình trống. Chọn tệp mô hình RVC/SVC bạn muốn sử dụng.
- **Bước 3: Cấu Hình Tham Số Âm Thanh**
 - o Tùy chỉnh các chỉ số quan trọng để tối ưu chất lượng:
 - Dịch chuyển cao độ: Nếu từ Nam sang Nữ hãy tăng lên 12, từ Nữ sang Nam hãy giảm xuống -12.
 - Kích thước đoạn: Thiết lập dựa trên cấu hình máy. Số càng nhỏ độ trễ càng thấp nhưng dễ gây giật lag nếu máy yếu.

- Làm sạch âm thanh: Kích hoạt để lọc bỏ tiếng ồn nền (quạt, tiếng xe) giúp lỗi xử lý giọng nói chuẩn xác hơn.

- **Bước 4: Kích Hoạt & Kiểm Tra**

- Bật tính năng Phát hiện giọng nói VAD để hệ thống chỉ hoạt động khi bạn nói (tránh tiếng nhiễu khi im lặng).
- Bắt đầu nói vào Micro và nghe âm thanh đã được chuyển đổi trả về qua thiết bị nghe lại (Loa/Tai nghe) ngay lập tức.

► **TÍNH NĂNG HUẤN LUYỆN MÔ HÌNH:**

- **Bước 1: Thiết Lập Cơ Bản & Kiến Trúc**

- Đặt Tên mô hình (không dấu, không khoảng trắng) và chọn Tốc độ lấy mẫu phù hợp với chất lượng dữ liệu đầu vào.
- Lưu ý: Ưu tiên chọn Phiên bản v2 và kiến trúc RVC để tối ưu tốc độ xử lý. Luôn bật Huấn luyện cao độ nếu bạn muốn mô hình có thể hát hoặc nói có ngữ điệu.

- **Bước 2: Chuẩn Bị & Tiền Xử Lý Dữ Liệu**

- Tải dữ liệu âm thanh lên và chọn chế độ Làm sạch dữ liệu để loại bỏ tạp âm.
- Sử dụng tính năng Tùy chỉnh cắt âm thanh (Automatic/Slicer2) để chia nhỏ các tệp dài thành các đoạn 3-5 giây. Việc này giúp mô hình học nhanh và chính xác hơn.
- Mẹo: Đừng quên Chuẩn hóa âm lượng để âm thanh đầu ra không bị lúc to lúc nhỏ.

- **Bước 3: Trích Xuất Đặc Trưng (F0 & Embedding)**

- Chọn Phương pháp trích xuất F0 (Khuyến dùng RMVPE).
- Tại mục Chế độ nhúng, hãy giữ mặc định Fairseq với mô hình ContentVec để tận dụng các bộ tiền huấn luyện (Pre-trained) sẵn có, giúp tiết kiệm thời gian huấn luyện.

- **Bước 4: Cấu Hình Thông Số Huấn Luyện**

- Tổng số kỷ nguyên (Epoch): Đặt số lượng vòng lặp (ví dụ: 200-500).
- Kích thước lô (Batch size): Điều chỉnh dựa trên dung lượng VRAM của card đồ họa. Nếu máy bị treo hoặc báo lỗi bộ nhớ, hãy giảm chỉ số này xuống.
- Trình tối ưu hóa: Sử dụng AdamW cho sự ổn định hoặc AdaBelief nếu bạn muốn mô hình hội tụ nhanh và rõ chữ hơn.

- **Bước 5: Theo Dõi & Xuất Mô Hình**

- Kiểm tra trạng thái qua tab Thông tin huấn luyện hoặc Terminal.
- Lưu ý: Luôn bật Lưu mọi mô hình hoặc đặt Tần suất lưu phù hợp để tránh mất dữ liệu nếu xảy ra sự cố nguồn điện hoặc phần cứng.
- Sau khi hoàn tất, hãy tạo Tệp chỉ mục (Index) để giọng nói AI giữ được hồn và sắc thái đặc trưng của người thật.

► **TÍNH NĂNG TẢI XUỐNG MÔ HÌNH:**

- **Bước 1: Chọn Phương Thức Tải**

- Tại mục Chọn cách tải mô hình, hãy xác định nguồn bạn có:
 - Sử dụng Đường dẫn liên kết nếu bạn có link tải trực tiếp.
 - Sử dụng Kho mô hình CSV hoặc Tìm kiếm nếu muốn khám phá các mẫu từ cộng đồng.
 - Chọn Tải lên nếu bạn đã có sẵn file mô hình (.zip, .pth) trong máy.

- **Bước 2: Tải Từ Link Liên Kết (HuggingFace, Drive,...)**

- Dán đường dẫn vào ô Đường dẫn liên kết đến mô hình.
- Hệ thống hỗ trợ hầu hết các nền tảng lưu trữ lớn như: HuggingFace, Google Drive, Mediafire, Pixeldrain và Mega.nz.
- Bạn có thể đặt lại Tên mô hình để dễ quản lý trong danh sách.

- **Bước 3: Khám Phá Kho Mô Hình Cộng Đồng**

- Từ kho CSV: Chọn tên nhân vật bạn thích trong danh sách và nhấn Nhận mô hình.
- Từ công cụ Tìm kiếm: Nhập tên nhân vật (ví dụ: tên ca sĩ, diễn viên) và nhấn Tìm kiếm. Sau khi kết quả hiện ra, chỉ cần click vào để hệ thống tự động tải về.

- Bước 4: Tải Lên Trực Tiếp Từ Máy Tính

- Nếu bạn nhận file từ bạn bè hoặc tải thủ công từ trước, hãy chọn mục Tải lên.
- Kéo và thả file mô hình vào vùng Thả mô hình vào đây để hệ thống tự động giải nén và đưa vào danh sách sử dụng.

Giải Thích Nâng Cao:

► CẤU HÌNH TÍNH NĂNG TÁCH NHẠC:

Mô tả: Sử dụng một số mô hình tách nhạc đơn giản để tách giọng và nhạc ra riêng để hỗ trợ việc chuyển đổi giọng nói.

Lưu ý: Mô hình Demucs chỉ được tối ưu tốt trên CPU thay vì GPU nên rất dễ gây ra lỗi hết bộ nhớ RAM và VRAM. Mô hình VR thì có thể cho ra thời lượng đầu ra nhỏ hơn tệp đầu vào nếu bị lệch tốc độ lấy mẫu.

1. Nhóm Tính Năng Bổ Trợ:

- **Khử nhiễu tách nhạc:** Áp dụng thêm bước lọc sau khi tách để âm thanh sạch hơn.

- *Mô hình MDX:* Tự động khử nhiễu.
- *Mô hình VR:* Cho phép chọn thêm một model khử nhiễu riêng biệt.
- *Lưu ý:* Không khả dụng với Demucs.

- **Tách giọng bè (Backing Vocals):** Tách riêng phần bè ra khỏi giọng chính.

- *Trạng thái:* Nếu tắt, đầu ra là **Original Vocals**. Nếu bật, đầu ra chia làm **Main Vocals** và **Backing Vocals**.

- **Tách vang (Reverb/Echo):** Tách tiếng vang từ giọng hát để hỗ trợ suy luận AI (VC) tốt hơn.

- *Lưu ý:* Với model VR, có thể chọn giữa tách tiếng Vang (Reverb) hoặc tiếng Vọng (Echo).

- **Tăng cường suy luận (VR Only):** Kéo dài thời gian kiểm thử để đạt chất lượng tách cao nhất. Chỉ dành cho model VR.

- **Xử lý dải cao (VR Only):** Tối ưu hóa các tần số cao ở đầu ra, giúp âm thanh sáng và rõ hơn. Chỉ dành cho model VR.

- **Hậu xử lí (VR Only):** Thêm một bước hậu xử lí ở đầu ra, giúp giảm tạp âm và cải thiện chất lượng. Chỉ dành cho model VR.

2. Nhóm mô hình xử lý:

Phần mô hình tách nhạc hỗ trợ 3 dòng mô hình chính: **MDX-NET**, **VR-ARCH**, và **Demucs**.

1. Mô hình Tách nhạc chính:

- **Tính năng:** Là bộ não chính thực hiện việc tách Giọng hát và Nhạc nền. Chất lượng của bước này ảnh hưởng trực tiếp đến tất cả các bước xử lý sau đó.
- **Cơ chế hiển thị:**
 - **Dropdown Model:** Cho phép chọn giữa các dòng **VR**, **MDX**, **Demucs**.
 - **Tùy biến theo dòng:** Khi chọn mỗi dòng, các thanh trượt tham số (Kích thước lô, Phân đoạn, Cửa sổ...) sẽ tự động thay đổi tương ứng như đã mô tả ở phần trước.

2. Mô hình Tách bè:

- **Tính năng:** Thực hiện tách lớp thứ hai sau khi đã có tệp Original Vocals, giúp lọc riêng giọng hát chính và đẩy phần bè sang một tệp riêng.
- **Cơ chế hiển thị:** Chỉ xuất hiện khi người dùng **Tích chọn (Enable)** tính năng "Tách giọng bè".

3. Mô hình Tách vang:

- **Tính năng:** Loại bỏ các hiệu ứng môi trường (Echo/Reverb) để giọng hát trở nên "khô" (Dry), cực kỳ quan trọng nếu bạn dùng giọng này để training hoặc suy luận AI (RVC).
- **Cơ chế hiển thị:** * Chỉ xuất hiện khi **Tích chọn** tính năng "Tách vang".
 - **Lưu ý đặc biệt:** Đối với dòng **VR**, hệ thống sẽ cung cấp thêm các model chuyên sâu để phân biệt giữa **Tiếng Vang (Reverb)** và **Tiếng Vọng (Echo)**.

4. Mô hình Khử nhiễu:

- **Tính năng:** Bước làm sạch cuối cùng để loại bỏ các tạp âm nhỏ phát sinh trong quá trình tách.
- **Cơ chế hiển thị:** Chỉ hiển thị khi chọn mô hình chính thuộc dòng **VR** và có bật "Khử tách nhạc".

3. Giải Thích Tham Số Kỹ Thuật

- **Số lượng dự đoán (Shifts):** (Chỉ Demucs) Số lần dự đoán lại để lấy trung bình. Càng cao càng sạch nhưng cực kỳ tốn phần cứng và chậm.
- **Kích thước lô (Batch Size):** Số lượng mẫu xử lý cùng lúc. Tăng để nhanh hơn nhưng dễ gây tràn bộ nhớ (VRAM).
- **Kích thước phân đoạn (Segment Size):** Độ dài đoạn âm thanh được cắt ra để xử lý. Giá trị lớn giúp giảm vết nối âm thanh nhưng đòi hỏi RAM cao.
- **Mức độ mạnh tay (Aggressive):** (Chỉ VR) Kiểm soát lực tách. Thấp thì giữ chi tiết nhưng dễ lẫn nhạc; Cao thì sạch nhạc nhưng dễ méo giọng.
- **Chồng chéo (Overlap):** Phần giao nhau giữa các phân đoạn để khi ghép lại âm thanh không bị ngắt quãng, tạo sự mượt mà.
- **Kích thước cửa sổ (Window Size):** (Chỉ VR) Kích thước cắt nhỏ trước khi đưa vào AI. Càng nhỏ thì AI nhìn nhận càng chi tiết, càng lớn thì tốc độ xử lý càng nhanh.
- **Độ dịch chuyển khung (Hop Length):** (Chỉ MDX) Quyết định độ phân giải thời gian của phổ âm thanh. Giá trị thấp cho độ chi tiết cao hơn.
- **Ngưỡng hậu xử lý (Post Process):** (Chỉ VR) Quyết định mức độ xử lý hậu kỳ của mô hình.

4. Thiết Lập Đầu Vào & Đầu Ra

- **Tải lên âm thanh:** Chọn tệp tin âm thanh trực tiếp từ thiết bị của bạn (hỗ trợ .mp3, .wav, .flac...).
- **Sử dụng đường dẫn YouTube:** Dán link video/audio từ YouTube, hệ thống sẽ tự động tải về và đưa vào quy trình tách.
- **Tốc độ lấy mẫu:** Tùy chọn tốc độ lấy mẫu của đầu ra, đối với một số mô hình có thể gây ra hiện tượng lệch tốc độ âm thanh.
- **Định dạng âm thanh:** Lựa chọn định dạng tệp xuất ra ví dụ như wav.
- **Đường dẫn đầu vào:** Lựa chọn tệp âm thanh đầu vào được sử dụng để tách nhạc, có thể sử dụng đường dẫn bên ngoài.
- **Đường dẫn thư mục đầu ra:** Lựa chọn thư mục sẽ chứa các tệp tải về từ Youtube, âm thanh sau khi tách nhạc (Giọng, Nhạc).
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử kết quả ngay sau khi xử lý xong mà không cần phải mở đến tệp để nghe thử.

► CẤU HÌNH TÍNH NĂNG CHUYỂN ĐỔI ÂM THANH:

Mô tả: Sử dụng để chuyển đổi giọng nói phi thời gian thực. Có thêm một số cài đặt tùy chỉnh giúp bạn chọn nhanh các tệp đã tách trước đó để sử dụng. Hỗ trợ mô hình RVC và SVC, có thêm một số các cài đặt bổ sung so với dự án gốc.

Lưu ý: Việc sử dụng các tối ưu ONNX có thể gây hiệu ứng ngược khi làm giảm chất lượng đầu ra, nhiều hoặc lỗi trên các đoạn âm thanh dài, nó chỉ được khuyến cáo sử dụng trên chuyển đổi giọng nói thời gian thực.

1. Thiết Lập Đầu Vào & Đầu Ra

- **Làm sạch âm thanh:** Kích hoạt bộ lọc nhiễu. Khi chọn, thanh trượt xuất hiện để bạn điều chỉnh mức độ lọc tạp âm.
- **Mức độ làm sạch âm thanh:** Điều chỉnh cường độ lọc. Giá trị cao lọc sạch hơn nhưng dễ mất chi tiết giọng; giá trị thấp giữ được ngữ âm tự nhiên nhưng có thể còn tạp âm.
- **Tải lên âm thanh:** Chọn tệp trực tiếp từ thiết bị (hỗ trợ .mp3, .wav, .flac...).
- **Sử dụng âm thanh vừa tách:** Dừng lại các bản tách nhạc trước đó. Nếu chỉ có 1 bản, hệ thống tự chuyển đổi; nếu có nhiều bản, bạn sẽ được chọn thủ công.
- **Đường dẫn đầu vào/đầu ra:** Chỉ hiện khi không dùng tính năng "Sử dụng âm thanh vừa tách", dùng để chỉ định nơi lấy và lưu tệp.
 - Nếu bạn muốn chuyển đổi theo lô (hay còn gọi là chuyển đổi hàng loạt), hãy chuẩn bị một thư mục chứa các tệp đầu vào và cung cấp đầu vào là thư mục đó, nó sẽ thực hiện chuyển đổi giọng theo lô thay vì chỉ 1 tệp duy nhất.
- **Cắt âm thanh:** Tự động chia nhỏ âm thanh dài để tránh lỗi bộ nhớ và lỗi trích xuất cao độ, sau đó tự ghép lại.
- **Định dạng âm thanh:** Lựa chọn định dạng tệp xuất ra ví dụ như wav.

2. Thiết Lập Sử Dụng Các Bản Tách Sẵn.

- **Chuyển đổi giọng gốc (Original Vocals):** Nếu chọn, hệ thống dùng Original Vocals (cần thiết khi bạn không thực hiện tách bè). Nếu không chọn, hệ thống mặc định chuyển đổi giọng Main Vocals.
- **Chuyển đổi giọng bè:** Chuyển đổi cả giọng hát bè. Lưu ý: Dễ gây lỗi âm thanh đầu ra nên không khuyến khích sử dụng.
- **Không kết hợp giọng bè:** Ngăn việc tự động trộn giọng bè vào giọng chính sau khi chuyển đổi xong.
- **Kết hợp nhạc nền:** Tự động trộn giọng đã chuyển đổi với nhạc nền có sẵn.

- **Chọn bản tách:** Xuất hiện khi có nhiều bản tách nhạc để bạn chọn đúng tệp cần chuyển đổi.

3. Điều Chỉnh Cao Độ F0

- **Dịch chuyển cao độ:** Điều chỉnh giọng để khớp với mô hình (Ví dụ: Giọng nữ sang mô hình nam thì chỉnh -12 và ngược lại +12). Nếu cùng giới tính, có thể chỉnh nhẹ để tối ưu chất giọng.
- **Tự động đề xuất cao độ:** Tính toán để đưa cao độ về mức phù hợp nhất với mô hình giọng nói.
- **Ngưỡng tần số ước tính:** (Chỉ hiện khi dùng Tự động đề xuất) Thanh trượt để hệ thống tính toán mức độ kéo cao độ chính xác hơn, đề xuất sử dụng 155 trở xuống đối với mô hình giọng Nam và 255 trở lên đối với mô hình giọng nữ.
- **Tự động điều chỉnh:** Tự động nắn cao độ F0 theo các nốt nhạc nó hoạt động giống như Autotune.
- **Mức độ điều chỉnh:** (Chỉ hiện khi dùng Tự động điều chỉnh) Điều chỉnh mức độ can thiệp vào nốt nhạc của F0.
- **Chế độ F0 ONNX:** Sử dụng mô hình ONNX để tăng tốc trích xuất cao độ (thay vì Pytorch), nhưng có thể lỗi với tệp quá dài.
- **Tải lên tệp F0 / Tệp F0:** Cho phép bạn tải lên tệp cao độ đã được chỉnh sửa thủ công theo ý muốn thay vì để mô hình tự tính.

4. Mô Hình & Chỉ Mục

- **Tệp mô hình:** Chọn mô hình RVC hoặc SVC. Với SVC, cần dùng mô hình thông số gốc và chuyển đổi lại trong tab "Thêm -> Chuyển đổi mô hình SVC".
- **Tệp chỉ mục (Index):** Giúp giọng nói giống mẫu huấn luyện hơn, giảm lỗi âm thanh. Khi có tệp này, thanh trượt "Ảnh hưởng của chỉ mục" sẽ hiện ra (Có thể nó sẽ không cần thiết nếu mô hình được huấn luyện với thời lượng dữ liệu lớn chất lượng cao).
- **Ảnh hưởng của chỉ mục:** Điều chỉnh mức độ tác động của tệp Index. Quá cao sẽ gây giọng robot/giả; quá thấp sẽ không thấy tác dụng.
- **Mã nhận dạng giọng nói:** (Chỉ hiện đối với mô hình đa giọng nói) Dùng để chọn nhân vật cụ thể trong các mô hình được huấn luyện đa giọng nói.

5. Cài Đặt Chuyên Sâu

- **Phương pháp trích xuất:** Chọn thuật toán chính để trích xuất F0. Nếu chọn **Hybrid**, bạn có thể trộn nhiều phương pháp với nhau theo tỷ lệ tùy chỉnh.
- **Mở khóa tất cả:** Hiển thị toàn bộ các phương pháp trích xuất cao độ để nghiên cứu chuyên sâu.

- **Phương pháp trích xuất HYBRID:** Cho phép bạn kết hợp của hai hoặc nhiều loại trích xuất khác nhau. Ví dụ: hybrid[pm+dio].
- **Mức trộn của các phương thức:** Mức trộn các cao độ của các phương thức trích xuất với nhau.
- **Sử dụng hiệu quả bộ nhớ:** Tối ưu hóa để tránh tràn bộ nhớ nhưng thời gian chuyển đổi sẽ lâu hơn.
- **Chế độ nhúng (Embedding):**
 - Fairseq: Dùng cho các mô hình pytorch truyền thống.
 - ONNX: Dùng cho các mô hình onnx để tối ưu tốc độ nhúng.
 - Transformers: Dùng cho các mô hình hiện đại và an toàn hơn.
 - Spin: Chất lượng cao nhất nhưng cần huấn luyện riêng.
 - Whisper: Đang trong giai đoạn thử nghiệm.
- **Mô hình nhúng:** Quyết định chất lượng giọng (thường dùng Hubert hoặc ContentVec). Nếu dùng mô hình khác Hubert và Contentvec thì bạn sẽ cần huấn luyện lại mô hình từ đầu để mô hình có thể phát âm bình thường.
- **Tên của mô hình:** Khi sử dụng mô hình nhúng tùy chỉnh sẽ được hiện thị để bạn nhập tên của mô hình nhúng, ví dụ: Nếu là Fairseq đầu vào nên là hubert_base.pt, nếu là ONNX đầu vào nên là hubert_base.onnx và nếu là Transformers đầu vào nên là hubert_base (Nơi chứa tệp mô hình và tệp cấu hình mô hình).
- **Hòa trộn tầng:** Trộn các tầng của mô hình nhúng để tạo độ "rò" âm thanh nhẹ để cải thiện chất lượng đầu ra.
- **Tầng hòa trộn:**
 - Tầng thấp (1-6): Chi tiết âm học.
 - Tầng giữa (7-9): Thông tin ngữ âm (RVC v1 thường dùng tầng 9).
 - Tầng cao (10-12): Nội dung ngôn ngữ, sạch hơn, tránh rò âm gốc (RVC v2 dùng tầng 12).
- **Tỉ lệ hòa trộn:** Mức độ rò âm sắc. Quá cao gây lạc điệu, quá thấp sẽ không có tác dụng.

6. Xử Lý Hậu Kỳ & Định Dạng

- **Lọc trung vị:** Kiểm soát nhiễu, làm mượt cao độ và độ tự tin của mô hình trích xuất.
- **Độ dịch chuyển khung:** Quyết định khoảng cách giữa các frame phân tích, ảnh hưởng trực tiếp đến độ mượt, độ chính xác theo thời gian, độ trễ và hiệu năng xử lý của hệ thống âm thanh.

- **Tỷ lệ trộn RMS:** Trộn năng lượng từ âm thanh gốc vào đầu ra để giọng nói sống động hơn (giá trị bằng 1 là không trộn).
- **Bảo vệ phụ âm:** Giá trị nhỏ giữ rõ tiếng thờ/phụ âm; giá trị lớn ngăn tạp âm nhưng làm giọng bị bí (mất chi tiết phụ âm).
- **Tỉ lệ độ ồn (Chỉ dành cho SVC):** Điều chỉnh mức tạp âm trong mô hình. Nên giữ nguyên mặc định.
- **Dịch chuyển cao độ và âm sắc:** Thay đổi tần số phổ trước khi chuyển đổi.
- **Dịch chuyển định dạng (Formant):** Điều chỉnh độ sáng/trầm của âm sắc (Cao hơn = nữ tính, thấp hơn = nam tính).
- **Lấy mẫu lại (Resample):** Thay đổi tần số lấy mẫu của tệp âm thanh đầu ra. Hữu ích khi bạn cần đồng bộ chất lượng âm thanh với các tiêu chuẩn phần mềm hoặc phần cứng khác nhau.
- **Định dạng âm thanh:** Chọn định dạng xuất file (ví dụ: .wav).
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử kết quả ngay sau khi xử lý xong mà không cần phải mò đến tệp để nghe thử.

7. Thiết Lập Cài Đặt Sẵn

- **Tệp cài đặt trước:** Lưu lại các thông số bạn đã chỉnh để không phải làm lại từ đầu khi tải lại trang.
- **Các cài đặt sẽ được lưu:** Chọn các mục cụ thể muốn lưu vào tệp preset.
- **Tên khi lưu tệp:** Đặt tên tùy ý cho tệp cài đặt của bạn.
- **Xuất/Tải lên tệp cài đặt:** Thao tác để lưu file ra máy hoặc nạp file preset từ máy vào hệ thống.

► CẤU HÌNH TÍNH NĂNG CHUYỂN ĐỔI ÂM THANH VỚI BỘ PHÁT HIỆN GIỌNG NÓI:

Mô tả: Sử dụng mô hình Speechbrain và VAD (Voice Activity Detector) để phân biệt các giọng nói trong một tệp giọng, sau đó sẽ tự động cắt riêng lẻ các giọng sang các tệp riêng và chuyển đổi bằng 2 mô hình giọng nói.

Lưu ý: Nó có thể không thực sự chính xác trong việc xác định và cắt các đoạn giọng và phần chuyển đổi nếu sử dụng các tối ưu ONNX có thể gây hiệu ứng ngược khi làm giảm chất lượng đầu ra, nhiều hoặc lỗi trên các đoạn âm thanh dài, nó chỉ được khuyến cáo sử dụng trên chuyển đổi giọng nói thời gian thực.

1. Cấu Hình Phát Hiện Giọng Nói

- **Số lượng giọng:** Số lượng giọng có trong tệp âm thanh giọng nói.
- **Độ nhạy VAD:** Độ nhạy của bộ phát hiện giọng nói, giá trị cao sẽ bắt giọng chuẩn xác nhưng có thể nhận dạng cả noise.
- **Kích thước khung:** Độ dài của các đoạn âm thanh được xử lý, giá trị thấp sẽ phản ứng nhanh hơn và giá trị cao sẽ mượt hơn.

2. Thiết Lập Đầu Vào & Đầu Ra

- **Làm sạch âm thanh:** Kích hoạt bộ lọc nhiễu. Khi chọn, thanh trượt xuất hiện để bạn điều chỉnh mức độ lọc tạp âm.
- **Mức độ làm sạch âm thanh:** Điều chỉnh cường độ lọc. Giá trị cao lọc sạch hơn nhưng dễ mất chi tiết giọng; giá trị thấp giữ được ngữ âm tự nhiên nhưng có thể còn tạp âm.
- **Tải lên âm thanh:** Chọn tệp trực tiếp từ thiết bị (hỗ trợ .mp3, .wav, .flac...).
- **Đường dẫn đầu vào/đầu ra:** Dùng để chỉ định nơi lấy và lưu tệp.
- **Định dạng âm thanh:** Lựa chọn định dạng tệp xuất ra ví dụ như wav.

3. Điều Chỉnh Cao Độ F0

- **Dịch chuyển cao độ:** Điều chỉnh giọng để khớp với mô hình (Ví dụ: Giọng nữ sang mô hình nam thì chỉnh -12 và ngược lại +12). Nếu cùng giới tính, có thể chỉnh nhẹ để tối ưu chất giọng.
- **Tự động đề xuất cao độ:** Tính toán để đưa cao độ về mức phù hợp nhất với mô hình giọng nói.
- **Ngưỡng tần số ước tính:** (Chỉ hiện khi dùng Tự động đề xuất) Thanh trượt để hệ thống tính toán mức độ kéo cao độ chính xác hơn, đề xuất sử dụng 155 trở xuống đối với mô hình giọng Nam và 255 trở lên đối với mô hình giọng nữ.
- **Tự động điều chỉnh:** Tự động nắn cao độ F0 theo các nốt nhạc nó hoạt động giống như Autotune.
- **Mức độ điều chỉnh:** (Chỉ hiện khi dùng Tự động điều chỉnh) Điều chỉnh mức độ can thiệp vào nốt nhạc của F0.
- **Chế độ F0 ONNX:** Sử dụng mô hình ONNX để tăng tốc trích xuất cao độ (thay vì Pytorch), nhưng có thể lỗi với tệp quá dài.

4. Mô Hình & Chỉ Mục

- **Tệp mô hình:** Chọn mô hình RVC hoặc SVC. Với SVC, cần dùng mô hình thông số gốc và chuyển đổi lại trong tab "Thêm -> Chuyển đổi mô hình SVC".

- **Tệp chỉ mục (Index):** Giúp giọng nói giống mẫu huấn luyện hơn, giảm lỗi âm thanh. Khi có tệp này, thanh trượt "Ảnh hưởng của chỉ mục" sẽ hiện ra (Có thể nó sẽ không cần thiết nếu mô hình được huấn luyện với thời lượng dữ liệu lớn chất lượng cao).
- **Ảnh hưởng của chỉ mục:** Điều chỉnh mức độ tác động của tệp Index. Quá cao sẽ gây giọng robot/giả; quá thấp sẽ không thấy tác dụng.
- **Mã nhận dạng giọng nói:** (Chỉ hiện đối với mô hình đa giọng nói) Dùng để chọn nhân vật cụ thể trong các mô hình được huấn luyện đa giọng nói.

5. Cài Đặt Chuyên Sâu

- **Phương pháp trích xuất:** Chọn thuật toán chính để trích xuất F0. Nếu chọn **Hybrid**, bạn có thể trộn nhiều phương pháp với nhau theo tỷ lệ tùy chỉnh.
- **Mở khóa tất cả:** Hiển thị toàn bộ các phương pháp trích xuất cao độ để nghiên cứu chuyên sâu.
- **Phương pháp trích xuất HYBRID:** Cho phép bạn kết hợp của hai hoặc nhiều loại trích xuất khác nhau. Ví dụ: hybrid[pm+dio].
- **Mức trộn của các phương thức:** Mức trộn các cao độ của các phương thức trích xuất với nhau.
- **Sử dụng hiệu quả bộ nhớ:** Tối ưu hóa để tránh tràn bộ nhớ nhưng thời gian chuyển đổi sẽ lâu hơn.
- **Chế độ nhúng (Embedding):**
 - Fairseq: Dùng cho các mô hình pytorch truyền thống.
 - ONNX: Dùng cho các mô hình onnx để tối ưu tốc độ nhúng.
 - Transformers: Dùng cho các mô hình hiện đại và an toàn hơn.
 - Spin: Chất lượng cao nhất nhưng cần huấn luyện riêng.
 - Whisper: Đang trong giai đoạn thử nghiệm.
- **Mô hình nhúng:** Quyết định chất lượng giọng (thường dùng Hubert hoặc ContentVec). Nếu dùng mô hình khác Hubert và Contentvec thì bạn sẽ cần huấn luyện lại mô hình từ đầu để mô hình có thể phát âm bình thường.
- **Hòa trộn tầng:** Trộn các tầng của mô hình nhúng để tạo độ "rò" âm thanh nhẹ để cải thiện chất lượng đầu ra.
- **Tên của mô hình:** Khi sử dụng mô hình nhúng tùy chỉnh sẽ được hiện thì để bạn nhập tên của mô hình nhúng, ví dụ: Nếu là Fairseq đầu vào nên là hubert_base.pt, nếu là ONNX đầu vào nên là hubert_base.onnx và nếu là Transformers đầu vào nên là hubert_base (Nơi chứa tệp mô hình và tệp cấu hình mô hình).

- **Tầng hòa trộn:**
 - *Tầng thấp (1-6):* Chi tiết âm học.
 - *Tầng giữa (7-9):* Thông tin ngữ âm (RVC v1 thường dùng tầng 9).
 - *Tầng cao (10-12):* Nội dung ngôn ngữ, sạch hơn, tránh rò âm gốc (RVC v2 dùng tầng 12).
- **Tỉ lệ hòa trộn:** Mức độ rò âm sắc. Quá cao gây lạc điệu, quá thấp sẽ không có tác dụng.

6. Xử Lý Hậu Kỳ & Định Dạng

- **Lọc trung vị:** Kiểm soát nhiễu, làm mượt cao độ và độ tự tin của mô hình trích xuất.
- **Độ dịch chuyển khung:** Quyết định khoảng cách giữa các frame phân tích, ảnh hưởng trực tiếp đến độ mượt, độ chính xác theo thời gian, độ trễ và hiệu năng xử lý của hệ thống âm thanh.
- **Tỷ lệ trộn RMS:** Trộn năng lượng từ âm thanh gốc vào đầu ra để giọng nói sống động hơn (giá trị bằng 1 là không trộn).
- **Bảo vệ phụ âm:** Giá trị nhỏ giữ rõ tiếng thở/phụ âm; giá trị lớn ngăn tạp âm nhưng làm giọng bị bí (mất chi tiết phụ âm).
- **Tỉ lệ độ ồn (Chỉ dành cho SVC):** Điều chỉnh mức tạp âm trong mô hình. Nên giữ nguyên mặc định.
- **Dịch chuyển cao độ và âm sắc:** Thay đổi tần số phổ trước khi chuyển đổi.
- **Dịch chuyển định dạng (Formant):** Điều chỉnh độ sáng/trầm của âm sắc (Cao hơn = nữ tính, thấp hơn = nam tính).
- **Lấy mẫu lại (Resample):** Thay đổi tần số lấy mẫu của tệp âm thanh đầu ra. Hữu ích khi bạn cần đồng bộ chất lượng âm thanh với các tiêu chuẩn phần mềm hoặc phần cứng khác nhau.
- **Định dạng âm thanh:** Chọn định dạng xuất file (ví dụ: .wav).
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử kết quả ngay sau khi xử lý xong mà không cần phải mở đến tệp để nghe thử.

► CẤU HÌNH TÍNH NĂNG CHUYỂN ĐỔI VĂN BẢN:

Mô tả: Sử dụng các dịch vụ như Microsoft Edge TTS và Google TTS để chuyển đổi văn bản thành giọng nói sau đó dùng mô hình giọng nói để chuyển đổi.

Lưu ý: Các dịch vụ chuyển đổi văn bản được hỗ trợ có giới hạn và có thể gặp rắc rối với việc sử dụng các đoạn văn bản cực dài và việc sử dụng các tối ưu ONNX có thể gây hiệu ứng ngược khi làm giảm chất lượng đầu ra, nhiều hoặc lỗi trên các đoạn âm thanh dài, nó chỉ được khuyến cáo sử dụng trên chuyển đổi giọng nói thời gian thực.

1. Cấu Hình Chuyển Đổi Văn Bản

- **Nhập dữ liệu từ tệp văn bản:** Hỗ trợ tải lên các tệp văn bản như .txt, .docx, .srt (Lưu ý: Tệp srt sẽ được xử lý đặc biệt để canh theo thời gian).
- **Chuyển đổi văn bản bằng google:** Sử dụng dịch vụ của Google để chuyển đổi văn bản thành giọng nói thay vì dịch vụ của Microsoft.
- **Văn bản cần đọc:** Nội dung, văn bản cần được chuyển đổi thành giọng nói. Khi tải tệp .txt hoặc .docx nội dung sẽ được tải lên hộp văn bản này (Lưu ý: có thể không hỗ trợ văn bản quá dài).
- **Tải lên tệp văn bản:** Chọn tệp văn bản trực tiếp từ thiết bị (hỗ trợ .txt, .docx, .srt).
- **Giọng nói của các nước:** Chọn giọng nói theo giới tính và ngôn ngữ của từng quốc gia phù hợp với văn bản được cung cấp, nó được dùng để chuyển đổi văn bản thành giọng nói.
- **Dịch chuyển cao độ:** Dùng để điều chỉnh cao độ giọng nói của dịch vụ TTS cung cấp.
- **Dịch:** Dùng API Miễn phí của Google để thực hiện việc dịch văn bản cần đọc thành ngôn ngữ khác, nó có thể gặp vấn đề với đoạn văn bản quá dài.
- **Tốc độ đọc:** Tốc độ của giọng nói được tạo ra từ dịch vụ TTS.

2. Thiết Lập Đầu Vào & Đầu Ra

- **Làm sạch âm thanh:** Kích hoạt bộ lọc nhiễu. Khi chọn, thanh trượt xuất hiện để bạn điều chỉnh mức độ lọc tạp âm.
- **Mức độ làm sạch âm thanh:** Điều chỉnh cường độ lọc. Giá trị cao lọc sạch hơn nhưng dễ mất chi tiết giọng; giá trị thấp giữ được ngữ âm tự nhiên nhưng có thể còn tạp âm.
- **Đường dẫn đầu ra:** Chỉ định đường dẫn đầu ra của tệp âm thanh TTS và âm thanh được chuyển đổi thông qua mô hình.
- **Cắt âm thanh:** Tự động chia nhỏ âm thanh dài để tránh lỗi bộ nhớ và lỗi trích xuất cao độ, sau đó tự ghép lại.
- **Định dạng âm thanh:** Lựa chọn định dạng tệp xuất ra ví dụ như wav.

3. Điều Chỉnh Cao Độ F0

- **Dịch chuyển cao độ:** Điều chỉnh giọng để khớp với mô hình (Ví dụ: Giọng nữ sang mô hình nam thì chỉnh -12 và ngược lại +12). Nếu cùng giới tính, có thể chỉnh nhẹ để tối ưu chất giọng.
- **Tự động đề xuất cao độ:** Tính toán để đưa cao độ về mức phù hợp nhất với mô hình giọng nói.
- **Ngưỡng tần số ước tính:** (Chỉ hiện khi dùng Tự động đề xuất) Thanh trượt để hệ thống tính toán mức độ kéo cao độ chính xác hơn, đề xuất sử dụng 155 trở xuống đối với mô hình giọng Nam và 255 trở lên đối với mô hình giọng nữ.
- **Tự động điều chỉnh:** Tự động nắn cao độ F0 theo các nốt nhạc nó hoạt động giống như Autotune.
- **Mức độ điều chỉnh:** (Chỉ hiện khi dùng Tự động điều chỉnh) Điều chỉnh mức độ can thiệp vào nốt nhạc của F0.
- **Chế độ F0 ONNX:** Sử dụng mô hình ONNX để tăng tốc trích xuất cao độ (thay vì Pytorch), nhưng có thể lỗi với tệp quá dài.
- **Tải lên tệp F0 / Tệp F0:** Cho phép bạn tải lên tệp cao độ đã được chỉnh sửa thủ công theo ý muốn thay vì để mô hình tự tính.

4. Mô Hình & Chỉ Mục

- **Tệp mô hình:** Chọn mô hình RVC hoặc SVC. Với SVC, cần dùng mô hình thông số gốc và chuyển đổi lại trong tab "Thêm -> Chuyển đổi mô hình SVC".
- **Tệp chỉ mục (Index):** Giúp giọng nói giống mẫu huấn luyện hơn, giảm lỗi âm thanh. Khi có tệp này, thanh trượt "Ảnh hưởng của chỉ mục" sẽ hiện ra (Có thể nó sẽ không cần thiết nếu mô hình được huấn luyện với thời lượng dữ liệu lớn chất lượng cao).
- **Ảnh hưởng của chỉ mục:** Điều chỉnh mức độ tác động của tệp Index. Quá cao sẽ gây giọng robot/giả; quá thấp sẽ không thấy tác dụng.
- **Mã nhận dạng giọng nói:** (Chỉ hiện đối với mô hình đa giọng nói) Dùng để chọn nhân vật cụ thể trong các mô hình được huấn luyện đa giọng nói.

5. Cài Đặt Chuyên Sâu

- **Phương pháp trích xuất:** Chọn thuật toán chính để trích xuất F0. Nếu chọn **Hybrid**, bạn có thể trộn nhiều phương pháp với nhau theo tỷ lệ tùy chỉnh.
- **Mở khóa tất cả:** Hiện thị toàn bộ các phương pháp trích xuất cao độ để nghiên cứu chuyên sâu.
- **Phương pháp trích xuất HYBRID:** Cho phép bạn kết hợp của hai hoặc nhiều loại trích xuất khác nhau. Ví dụ: hybrid[pm+dio].

- **Mức trộn của các phương thức:** Mức trộn các cao độ của các phương thức trích xuất với nhau.
- **Sử dụng hiệu quả bộ nhớ:** Tối ưu hóa để tránh tràn bộ nhớ nhưng thời gian chuyển đổi sẽ lâu hơn.
- **Chế độ nhúng (Embedding):**
 - Fairseq: Dùng cho các mô hình pytorch truyền thống.
 - ONNX: Dùng cho các mô hình onnx để tối ưu tốc độ nhúng.
 - Transformers: Dùng cho các mô hình hiện đại và an toàn hơn.
 - Spin: Chất lượng cao nhất nhưng cần huấn luyện riêng.
 - Whisper: Đang trong giai đoạn thử nghiệm.
- **Mô hình nhúng:** Quyết định chất lượng giọng (thường dùng Hubert hoặc ContentVec). Nếu dùng mô hình khác Hubert và Contentvec thì bạn sẽ cần huấn luyện lại mô hình từ đầu để mô hình có thể phát âm bình thường.
- **Tên của mô hình:** Khi sử dụng mô hình nhúng tùy chỉnh sẽ được hiện thì để bạn nhập tên của mô hình nhúng, ví dụ: Nếu là Fairseq đầu vào nên là hubert_base.pt, nếu là ONNX đầu vào nên là hubert_base.onnx và nếu là Transformers đầu vào nên là hubert_base (Nơi chứa tệp mô hình và tệp cấu hình mô hình).
- **Hòa trộn tầng:** Trộn các tầng của mô hình nhúng để tạo độ "rò" âm thanh nhẹ để cải thiện chất lượng đầu ra.
- **Tầng hòa trộn:**
 - Tầng thấp (1-6): Chi tiết âm học.
 - Tầng giữa (7-9): Thông tin ngữ âm (RVC v1 thường dùng tầng 9).
 - Tầng cao (10-12): Nội dung ngôn ngữ, sạch hơn, tránh rò âm gốc (RVC v2 dùng tầng 12).
- **Tỉ lệ hòa trộn:** Mức độ rò âm sắc. Quá cao gây lạc điệu, quá thấp sẽ không có tác dụng.

6. Xử Lý Hậu Kỳ & Định Dạng

- **Lọc trung vị:** Kiểm soát nhiễu, làm mượt cao độ và độ tự tin của mô hình trích xuất.
- **Độ dịch chuyển khung:** Quyết định khoảng cách giữa các frame phân tích, ảnh hưởng trực tiếp đến độ mượt, độ chính xác theo thời gian, độ trễ và hiệu năng xử lý của hệ thống âm thanh.
- **Tỷ lệ trộn RMS:** Trộn năng lượng từ âm thanh gốc vào đầu ra để giọng nói sống động hơn (giá trị bằng 1 là không trộn).

- **Bảo vệ phụ âm:** Giá trị nhỏ giữ rõ tiếng thờ/phụ âm; giá trị lớn ngăn tạp âm nhưng làm giọng bị bí (mất chi tiết phụ âm).
- **Tỉ lệ độ ồn (Chỉ dành cho SVC):** Điều chỉnh mức tạp âm trong mô hình. Nên giữ nguyên mặc định.
- **Dịch chuyển cao độ và âm sắc:** Thay đổi tần số phổ trước khi chuyển đổi.
- **Dịch chuyển định dạng (Formant):** Điều chỉnh độ sáng/trầm của âm sắc (Cao hơn = nữ tính, thấp hơn = nam tính).
- **Lấy mẫu lại (Resample):** Thay đổi tần số lấy mẫu của tệp âm thanh đầu ra. Hữu ích khi bạn cần đồng bộ chất lượng âm thanh với các tiêu chuẩn phần mềm hoặc phần cứng khác nhau.
- **Định dạng âm thanh:** Chọn định dạng xuất file (ví dụ: .wav).
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử kết quả ngay sau khi xử lý xong mà không cần phải mở đến tệp để nghe thử.

7. Thiết Lập Cài Đặt Sẵn

- **Tệp cài đặt trước:** Lưu lại các thông số bạn đã chỉnh để không phải làm lại từ đầu khi tải lại trang.
- **Các cài đặt sẽ được lưu:** Chọn các mục cụ thể muốn lưu vào tệp preset.
- **Tên khi lưu tệp:** Đặt tên tùy ý cho tệp cài đặt của bạn.
- **Xuất/Tải lên tệp cài đặt:** Thao tác để lưu file ra máy hoặc nạp file preset từ máy vào hệ thống.

► CẤU HÌNH TÍNH NĂNG HIỆU ỨNG ÂM THANH:

Mô tả: Sử dụng pedalboard để cho phép bạn chỉnh sửa và nâng cao chất lượng âm thanh đầu ra bằng cách áp dụng các hiệu ứng xử lý tín hiệu khác nhau.

Lưu ý: Nếu bạn cấu hình sai hoặc là kéo mọi thứ quá cao nó có thể tạo ra tiếng rất chói tai nên hãy cẩn thận với cách bạn cấu hình.

1. Thiết Lập Đầu Vào & Đầu Ra

- **Tải lên âm thanh:** Chọn tệp trực tiếp từ thiết bị (hỗ trợ .mp3, .wav, .flac...).
- **Đường dẫn đầu vào/đầu ra:** Dùng để chỉ định nơi lấy và lưu tệp.
- **Định dạng âm thanh:** Lựa chọn định dạng tệp xuất ra ví dụ như wav.

- **Kết hợp nhạc nền:** Dùng khi cần kết hợp thêm một tệp âm thanh vào, có thể là nhạc nền hoặc giọng bè.
- **Âm lượng tệp âm thanh chính:** Mức âm lượng của tệp âm thanh chính khi được kết hợp với tệp âm thanh kết hợp.
- **Âm lượng tệp âm thanh kết hợp:** Mức âm lượng của tệp âm thanh kết hợp khi được kết hợp với tệp âm thanh chính.
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử kết quả ngay sau khi xử lý xong mà không cần phải mò đến tệp để nghe thử.

2. Thiết Lập Cài Đặt Sẵn

- **Tệp cài đặt trước:** Lưu lại các thông số bạn đã chỉnh để không phải làm lại từ đầu khi tải lại trang.
- **Các cài đặt sẽ được lưu:** Có cấu tạo khác với Tệp cài đặt trước của các tab chuyển đổi âm thanh, hãy điều chỉnh các cài đặt bên ngoài và bấm **Xuất tệp cài đặt trước**.
- **Tên khi lưu tệp:** Đặt tên tùy ý cho tệp cài đặt của bạn.
- **Xuất/Tải lên tệp cài đặt:** Thao tác để lưu file ra máy hoặc nạp file preset từ máy vào hệ thống.

3. Các Cài Đặt Hiệu Ứng

- **Hiệu ứng vọng âm:** Tạo ra không gian cho âm thanh, giúp giọng hát hoặc nhạc cụ không bị "khô". Bạn có thể điều chỉnh các thông số để tạo cảm giác âm thanh đang ở trong phòng nhỏ, khán phòng hay nhà thờ.
- **Hiệu ứng hòa âm:** Làm dày âm thanh bằng cách tạo ra các bản sao có độ lệch nhẹ về cao độ và thời gian. Hiệu ứng này tạo ra cảm giác như có nhiều nguồn âm cùng phát một lúc (ví dụ: **một dàn đồng ca thay vì một người hát đơn lẻ**).
- **Hiệu ứng độ trễ:** Lặp lại tín hiệu âm thanh sau một khoảng thời gian nhất định.
- **Hiệu ứng xoay pha:** Tạo ra sự biến đổi về màu sắc âm thanh bằng cách lọc các dải tần số thông qua việc thay đổi pha của tín hiệu. Hiệu ứng này mang lại âm thanh có tính chất "lướt" hoặc "xoáy" đặc trưng, thường dùng để tạo điểm nhấn ảo giác.
- **Hiệu ứng nén:** Kiểm soát dải động của âm thanh bằng cách giảm bớt âm lượng của những đoạn quá to và nâng những đoạn quá nhỏ. Điều này giúp âm thanh trở nên ổn định, chắc chắn và chuyên nghiệp hơn trong bản phối.
- **Tùy chọn thêm:** Cung cấp thêm một số cài đặt hiệu ứng bổ sung như:
 - *Hiệu ứng mờ dần: Tự động tạo hiệu ứng âm lượng lớn dần ở đầu hoặc nhỏ dần ở cuối đoạn âm thanh, giúp các điểm nối âm trở nên mượt mà, tránh tình trạng âm thanh bị ngắt đột ngột.*

- **Âm trầm và âm cao:** Bộ lọc đơn giản để tăng cường hoặc giảm bớt dải tần số thấp (tạo độ ấm, lực) và dải tần số cao (tạo độ sáng, sắc nét) cho giọng nói hoặc âm nhạc.
- **Giới hạn ngưỡng:** Một dạng nén âm mạnh giúp ngăn chặn tín hiệu vượt quá mức. Đây là "chốt chặn" cuối cùng để bảo vệ loa và tai người nghe không bị chói tai hay rè khi có âm lượng tăng đột biến.
- **Lấy mẫu lại:** Thay đổi tần số lấy mẫu của tệp âm thanh đầu ra. Hữu ích khi bạn cần đồng bộ chất lượng âm thanh với các tiêu chuẩn phần mềm hoặc phần cứng khác nhau.
- **Hiệu ứng nhiễu âm:** Thêm một lớp tạp âm nền (như tiếng xì trắng hoặc nhiễu hạt). Thường dùng trong thiết kế âm thanh để tạo cảm giác cổ điển hoặc giả lập môi trường thực tế.
- **Cường độ âm:** Điều chỉnh mức âm lượng tổng thể của tín hiệu trước khi xuất ra.
- **Hiệu ứng giảm bits:** Giảm độ phân giải bit của âm thanh để tạo ra chất âm thô ráp, méo kỹ thuật số kiểu các máy chơi game cũ hoặc thiết bị điện tử cổ điển.
- **Hiệu ứng méo âm:** Làm biến dạng sóng âm để tạo ra chất âm gai góc, mạnh mẽ. Thường được sử dụng để tăng thêm "cá tính" cho giọng hát hoặc nhạc cụ điện tử.
- **Dịch chuyển cao độ:** Đây là công cụ quan trọng nhất để thay đổi tông giọng mà không làm thay đổi tốc độ của đoạn âm thanh.

4. Hiệu Ứng Mờ Dần (Fade)

- **Hiệu ứng mờ dần vào (ms):** Thiết lập thời gian (mili giây) để âm lượng tăng từ 0 đến mức bình thường.
- **Hiệu ứng mờ dần ra (ms):** Thiết lập thời gian để âm lượng giảm dần về 0. Giúp kết thúc đoạn âm thanh tự nhiên hơn, không bị ngắt quãng gây khó chịu.

5. Âm Trầm Và Âm Cao (Bass/Treble)

- **Độ khuếch đại âm trầm (dB):** Tăng cường dải tần thấp để giọng nói có độ dày, ấm và nội lực hơn.
- **Tần số cắt bộ lọc thông thấp (Hz):** Xác định ngưỡng tần số mà tại đó các âm cao bị loại bỏ, giúp tập trung hoàn toàn vào dải trầm (càng thấp âm thanh càng tối).
- **Độ khuếch đại âm cao (dB):** Tăng cường độ sáng và sự rõ nét cho các âm vắn, phụ âm trong giọng hát.
- **Tần số cắt bộ lọc thông cao (Hz):** Loại bỏ các tần số thấp không cần thiết (như tiếng ù nền) để giữ lại dải âm thanh thanh thoát.

6. Giới Hạn Ngưỡng (Limiter)

- **Ngưỡng giới hạn (dB):** Thiết lập mức trần âm lượng tối đa mà tín hiệu không được phép vượt qua.
- **Thời gian thả (ms):** Tốc độ để âm lượng trở lại trạng thái bình thường sau khi bị nén lại bởi bộ giới hạn.

7. Lấy Mẫu Lại (Resample)

- **Lựa chọn tần số:** Các tùy chọn từ 8000Hz (chất lượng điện thoại) đến 96000Hz (chất lượng phòng thu cao cấp).
- **Lưu ý:** Nếu chọn 0, hệ thống sẽ giữ nguyên tốc độ lấy mẫu gốc. Cần kiểm tra định dạng file đầu ra vì một số định dạng không hỗ trợ trên 48000Hz.

8. Hiệu Ứng Nhiễu Âm & Cường Độ Âm

- **Hiệu ứng nhiễu âm:** Thêm tạp âm trắng hoặc nhiễu kỹ thuật số để tạo hiệu ứng âm thanh cổ điển hoặc méo tiếng nghệ thuật.
- **Cường độ âm:** Dùng để tăng hoặc giảm biên độ tín hiệu tổng thể của âm thanh.

9. Hiệu Ứng Giảm Bits & Hiệu Ứng Méo Âm

- **Ngưỡng cắt (Hiệu ứng méo âm):** Cắt bỏ các đỉnh sóng vượt quá ngưỡng để tạo ra âm thanh méo (distortion) mạnh mẽ.
- **Độ sâu bit (Hiệu ứng giảm bits):** Giảm số lượng bit dữ liệu (từ 24 xuống 1). Giá trị càng thấp, âm thanh càng trở nên thô ráp, "vỡ" hạt, mang đậm phong cách kỹ thuật số đời đầu.

10. Hiệu Ứng Vọng Âm (Reverb)

- **Chế độ đóng băng (Freeze Mode):** Khi được bật, âm thanh vang sẽ được lặp lại liên tục và vô tận, tạo ra một lớp nền âm thanh (pad) kéo dài.
- **Kích thước phòng (Room Size):** Điều chỉnh thể tích của không gian ảo. Giá trị càng cao (tiến về 1) tạo cảm giác âm thanh đang ở trong không gian rộng lớn như nhà hát hoặc hang động.
- **Giảm âm (Damping):** Kiểm soát mức độ hấp thụ âm thanh của các bề mặt trong phòng. Chỉ số cao sẽ làm giảm các tần số cao trong tiếng vang, khiến âm thanh nghe "tối" và ấm hơn.
- **Mức độ tín hiệu vang (Wet Level):** Điều chỉnh âm lượng của phần âm thanh đã qua xử lý hiệu ứng.

- **Mức độ tín hiệu gốc (Dry Level):** Điều chỉnh âm lượng của âm thanh nguyên bản không có hiệu ứng. Sự cân bằng giữa Wet và Dry sẽ quyết định độ "xa/gần" của nguồn âm.
- **Chiều rộng âm thanh (Width):** Điều chỉnh độ rộng của không gian stereo. Giá trị 1 mang lại không gian âm thanh bao quát và rộng mở nhất.

11. Hiệu Ứng Hòa Âm (Chorus)

- **Độ sâu hòa âm (Depth):** Quyết định mức độ biến đổi của hiệu ứng. Chỉ số cao sẽ tạo ra sự thay đổi rõ rệt về sắc thái âm thanh, làm âm thanh trở nên "dày" hơn.
- **Tần số (Rate):** Điều chỉnh tốc độ dao động của hiệu ứng hòa âm. Tần số thấp tạo sự chuyển động chậm rãi, mượt mà; tần số cao tạo hiệu ứng rung động nhanh.
- **Trộn tín hiệu (Mix):** Tỷ lệ hòa trộn giữa âm thanh gốc và âm thanh đã qua xử lý hòa âm.
- **Độ trễ trung tâm (Center Delay):** Thiết lập khoảng thời gian trễ cơ bản (tính bằng mili giây) giữa các kênh âm thanh để tạo ra sự sai lệch đặc trưng của hiệu ứng Chorus.
- **Phản hồi (Feedback):** Điều chỉnh lượng tín hiệu sau khi xử lý được quay ngược trở lại đầu vào. Các giá trị khác 0 sẽ làm hiệu ứng trở nên sắc nét và có tính chất "kim khí" hơn.

12. Hiệu Ứng Độ Trễ (Delay)

- **Thời gian trễ:** Điều chỉnh khoảng thời gian giữa âm thanh gốc và âm thanh lặp lại. Giá trị thấp tạo hiệu ứng dày âm, giá trị cao tạo tiếng vang rõ rệt.
- **Phản hồi độ trễ:** Kiểm soát số lần âm thanh lặp lại. Càng kéo về phía bên phải (giá trị 1), tiếng vang sẽ kéo dài hơn trước khi biến mất hoàn toàn.
- **Trộn tín hiệu độ trễ:** Tỷ lệ hòa quyện giữa âm thanh gốc và tiếng vang. Điều này giúp bạn kiểm soát việc tiếng vang sẽ nằm ẩn dưới nền hay nổi bật hẳn lên.

13. Hiệu Ứng Xoay Pha (Phaser)

- **Độ sâu:** Điều chỉnh mức độ ảnh hưởng của hiệu ứng lên tín hiệu gốc. Độ sâu càng lớn, âm thanh "xoáy" càng mạnh và rõ rệt.
- **Tần số:** Tốc độ di chuyển của bộ lọc pha. Tốc độ thấp tạo cảm giác lướt nhẹ, tốc độ cao tạo hiệu ứng rung động nhanh.
- **Trộn tín hiệu:** Cân bằng giữa âm thanh chưa xử lý và âm thanh đã qua bộ lọc xoay pha.
- **Tần số trung tâm:** Điểm tần số mà hiệu ứng xoay pha sẽ tập trung tác động vào nhiều nhất.

- **Phản hồi:** Đưa một phần tín hiệu đã xử lý quay ngược lại đầu vào, giúp âm thanh trở nên sắc bén, mang hơi hướng "điện tử" hoặc "kim loại" hơn.

14. Hiệu Ứng Nén (Compressor)

- **Ngưỡng nén (Threshold):** Thiết lập mức âm lượng mà tại đó bộ nén bắt đầu hoạt động. Khi âm thanh vượt quá ngưỡng này, nó sẽ bị giảm âm lượng xuống.
- **Tỉ lệ nén (Ratio):** Quyết định mức độ giảm âm lượng sau khi vượt ngưỡng. Tỉ lệ cao sẽ giữ âm lượng cực kỳ chặt chẽ, không cho phép tiếng quá to thoát ra.
- **Thời gian tấn công (Attack):** Tốc độ phản ứng của bộ nén ngay khi âm thanh vượt ngưỡng. Thời gian ngắn (mili giây nhỏ) sẽ nén ngay lập tức, thời gian dài sẽ giữ lại độ sắc nét (transients) ban đầu của âm thanh.
- **Thời gian thả (Release):** Tốc độ bộ nén ngừng hoạt động và đưa âm thanh trở lại mức bình thường sau khi tín hiệu đã xuống dưới ngưỡng nén.

► CẤU HÌNH TÍNH NĂNG HIỆU ỨNG KỶ QUẶC:

Mô tả: Cung cấp một số tập hợp tùy chỉnh hiệu ứng âm thanh theo một hướng thật kỳ quặc.

Lưu ý: Nó chỉ cung cấp một số cài đặt kỳ quặc đối với âm thanh, không thực sự khuyến khích sử dụng vì nó có thể không chính xác như cách nó được miêu tả.

1. Thiết Lập Đầu Vào & Đầu Ra

- **Tải lên âm thanh:** Chọn tệp trực tiếp từ thiết bị (hỗ trợ .mp3, .wav, .flac...).
- **Đường dẫn đầu vào/đầu ra:** Dùng để chỉ định nơi lấy và lưu tệp.
- **Định dạng âm thanh:** Lựa chọn định dạng tệp xuất ra ví dụ như wav.
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử kết quả ngay sau khi xử lý xong mà không cần phải mò đến tệp để nghe thử.

2. Các Hiệu Ứng Kỳ Quặc

- **Ngẫu Nhiên:** Chọn ngẫu nhiên các hiệu ứng hiện có.
- **Vỡ Âm:** Nhân tín hiệu gốc với một mảng giá trị ngẫu nhiên (0.5 - 0.8). Âm thanh bị sụt giảm biên độ liên tục, tạo cảm giác như bị nén quá mức hoặc nhiễu điện.
- **Kinh Dị:** Hạ tông sâu (-1.5 đến -3.5 steps) kết hợp cộng thêm nhiễu Gaussian. Giọng nói trở nên dày, trầm đục và mang sắc thái đáng sợ.

- **Người Máy:** Nâng tông (+3) và tăng tốc độ (1.2x). Tạo ra âm thanh mang tính chất cơ học, dứt khoát như giọng nói của trí tuệ nhân tạo.
- **Em bé:** Đẩy cao độ lên rất cao (+8 steps) và tăng tốc độ xử lý (1.3x). Kết quả cho ra dải tần số cao đặc trưng của giọng trẻ con.
- **Trầm:** Hạ tông (-3 steps) và giảm tốc độ xuống 0.75x. Giọng nói trở nên chậm chạp, nặng nề và mang nhiều uy lực hơn.
- **Giật Giọng:** Điều biến biên độ bằng hàm Sin theo chu kỳ. Âm thanh dao động to nhỏ liên tục, tạo hiệu ứng rung chấn đều đặn.
- **Người Già:** Kết hợp hạ tông sâu, giảm tốc và rung cao độ (Vibrato). Giọng nói mang sắc thái run rẩy và yếu ớt của người cao tuổi.
- **Vọng Âm:** Trộn tín hiệu gốc với bản sao của chính nó được làm trễ nửa giây. Tạo không gian âm thanh rộng, giống như có tiếng vang trong phòng trống.
- **Quỹ Dữ:** Nâng tông nhẹ (+2) nhưng chèn thêm sóng Sin tần số thấp vào toàn bộ dải âm. Tạo ra tiếng ù nền kỳ quái và biến dạng phi tự nhiên.
- **Méo Giọng:** Sử dụng bộ lọc Vibrato tần số cao (8Hz). Cao độ bị thay đổi liên tục một cách chóng mặt, gây cảm giác âm thanh bị biến dạng cơ học.
- **Bán Hàng Trực Tuyến:** Nâng tông (+4) và tăng tốc mạnh (1.25x). Tạo cảm giác hối hả, gấp gáp và đầy năng lượng.
- **Kéo Lê:** Phân mảnh âm thanh thành các khung nhỏ và kéo dài biên độ ở mỗi khung. Tạo hiệu ứng âm thanh bị kẹt, lê thê và vấp đĩa.
- **Khó Chịu:** Nối một đoạn sóng Sin ngắn với biên độ được nâng lên 0.04 vào cuối tệp. Tạo ra tiếng "tút" chói tai và rõ rệt hơn ngay sau khi âm thanh kết thúc.
- **Rè:** Cộng trực tiếp nhiều trắng vào mảng dữ liệu. Âm thanh giống như phát ra từ radio cũ bị mất sóng hoặc loa hỏng màng rung.
- **Lỗi Mạng:** Chia nhỏ âm thanh thành các đoạn dài 0.8 giây và xáo trộn vị trí ngẫu nhiên. Các cụm từ bị đảo lộn theo khối lớn hơn, mô phỏng tình trạng giật lag mạng nghiêm trọng.
- **Rối Loạn:** Đảo ngược dữ liệu trong từng khối 300ms xen kẽ. Tạo ra chuỗi âm thanh nghịch đảo kỳ quái nhưng vẫn giữ được nhịp điệu của câu nói.

► CẤU HÌNH TÍNH NĂNG CHUYỂN ĐỔI GIỌNG NÓI THỜI GIAN THỰC:

Mô tả: Sử dụng để chuyển đổi giọng nói thời gian thực. Cho phép chuyển đổi giọng nói từ micro và trả về loa hoặc một thiết bị đầu ra.

Lưu ý: Việc sử dụng kích thước đoạn quá nhỏ hoặc quá lớn có thể gây treo phần lõi chuyển đổi bên trong. Bạn cũng nên chú ý đến cấu hình máy, nếu cấu hình khá thấp bạn có thể thử dùng các tối ưu ONNX để giảm độ trễ thời gian thực.

1. Thiết Lập Đầu Vào & Đầu Ra

- **Độ trễ thời gian thực:** Một hộp văn bản dùng để hiển thị độ trễ của lõi chuyển đổi thời gian thực, âm lượng đầu vào và tốc độ phản hồi (Đối với chế độ Client).
- **Sử dụng thiết bị nghe lại:** Cung cấp thêm một đầu ra âm thanh khác có thể là loa hoặc là cáp đầu vào ảo.
- **Thiết bị âm thanh đầu vào/đầu ra/nghe lại:** Lựa chọn các cổng âm thanh đầu vào, đầu ra âm thanh. Có thể thiết lập các thiết bị đầu vào từ Microphone, đầu ra có thể là loa, cáp ảo hoặc là cả hai. Khuyến dùng các thiết bị có chữ WASAPI hoặc ASIO.
- **Kênh ASIO đầu vào/đầu ra/nghe lại:** Hiển thị đối với thiết bị của ASIO, dùng để thiết lập kênh của thiết bị.
- **Thiết bị đầu ra/nghe lại Stereo:** Khi được sử dụng nó sẽ phát âm thanh ra 2 kênh thay vì một kênh ASIO duy nhất.
- **Âm lượng đầu vào/đầu ra/nghe lại:** Mức độ âm lượng của các thiết bị đầu vào, đầu ra và thiết bị nghe lại. Mức âm lượng cao quá có thể gây tác động lên tai nếu sử dụng tai nghe.
- **Tốc độ lấy mẫu thiết bị đầu vào/đầu ra/nghe lại:** Tốc độ lấy mẫu của các thiết bị, hãy chắc chắn rằng thiết bị hoàn toàn hỗ trợ và tốc độ này phù hợp với thiết bị.
- **Giảm trễ thiết bị:** Đối với thiết bị đặc biệt như WASAPI có thể giảm bớt quá trình xử lý để giảm bớt độ trễ của thiết bị.
- **Làm sạch âm thanh:** Kích hoạt bộ lọc nhiễu. Khi chọn, thanh trượt xuất hiện để bạn điều chỉnh mức độ lọc tạp âm.
- **Mức độ làm sạch âm thanh:** Điều chỉnh cường độ lọc. Giá trị cao lọc sạch hơn nhưng dễ mất chi tiết giọng; giá trị thấp giữ được ngữ âm tự nhiên nhưng có thể còn tạp âm.
- **Hiệu Ứng Âm Thanh:** Khi được chọn sẽ hiển thị các cài đặt hiệu ứng âm thanh sử dụng.

2. Cấu Hình Ghi Âm

- **Đường dẫn đầu ra:** Dùng để chỉ định nơi lưu tệp ghi âm.
- **Định dạng âm thanh:** Lựa chọn định dạng tệp xuất ra ví dụ như wav.

- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử kết quả ngay sau khi xử lý xong mà không cần phải mò đến tệp để nghe thử.

3. Cấu Hình Phát Hiện Giọng Nói VAD

- **Phát hiện giọng nói VAD:** Sử dụng một bộ nhận dạng giọng nói để phát hiện và nhận dạng giọng nói, nếu không có giọng nói sẽ thực hiện việc tắt tiếng đầu ra (Lưu ý: Nó chỉ tắt tiếng đầu ra nhưng để ổn định độ trễ thì quá trình chuyển đổi hoạt động nhưng không được trả về đầu ra).
- **Độ nhạy VAD:** Độ nhạy của bộ phát hiện giọng nói, giá trị cao sẽ bắt giọng chuẩn xác nhưng có thể nhận dạng cả noise.
- **Kích thước khung:** Độ dài của các đoạn âm thanh được xử lý, giá trị thấp sẽ phản ứng nhanh hơn và giá trị cao sẽ mượt hơn.

4. Cấu Hình Thời Gian Thực

- **Kích thước đoạn:** Thiết lập độ dài của mỗi phân đoạn âm thanh mà hệ thống sẽ lấy ra để xử lý. Giá trị nhỏ giúp giảm độ trễ thời gian thực nhưng có thể chất lượng không được cao vì hệ thống có quá ít dữ liệu để hiểu, giá trị cao tăng độ trễ nhưng chất lượng có thể cao hơn. Hãy chọn giá trị phù hợp với thiết bị của bạn vì nếu độ trễ quá cao sẽ gây ngắt quãng âm thanh hoặc gây lỗi treo lỗi xử lý thời gian thực bên trong.
- **Ngưỡng im lặng:** Thiết lập mức cường độ âm thanh mà tại đó hệ thống sẽ coi là "không có tiếng" (im lặng). Nếu âm lượng đầu vào nhỏ hơn mức độ ngưỡng nó sẽ được coi là im lặng và tắt tiếng.
- **Mức đệm thêm:** Khoảng thời gian âm thanh được thêm vào các đoạn âm thanh được xử lý trong hệ thống, nó được dùng nhằm mục đích tránh tình trạng âm thanh bị mất hoặc bị ngắt đột ngột hoặc khi chuyển đổi qua lại giữa các khối dữ liệu. Nó giúp duy trì sự liên tục của tín hiệu.
- **Chồng chéo mờ dần:** Độ dài của khoảng thời gian mà hai phân đoạn âm thanh kế tiếp nhau được trộn lẫn. Khi hệ thống ghép các đoạn âm thanh lại với nhau, việc này giúp loại bỏ các tiếng "click" hoặc "pop" tại điểm nối.

5. Điều Chỉnh Cao Độ F0

- **Dịch chuyển cao độ:** Điều chỉnh giọng để khớp với mô hình (Ví dụ: Giọng nữ sang mô hình nam thì chỉnh -12 và ngược lại +12). Nếu cùng giới tính, có thể chỉnh nhẹ để tối ưu chất giọng.
- **Tự động đề xuất cao độ:** Tính toán để đưa cao độ về mức phù hợp nhất với mô hình giọng nói.

- **Ngưỡng tần số ước tính:** (Chỉ hiện khi dùng Tự động đề xuất) Thanh trượt để hệ thống tính toán mức độ kéo cao độ chính xác hơn, đề xuất sử dụng 155 trở xuống đối với mô hình giọng Nam và 255 trở lên đối với mô hình giọng nữ.
- **Tự động điều chỉnh:** Tự động nắn cao độ F0 theo các nốt nhạc nó hoạt động giống như Autotune.
- **Mức độ điều chỉnh:** (Chỉ hiện khi dùng Tự động điều chỉnh) Điều chỉnh mức độ can thiệp vào nốt nhạc của F0.
- **Chế độ F0 ONNX:** Sử dụng mô hình ONNX để tăng tốc trích xuất cao độ (thay vì Pytorch), phù hợp với máy yếu hoặc là cần độ trễ thấp.

6. Mô Hình & Chỉ Mục

- **Tệp mô hình:** Chọn mô hình RVC hoặc SVC. Với SVC, cần dùng mô hình thông số gốc và chuyển đổi lại trong tab "Thêm -> Chuyển đổi mô hình SVC".
- **Tệp chỉ mục (Index):** Giúp giọng nói giống mẫu huấn luyện hơn, giảm lỗi âm thanh. Khi có tệp này, thanh trượt "Ảnh hưởng của chỉ mục" sẽ hiện ra (Có thể nó sẽ không cần thiết nếu mô hình được huấn luyện với thời lượng dữ liệu lớn chất lượng cao).
- **Ảnh hưởng của chỉ mục:** Điều chỉnh mức độ tác động của tệp Index. Quá cao sẽ gây giọng robot/giả; quá thấp sẽ không thấy tác dụng.
- **Mã nhận dạng giọng nói:** (Chỉ hiện đối với mô hình đa giọng nói) Dùng để chọn nhân vật cụ thể trong các mô hình được huấn luyện đa giọng nói.

7. Cài Đặt Chuyên Sâu

- **Phương pháp trích xuất:** Chọn thuật toán chính để trích xuất F0. Nếu chọn **Hybrid**, bạn có thể trộn nhiều phương pháp với nhau theo tỷ lệ tùy chỉnh.
- **Mở khóa tất cả:** Hiển thị toàn bộ các phương pháp trích xuất cao độ để nghiên cứu chuyên sâu.
- **Chế độ nhúng (Embedding):**
 - Fairseq: Dùng cho các mô hình pytorch truyền thống.
 - ONNX: Dùng cho các mô hình onnx để tối ưu tốc độ nhúng.
 - Transformers: Dùng cho các mô hình hiện đại và an toàn hơn.
 - Spin: Chất lượng cao nhất nhưng cần huấn luyện riêng.
 - Whisper: Đang trong giai đoạn thử nghiệm.
- **Mô hình nhúng:** Quyết định chất lượng giọng (thường dùng Hubert hoặc ContentVec). Nếu dùng mô hình khác Hubert và Contentvec thì bạn sẽ cần huấn luyện lại mô hình từ đầu để mô hình có thể phát âm bình thường.

- **Tên của mô hình:** Khi sử dụng mô hình nhúng tùy chỉnh sẽ được hiện thị để bạn nhập tên của mô hình nhúng, ví dụ: Nếu là Fairseq đầu vào nên là hubert_base.pt, nếu là ONNX đầu vào nên là hubert_base.onnx và nếu là Transformers đầu vào nên là hubert_base (Nơi chứa tệp mô hình và tệp cấu hình mô hình).
- **Hòa trộn tầng:** Trộn các tầng của mô hình nhúng để tạo độ "rò" âm thanh nhẹ để cải thiện chất lượng đầu ra.
- **Tầng hòa trộn:**
 - *Tầng thấp (1-6):* Chi tiết âm học.
 - *Tầng giữa (7-9):* Thông tin ngữ âm (RVC v1 thường dùng tầng 9).
 - *Tầng cao (10-12):* Nội dung ngôn ngữ, sạch hơn, tránh rò âm gốc (RVC v2 dùng tầng 12).
- **Tỉ lệ hòa trộn:** Mức độ rò âm sắc. Quá cao gây lạc điệu, quá thấp sẽ không có tác dụng.

8. Xử Lý Hậu Kỳ

- **Lọc trung vị:** Kiểm soát nhiễu, làm mượt cao độ và độ tự tin của mô hình trích xuất.
- **Độ dịch chuyển khung:** Quyết định khoảng cách giữa các frame phân tích, ảnh hưởng trực tiếp đến độ mượt, độ chính xác theo thời gian, độ trễ và hiệu năng xử lý của hệ thống âm thanh.
- **Tỷ lệ trộn RMS:** Trộn năng lượng từ âm thanh gốc vào đầu ra để giọng nói sống động hơn (giá trị bằng 1 là không trộn).
- **Bảo vệ phụ âm:** Giá trị nhỏ giữ rõ tiếng thở/phụ âm; giá trị lớn ngăn tạp âm nhưng làm giọng bị bí (mất chi tiết phụ âm).
- **Tỉ lệ độ ồn (Chỉ dành cho SVC):** Điều chỉnh mức tạp âm trong mô hình. Nên giữ nguyên mặc định.

9. Cài Đặt Hiệu Ứng

- **Hiệu ứng vọng âm:** Tạo ra không gian cho âm thanh, giúp giọng hát hoặc nhạc cụ không bị "khô". Bạn có thể điều chỉnh các thông số để tạo cảm giác âm thanh đang ở trong phòng nhỏ, khán phòng hay nhà thờ.
- **Hiệu ứng hòa âm:** Làm dày âm thanh bằng cách tạo ra các bản sao có độ lệch nhẹ về cao độ và thời gian. Hiệu ứng này tạo ra cảm giác như có nhiều nguồn âm cùng phát một lúc (ví dụ: **một dàn đồng ca thay vì một người hát đơn lẻ**).
- **Hiệu ứng độ trễ:** Lặp lại tín hiệu âm thanh sau một khoảng thời gian nhất định.

- **Hiệu ứng xoay pha:** Tạo ra sự biến đổi về màu sắc âm thanh bằng cách lọc các dải tần số thông qua việc thay đổi pha của tín hiệu. Hiệu ứng này mang lại âm thanh có tính chất "lướt" hoặc "xoáy" đặc trưng, thường dùng để tạo điểm nhấn ảo giác.
- **Hiệu ứng nén:** Kiểm soát dải động của âm thanh bằng cách giảm bớt âm lượng của những đoạn quá to và nâng những đoạn quá nhỏ. Điều này giúp âm thanh trở nên ổn định, chắc chắn và chuyên nghiệp hơn trong bản phối.
- **Giới hạn ngưỡng:** Một dạng nén âm mạnh giúp ngăn chặn tín hiệu vượt quá mức. Đây là "chốt chặn" cuối cùng để bảo vệ loa và tai người nghe không bị chói tai hay rè khi có âm lượng tăng đột biến.
- **Hiệu ứng nhiễu âm:** Thêm một lớp tạp âm nền (như tiếng xì trắng hoặc nhiễu hạt). Thường dùng trong thiết kế âm thanh để tạo cảm giác cổ điển hoặc giả lập môi trường thực tế.
- **Dịch chuyển cao độ:** Đây là công cụ quan trọng nhất để thay đổi tông giọng mà không làm thay đổi tốc độ của đoạn âm thanh.
- **Cường độ âm:** Điều chỉnh mức âm lượng tổng thể của tín hiệu trước khi xuất ra.
- **Hiệu ứng giảm bits:** Giảm độ phân giải bit của âm thanh để tạo ra chất âm thô ráp, méo kỹ thuật số kiểu các máy chơi game cũ hoặc thiết bị điện tử cổ điển.
- **Hiệu ứng méo âm:** Làm biến dạng sóng âm để tạo ra chất âm gai góc, mạnh mẽ. Thường được sử dụng để tăng thêm "cá tính" cho giọng hát hoặc nhạc cụ điện tử.

10. Giới Hạn Ngưỡng (Limiter)

- **Ngưỡng giới hạn (dB):** Thiết lập mức trần âm lượng tối đa mà tín hiệu không được phép vượt qua.
- **Thời gian thả (ms):** Tốc độ để âm lượng trở lại trạng thái bình thường sau khi bị nén lại bởi bộ giới hạn.

11. Hiệu Ứng Nhiễu Âm & Cường Độ Âm

- **Hiệu ứng nhiễu âm:** Thêm tạp âm trắng hoặc nhiễu kỹ thuật số để tạo hiệu ứng âm thanh cổ điển hoặc méo tiếng nghệ thuật.
- **Cường độ âm:** Dùng để tăng hoặc giảm biên độ tín hiệu tổng thể của âm thanh.

12. Hiệu Ứng Giảm Bits & Hiệu Ứng Méo Âm

- **Ngưỡng cắt (Hiệu ứng méo âm):** Cắt bỏ các đỉnh sóng vượt quá ngưỡng để tạo ra âm thanh méo (distortion) mạnh mẽ.
- **Độ sâu bit (Hiệu ứng giảm bits):** Giảm số lượng bit dữ liệu (từ 24 xuống 1). Giá trị càng thấp, âm thanh càng trở nên thô ráp, "vỡ" hạt, mang đậm phong cách kỹ thuật số đời đầu.

13. Hiệu Ứng Vọng Âm (Reverb)

- **Chế độ đóng băng (Freeze Mode):** Khi được bật, âm thanh vang sẽ được lặp lại liên tục và vô tận, tạo ra một lớp nền âm thanh (pad) kéo dài.
- **Kích thước phòng (Room Size):** Điều chỉnh thể tích của không gian ảo. Giá trị càng cao (tiến về 1) tạo cảm giác âm thanh đang ở trong không gian rộng lớn như nhà hát hoặc hang động.
- **Giảm âm (Damping):** Kiểm soát mức độ hấp thụ âm thanh của các bề mặt trong phòng. Chỉ số cao sẽ làm giảm các tần số cao trong tiếng vang, khiến âm thanh nghe "tối" và ấm hơn.
- **Mức độ tín hiệu vang (Wet Level):** Điều chỉnh âm lượng của phần âm thanh đã qua xử lý hiệu ứng.
- **Mức độ tín hiệu gốc (Dry Level):** Điều chỉnh âm lượng của âm thanh nguyên bản không có hiệu ứng. Sự cân bằng giữa Wet và Dry sẽ quyết định độ "xa/gần" của nguồn âm.
- **Chiều rộng âm thanh (Width):** Điều chỉnh độ rộng của không gian stereo. Giá trị 1 mang lại không gian âm thanh bao quát và rộng mở nhất.

14. Hiệu Ứng Hòa Âm (Chorus)

- **Độ sâu hòa âm (Depth):** Quyết định mức độ biến đổi của hiệu ứng. Chỉ số cao sẽ tạo ra sự thay đổi rõ rệt về sắc thái âm thanh, làm âm thanh trở nên "dày" hơn.
- **Tần số (Rate):** Điều chỉnh tốc độ dao động của hiệu ứng hòa âm. Tần số thấp tạo sự chuyển động chậm rãi, mượt mà; tần số cao tạo hiệu ứng rung động nhanh.
- **Trộn tín hiệu (Mix):** Tỷ lệ hòa trộn giữa âm thanh gốc và âm thanh đã qua xử lý hòa âm.
- **Độ trễ trung tâm (Center Delay):** Thiết lập khoảng thời gian trễ cơ bản (tính bằng mili giây) giữa các kênh âm thanh để tạo ra sự sai lệch đặc trưng của hiệu ứng Chorus.
- **Phản hồi (Feedback):** Điều chỉnh lượng tín hiệu sau khi xử lý được quay ngược trở lại đầu vào. Các giá trị khác 0 sẽ làm hiệu ứng trở nên sắc nét và có tính chất "kim khí" hơn.

15. Hiệu Ứng Độ Trễ (Delay)

- **Thời gian trễ:** Điều chỉnh khoảng thời gian giữa âm thanh gốc và âm thanh lặp lại. Giá trị thấp tạo hiệu ứng dày âm, giá trị cao tạo tiếng vang rõ rệt.
- **Phản hồi độ trễ:** Kiểm soát số lần âm thanh lặp lại. Càng kéo về phía bên phải (giá trị 1), tiếng vang sẽ kéo dài hơn trước khi biến mất hoàn toàn.

- **Trộn tín hiệu độ trễ:** Tỷ lệ hòa quyện giữa âm thanh gốc và tiếng vang. Điều này giúp bạn kiểm soát việc tiếng vang sẽ nằm ẩn dưới nền hay nổi bật hẳn lên.

16. Hiệu Ứng Xoay Pha (Phaser)

- **Độ sâu:** Điều chỉnh mức độ ảnh hưởng của hiệu ứng lên tín hiệu gốc. Độ sâu càng lớn, âm thanh "xoáy" càng mạnh và rõ rệt.
- **Tần số:** Tốc độ di chuyển của bộ lọc pha. Tốc độ thấp tạo cảm giác lướt nhẹ, tốc độ cao tạo hiệu ứng rung động nhanh.
- **Trộn tín hiệu:** Cân bằng giữa âm thanh chưa xử lý và âm thanh đã qua bộ lọc xoay pha.
- **Tần số trung tâm:** Điểm tần số mà hiệu ứng xoay pha sẽ tập trung tác động vào nhiều nhất.
- **Phản hồi:** Đưa một phần tín hiệu đã xử lý quay ngược lại đầu vào, giúp âm thanh trở nên sắc bén, mang hơi hướng "điện tử" hoặc "kim loại" hơn.

17. Hiệu Ứng Nén (Compressor)

- **Ngưỡng nén (Threshold):** Thiết lập mức âm lượng mà tại đó bộ nén bắt đầu hoạt động. Khi âm thanh vượt quá ngưỡng này, nó sẽ bị giảm âm lượng xuống.
- **Tỉ lệ nén (Ratio):** Quyết định mức độ giảm âm lượng sau khi vượt ngưỡng. Tỉ lệ cao sẽ giữ âm lượng cực kỳ chặt chẽ, không cho phép tiếng quá to thoát ra.
- **Thời gian tấn công (Attack):** Tốc độ phản ứng của bộ nén ngay khi âm thanh vượt ngưỡng. Thời gian ngắn (mili giây nhỏ) sẽ nén ngay lập tức, thời gian dài sẽ giữ lại độ sắc nét (transients) ban đầu của âm thanh.
- **Thời gian thả (Release):** Tốc độ bộ nén ngừng hoạt động và đưa âm thanh trở lại mức bình thường sau khi tín hiệu đã xuống dưới ngưỡng nén.

18. Thiết Lập Cài Đặt Sẵn

- **Tệp cài đặt trước:** Lưu lại các thông số bạn đã chỉnh để không phải làm lại từ đầu khi tải lại trang.
- **Các cài đặt sẽ được lưu:** Chọn các mục cụ thể muốn lưu vào tệp preset.
- **Tên khi lưu tệp:** Đặt tên tùy ý cho tệp cài đặt của bạn.
- **Xuất/Tải lên tệp cài đặt:** Thao tác để lưu file ra máy hoặc nạp file preset từ máy vào hệ thống.

► CẤU HÌNH HUẤN LUYỆN MÔ HÌNH:

Mô tả: Dùng để tự huấn luyện mô hình giọng nói hoặc mô hình huấn luyện trước của riêng bản thân, nó còn được dùng để thử nghiệm các cấu hình huấn luyện khác nhau.

⚠ Lưu ý quan trọng trước khi huấn luyện (Training)

Việc huấn luyện mô hình yêu cầu sự chuẩn bị kỹ lưỡng về cả phần cứng lẫn kiến thức kỹ thuật. Trước khi bắt đầu, bạn cần lưu ý các điểm sau:

1. Tài nguyên và Thời gian

- Quá trình huấn luyện cực kỳ tiêu tốn tài nguyên phần cứng (CPU, GPU, RAM) và có thể kéo dài nhiều giờ hoặc nhiều ngày. Hãy đảm bảo hệ thống tản nhiệt hoạt động tốt và cấu hình máy đủ đáp ứng cường độ làm việc cao liên tục.

2. Sự tương thích của mô hình tiền huấn luyện (Pre-trained Model)

- Hầu hết các bộ **Pre-trained** hiện nay được tối ưu hóa cho cấu hình mặc định (thường là bộ nhúng **ContentVec** và **HuBERT** và bộ mã hóa tiêu chuẩn).
- **Nếu thay đổi cấu hình:** Nếu bạn chọn bộ nhúng, bộ mã hóa khác với mặc định, các tệp Pre-trained sẽ không còn tác dụng.
- **Hệ quả:** Bạn sẽ phải huấn luyện hoàn toàn từ đầu. Nếu cố tình sử dụng Pre-trained không tương thích, mô hình đầu ra sẽ bị lỗi hoặc âm thanh thu được chỉ là tạp âm.

3. Giới hạn hỗ trợ

Hiện tại, một số tùy chọn đặc biệt sau đây có thể **chưa có mô hình tiền huấn luyện**:

- Một số bộ mã hóa (Vocoder) hoặc mô hình nhúng (Embedder) mới.
- Một số tốc độ lấy mẫu (Sample Rate) không phổ biến như 24K, 44.1K.
- Không dùng tùy chọn huấn luyện cao độ (Pitch Guidance).

Lời khuyên: Nếu là người mới, hãy ưu tiên sử dụng các thiết lập mặc định để tận dụng sức mạnh của bộ Pre-trained, giúp tiết kiệm thời gian và đảm bảo chất lượng đầu ra tốt nhất.

1. Thiết Lập Cơ Bản

- **Tên của mô hình:** Đây là tên định danh cho mô hình khi huấn luyện. Lưu ý không sử dụng ký tự đặc biệt hay dấu cách. Khi lưu, hệ thống sẽ tự động đính kèm số lượng kỷ nguyên (epoch) và số bước (step) vào tên tệp.

- **Tốc độ lấy mẫu:** Tần số lấy mẫu âm thanh của mô hình. Tốc độ cao hơn sẽ cho chất lượng giọng nói tốt hơn, nhưng yêu cầu dữ liệu đầu vào phải có chất lượng tương ứng hoặc cao hơn.
- **Huấn luyện cao độ:** Giúp mô hình hiểu tín hiệu F0. Đây là tùy chọn bắt buộc để có chất lượng tốt; việc tắt tính năng này chỉ dành cho các thử nghiệm cũ.
- **Phiên bản mô hình:**
 - **v1:** Sử dụng tầng (layer) 9, điều chỉnh xuống mức 256. Dù cộng đồng đánh giá có âm sắc hay nhưng do kiến trúc cũ và dễ bị rò rỉ âm sắc nên không còn được khuyến cáo.
 - **v2:** Sử dụng tầng 12, giảm thiểu rò rỉ âm sắc và hỗ trợ các mô hình mới tốt hơn.
- **Kiến trúc mô hình:**
 - **SVC:** Đỉnh cao về độ tự nhiên, giữ được sắc thái biểu cảm và độ rung của giọng hát nhưng cực kỳ tiêu tốn tài nguyên phần cứng.
 - **RVC:** Tập trung vào mô phỏng âm sắc. Mặc dù đôi khi âm thanh hơi "phẳng", nhưng khi kết hợp với các thuật toán hiện đại như RMVPE, chất lượng đã tiệm cận rất gần với SVC.

2. Tiền Xử Lý Dữ Liệu

- **Làm sạch dữ liệu:** Bước lọc nhiễu giúp loại bỏ tạp âm khỏi bộ dữ liệu gốc, ngăn chặn việc mô hình học cả những âm thanh không mong muốn.
- **Mức độ làm sạch âm thanh:** Điều chỉnh cường độ lọc nhiễu. Giá trị cao giúp sạch tạp âm nhưng dễ làm mất chi tiết giọng; giá trị thấp giữ được chi tiết nhưng có thể còn sót nhiễu.
- **Xử lý hậu kỳ:** Sử dụng các kỹ thuật tín hiệu kỹ thuật số để làm mịn âm thanh và loại bỏ các lỗi nhỏ sau khi xử lý.
- **Chuẩn hóa âm lượng:** Điều chỉnh âm lượng tệp âm thanh về mức mục tiêu ổn định. Gồm các tùy chọn: none (không dùng), pre (trước khi xử lý) và post (sau khi xử lý).
- **Tải lên dữ liệu huấn luyện:** Khi kích hoạt, một hộp thoại sẽ xuất hiện để bạn nạp các tệp âm thanh vào thư mục huấn luyện.
- **Thả âm thanh vào đây:** Khu vực để bạn kéo thả các tệp âm thanh dùng để huấn luyện vào hệ thống.
- **Tùy chỉnh cắt âm thanh:** Phương pháp chia nhỏ tệp âm thanh dài thành các đoạn ngắn:
 - **Automatic:** Dùng Slicer2 tự động nhận diện đoạn im lặng để cắt.

- **Simple:** Cho phép tự chỉnh thời lượng cắt qua thanh trượt.
- **Skip:** Chỉ dùng khi bạn đã tự cắt dữ liệu thủ công từ trước.
- **Chiều dài đoạn (s) & Chiều dài chồng chéo:** (Dùng cho chế độ Simple) Xác định độ dài mỗi lát cắt (ví dụ 3 giây) và phần lặp lại giữa các đoạn để tránh mất thông tin tại điểm cắt.
- **Số tập im lặng:** Số lượng đoạn âm thanh không có tiếng mà bạn muốn mô hình học để nhận diện khoảng lặng chính xác hơn.
- **Thông tin phần tiền xử lý trước:** Hiển thị trạng thái quá trình chuẩn bị dữ liệu (có thể theo dõi qua terminal).

3. Trích Xuất Cao Độ & Nhúng

- **Chế độ F0 ONNX:** Sử dụng mô hình ONNX để tăng tốc trích xuất cao độ (thay vì Pytorch), nhưng có thể lỗi với tệp quá dài.
- **Phương pháp trích xuất:** Chọn thuật toán chính để trích xuất F0. Nếu chọn **Hybrid**, bạn có thể trộn nhiều phương pháp với nhau theo tỷ lệ tùy chỉnh.
- **Mở khóa tất cả:** Hiển thị toàn bộ các phương pháp trích xuất cao độ để nghiên cứu chuyên sâu.
- **Phương pháp trích xuất HYBRID:** Cho phép bạn kết hợp của hai hoặc nhiều loại trích xuất khác nhau. Ví dụ: hybrid[pm+dio].
- **Mức trộn của các phương thức:** Mức trộn các cao độ của các phương thức trích xuất với nhau.
- **Tự động điều chỉnh:** Tự động nắn cao độ F0 theo các nốt nhạc nó hoạt động giống như Autotune.
- **Mức độ điều chỉnh:** (Chỉ hiện khi dùng Tự động điều chỉnh) Điều chỉnh mức độ can thiệp vào nốt nhạc của F0.
- **Độ dịch chuyển khung:** Quyết định khoảng cách giữa các frame phân tích, ảnh hưởng trực tiếp đến độ mượt, độ chính xác theo thời gian, độ trễ và hiệu năng xử lý của hệ thống âm thanh.
- **Chế độ nhúng (Embedding):**
 - Fairseq: Dùng cho các mô hình pytorch truyền thống.
 - ONNX: Dùng cho các mô hình onnx để tối ưu tốc độ nhúng.
 - Transformers: Dùng cho các mô hình hiện đại và an toàn hơn.
 - Spin: Chất lượng cao nhất nhưng cần huấn luyện riêng.
 - Whisper: Đang trong giai đoạn thử nghiệm.

- **Mô hình nhúng:** Quyết định chất lượng giọng (thường dùng Hubert hoặc ContentVec). Nếu dùng mô hình khác Hubert và Contentvec thì bạn sẽ cần huấn luyện lại mô hình từ đầu để mô hình có thể phát âm bình thường.
- **Tên của mô hình:** Khi sử dụng mô hình nhúng tùy chỉnh sẽ được hiện thị để bạn nhập tên của mô hình nhúng, ví dụ: Nếu là Fairseq đầu vào nên là hubert_base.pt, nếu là ONNX đầu vào nên là hubert_base.onnx và nếu là Transformers đầu vào nên là hubert_base (Nơi chứa tệp mô hình và tệp cấu hình mô hình).
- **Hòa trộn tầng:** Trộn các tầng của mô hình nhúng để tạo độ "rò" âm thanh nhẹ để cải thiện chất lượng đầu ra.
- **Tầng hòa trộn:**
 - *Tầng thấp (1-6):* Chi tiết âm học.
 - *Tầng giữa (7-9):* Thông tin ngữ âm (RVC v1 thường dùng tầng 9).
 - *Tầng cao (10-12):* Nội dung ngôn ngữ, sạch hơn, tránh rò âm gốc (RVC v2 dùng tầng 12).
- **Tỉ lệ hòa trộn:** Mức độ rò âm sắc. Quá cao gây lạc điệu, quá thấp sẽ không có tác dụng.
- **Thông tin phần trích xuất dữ liệu:** Hiện thị tiến trình trích xuất cao độ, nhúng và năng lượng.

4. Cấu Hình Huấn Luyện

- **Tổng số kỷ nguyên:** Tổng số lần mô hình học lại toàn bộ dữ liệu. Đạt con số này quá trình sẽ dừng.
- **Tần suất lưu:** Số kỷ nguyên mỗi lần hệ thống thực hiện lưu bản sao mô hình (ví dụ: cứ mỗi 50 kỷ nguyên lưu 1 lần).
- **Kích thước lô (Batch size):** Số lượng mẫu xử lý trong một lần cập nhật trọng số. Số càng lớn càng tiêu tốn nhiều VRAM.
- **Số GPU được sử dụng:** Chỉ số của GPU được sử dụng, có thể xem ở phần "Thông tin GPU". Nếu thiết bị của bạn có 2 GPU có thể cấu hình thành 0-1 để tận dụng 2 GPU cho việc huấn luyện.
- **Thông tin của GPU:** Thông tin như chỉ số, tên và số dung lượng VRAM của GPU.
- **Số lõi xử lý có thể sử dụng:** Số luồng CPU tham gia vào quá trình tiền xử lý và trích xuất.
- **Lưu mô hình vào đệm:** Đưa dữ liệu lên VRAM để huấn luyện nhanh hơn, nhưng yêu cầu card đồ họa có bộ nhớ cực lớn.

- **Sử dụng hiệu quả bộ nhớ:** Giảm tải bộ nhớ bằng cách không lưu các phép tính tạm thời, nhưng sẽ làm thời gian huấn luyện lâu hơn do phải tính toán lại.
- **Trình tối ưu hóa:** Cách mô hình tự sửa lỗi:
 - **AdamW (Mặc định):** Cân bằng nhất giữa tốc độ và độ chính xác. An toàn cho 99% trường hợp.
 - **RAdam:** Giúp quá trình bắt đầu huấn luyện ổn định hơn, khó bị lỗi nếu bạn đặt Learning Rate hơi cao.
 - **AnyPrecisionAdamW:** Tối ưu cho việc huấn luyện với độ chính xác hỗn hợp (Mixed Precision), giúp tiết kiệm VRAM mà vẫn giữ chất lượng. Phù hợp khi bạn sử dụng Bfloat16.
 - **AdaBelief/AdaBeliefV2 (Được cung cấp bởi Politrees):** Kết hợp tốc độ hội tụ nhanh của Adam và khả năng tổng quát hóa cao của SGD bằng cách điều chỉnh bước nhảy dựa trên độ lệch giữa gradient thực tế và dự đoán.
- **Sử dụng giảm LR Cosine (Được cung cấp bởi Politrees):** Giảm tốc độ học theo hình sóng cosin (nhanh lúc đầu để nắm tổng quát, chậm lúc sau để tinh chỉnh), giúp phát âm rõ chữ và tránh lỗi âm thanh lạ.
- **Sử dụng tổn thất Mel đa tần:** So sánh giọng AI và giọng thật ở nhiều dải tần, giúp giọng hát dày dặn, không bị đục và giữ được độ "sáng".
- **Kiểm tra quá sức:** Trong quá trình huấn luyện, nếu mô hình học quá nhiều dữ liệu, nó có thể bị giảm chất lượng và tạo ra âm thanh kém tự nhiên. Chế độ này giúp tự động phát hiện tình trạng đó và dừng huấn luyện hoặc chọn ra mô hình tốt nhất.
- **Ngưỡng huấn luyện quá sức:** Số kỷ nguyên tối đa cho phép mức thất thoát không giảm trước khi hệ thống tự ngắt.
- **Thuật toán chỉ mục (Index):** Cách tạo tệp hỗ trợ truy xuất âm sắc:
 - **Auto:** Tự động chọn 1 trong 2 thuật toán tạo chỉ mục tốt nhất.
 - **Kmeans:** Sử dụng thuật toán để nén lại các Vector và loại bỏ dữ liệu dư thừa giúp giảm thời gian và mức độ tiêu thụ RAM.
 - **Faiss:** Chỉ dùng Faiss thuần để tạo tệp chỉ mục.
- **Thuật toán xác định:** Dùng khi muốn đảm bảo kết quả huấn luyện luôn giống hệt nhau nếu dùng chung dữ liệu.
- **Thuật toán điểm chuẩn:** Dùng để chạy thử các phương pháp tính toán khác nhau trên GPU để chọn ra cách nhanh nhất cho cấu hình máy của bạn.
- **Thông tin phần huấn luyện:** Dùng để hiển thị các thông tin của phần huấn luyện

5. Tùy Chỉnh Nâng Cao & Lưu Trữ

- **Bộ mã hóa:** Chuyển dữ liệu số thành sóng âm:

- **Default (HiFiGAN-NSF):** Tiêu chuẩn và ổn định nhất. Phù hợp cho mọi nhu cầu cơ bản và có thể sử dụng trong các phiên bản fork khác của RVC.
 - **MRF-HiFiGAN (Chưa có bộ huấn luyện trước):** Cải thiện độ trung thực so với bản mặc định.
 - **RefineGAN (Hiện chỉ tốt với các tốc độ lấy mẫu thấp):** Công nghệ mới, khử nhiễu tốt và âm thanh trong trẻo.
 - **BigVGAN:** "Trùm cuối" chất lượng nhưng cực kỳ nặng, Ngốn siêu nhiều tài nguyên tính toán.
- **Huấn luyện năng lượng (Thử nghiệm - Không khuyên dùng):** Thêm lớp học năng lượng nhưng gần như không tăng chất lượng.
 - **Tùy chỉnh bộ tham chiếu:** Nếu sử dụng Tensorboard, tùy chọn này có thể cho bạn thay đổi bộ tham chiếu âm thanh từ dữ liệu huấn luyện sang bộ tham chiếu âm thanh tùy chỉnh.
 - **Tên của bộ tham chiếu:** Tên của bộ tham chiếu được tạo trước, chọn nó để thay thế việc sử dụng dữ liệu huấn luyện làm bộ tham chiếu.
 - **Tùy chỉnh huấn luyện:** Cho phép chọn các tệp tiền huấn luyện D và G cụ thể.
 - **Chỉ lưu mới nhất:** Chỉ giữ lại bản sao trọng số mới nhất để tiết kiệm dung lượng ổ cứng.
 - **Lưu mọi mô hình:** Lưu tất cả các bản sao sau mỗi mốc tần suất lưu. Nếu tắt, chỉ lưu kết quả cuối cùng (Luôn nên sử dụng vì nếu quá trình huấn luyện đang diễn ra mà mất điện có thể gây mất luôn kết quả).
 - **Làm sạch huấn luyện:** Xóa mọi tệp dữ liệu cũ để bắt đầu huấn luyện lại từ con số không.
 - **Không dùng huấn luyện:** Tắt tất cả bộ huấn luyện trước. Huấn luyện từ đầu mà không dựa trên bất kỳ mô hình mẫu nào.
 - **Tên chủ mô hình:** Ghi danh tính người tạo vào trong tệp mô hình.
 - **Tùy chọn thư mục:** Cho phép thay đổi đường dẫn đầu vào dữ liệu huấn luyện.
 - **Thư mục chứa dữ liệu:** Đường dẫn của thư mục chứa dữ liệu dùng để huấn luyện.
 - **Tệp mô hình huấn luyện trước D/G:** Chọn mô hình huấn luyện trước.
 - **Xuất tệp:** Sau khi hoàn tất, bạn có thể chọn **Tệp mô hình** và **Tệp chỉ mục** để nén thành file ZIP, sẵn sàng chuyển sang các dự án khác.

► CẤU HÌNH TÍNH NĂNG TẠO DỮ LIỆU:

Mô tả: Tính năng này dùng để tạo nhanh một tập dữ liệu huấn luyện từ các đường dẫn youtube, sử dụng khi bạn nghe một bản cover hay một cái gì đó chứa giọng nói và bạn muốn dùng nó để huấn luyện nhưng lại không có nhiều thời gian xử lý tập dữ liệu.

Lưu ý: Nó chỉ tạo được tập dữ liệu với tính tương đối không chắc chắn chính xác và có thể có nhiễu, không khuyến cáo sử dụng nó để tạo mô hình chất lượng cao.

1. Nhóm Tính Năng Bổ Trợ:

- **Tách Nhạc:** Có quyết định tách nhạc nền hay không.

- **Khử nhiễu tách nhạc:** Áp dụng thêm bước lọc sau khi tách để âm thanh sạch hơn.

- *Mô hình MDX:* Tự động khử nhiễu.
- *Mô hình VR:* Cho phép chọn thêm một model khử nhiễu riêng biệt.
- *Lưu ý:* Không khả dụng với Demucs.

- **Tách vang (Reverb/Echo):** Tách tiếng vang từ giọng hát để hỗ trợ suy luận AI (VC) tốt hơn.

- *Lưu ý:* Với model VR, có thể chọn giữa tách tiếng Vang (Reverb) hoặc tiếng Vọng (Echo).

- **Tăng cường suy luận (VR Only):** Kéo dài thời gian kiểm thử để đạt chất lượng tách cao nhất. Chỉ dành cho model VR.

- **Xử lý dải cao (VR Only):** Tối ưu hóa các tần số cao ở đầu ra, giúp âm thanh sáng và rõ hơn. Chỉ dành cho model VR.

- **Hậu xử lý (VR Only):** Thêm một bước hậu xử lý ở đầu ra, giúp giảm tạp âm và cải thiện chất lượng. Chỉ dành cho model VR.

2. Nhóm mô hình xử lý:

Phần mô hình tách nhạc hỗ trợ 3 dòng mô hình chính: **MDX-NET**, **VR-ARCH**, và **Demucs**.

1. Mô hình Tách nhạc:

- **Tính năng:** Là bộ não chính thực hiện việc tách Giọng hát và Nhạc nền. Chất lượng của bước này ảnh hưởng trực tiếp đến tất cả các bước xử lý sau đó.
- **Cơ chế hiển thị:**
 - **Dropdown Model:** Cho phép chọn giữa các dòng **VR**, **MDX**, **Demucs**.
 - **Tùy biến theo dòng:** Khi chọn mỗi dòng, các thanh trượt tham số (Kích thước lô, Phân đoạn, Cửa sổ...) sẽ tự động thay đổi tương ứng như đã mô tả ở phần trước.

2. Mô hình Tách vang:

- **Tính năng:** Loại bỏ các hiệu ứng môi trường (Echo/Reverb) để giọng hát trở nên "khô" (Dry), cực kỳ quan trọng nếu bạn dùng giọng này để training hoặc suy luận AI (RVC).
- **Cơ chế hiển thị:** * Chỉ xuất hiện khi **Tích chọn** tính năng "Tách vang".
 - **Lưu ý đặc biệt:** Đối với dòng **VR**, hệ thống sẽ cung cấp thêm các model chuyên sâu để phân biệt giữa **Tiếng Vang (Reverb)** và **Tiếng Vọng (Echo)**.

3. Mô hình Khử nhiễu:

- **Tính năng:** Bước làm sạch cuối cùng để loại bỏ các tạp âm nhỏ phát sinh trong quá trình tách.
- **Cơ chế hiển thị:** Chỉ hiển thị khi chọn mô hình chính thuộc dòng **VR** và có bật "Khử tách nhạc".

3. Giải Thích Tham Số Kỹ Thuật

- **Số lượng dự đoán (Shifts):** (Chỉ Demucs) Số lần dự đoán lại để lấy trung bình. Càng cao càng sạch nhưng cực kỳ tốn phần cứng và chậm.
- **Kích thước lô (Batch Size):** Số lượng mẫu xử lý cùng lúc. Tăng để nhanh hơn nhưng dễ gây tràn bộ nhớ (VRAM).
- **Kích thước phân đoạn (Segment Size):** Độ dài đoạn âm thanh được cắt ra để xử lý. Giá trị lớn giúp giảm vết nối âm thanh nhưng đòi hỏi RAM cao.
- **Mức độ mạnh tay (Aggressive):** (Chỉ VR) Kiểm soát lực tách. Thấp thì giữ chi tiết nhưng dễ lẫn nhạc; Cao thì sạch nhạc nhưng dễ méo giọng.
- **Chồng chéo (Overlap):** Phần giao nhau giữa các phân đoạn để khi ghép lại âm thanh không bị ngắt quãng, tạo sự mượt mà.
- **Kích thước cửa sổ (Window Size):** (Chỉ VR) Kích thước cắt nhỏ trước khi đưa vào AI. Càng nhỏ thì AI nhìn nhận càng chi tiết, càng lớn thì tốc độ xử lý càng nhanh.
- **Độ dịch chuyển khung (Hop Length):** (Chỉ MDX) Quyết định độ phân giải thời gian của phổ âm thanh. Giá trị thấp cho độ chi tiết cao hơn.
- **Ngưỡng hậu xử lý (Post Process):** (Chỉ VR) Quyết định mức độ xử lý hậu kỳ của mô hình.

4. Thiết Lập Đầu Vào & Đầu Ra

- **Đường dẫn liên kết đến âm thanh:** Dán link video/audio từ YouTube, hệ thống sẽ tự động tải về hoặc thay bằng đường dẫn đến tệp âm thanh. Nếu cần sử dụng nhiều đường dẫn thì có thể dùng dấu phẩy để dùng nhiều
 ` <https://www.youtube.com/watch?v=...>, <https://www.youtube.com/watch?v=...> `
- **Tốc độ lấy mẫu:** Tùy chọn tốc độ lấy mẫu của tập dữ liệu.

- **Đầu ra dữ liệu:** Đường dẫn thư mục đầu ra của tập dữ liệu.
- **Làm sạch âm thanh:** Kích hoạt bộ lọc nhiễu. Khi chọn, thanh trượt xuất hiện để bạn điều chỉnh mức độ lọc tạp âm.
- **Mức độ làm sạch âm thanh:** Điều chỉnh cường độ lọc. Giá trị cao lọc sạch hơn nhưng dễ mất chi tiết giọng; giá trị thấp giữ được ngữ âm tự nhiên nhưng có thể còn tạp âm.
- **Bỏ qua giây:** Thực hiện việc bỏ qua số giây bắt đầu hoặc kết thúc. Khi sử dụng các âm thanh có tiếng mở đầu thì việc bỏ qua các tiếng mở đầu đó sẽ tốt hơn cho việc huấn luyện phía sau.
- **Bỏ qua phần đầu:** Số giây cần bỏ qua ở phần đầu dữ liệu, 0 nghĩa là không bỏ qua và nếu cần áp dụng cho nhiều tệp thì sử dụng dấu phẩy `0, 0,...`. Lưu ý phần này nên là số nguyên.
- **Bỏ qua phần cuối:** Số giây cần bỏ qua ở phần cuối dữ liệu, 0 nghĩa là không bỏ qua và nếu cần áp dụng cho nhiều tệp thì sử dụng dấu phẩy `0, 0,...`. Lưu ý phần này nên là số nguyên.
- **Thông tin tạo dữ liệu:** Hiển thị tiến trình tạo tập dữ liệu.

► CẤU HÌNH TÍNH NĂNG TẠO BỘ THAM CHIẾU:

Mô tả: Trong quá trình huấn luyện, mặc định hệ thống sẽ trích xuất chính dữ liệu huấn luyện để làm mẫu thử trên Tensorboard. Điều này dễ dẫn đến cái nhìn chủ quan vì mô hình đang tự đánh giá trên dữ liệu nó đã "học thuộc". Tính năng mới cho phép bạn tùy chỉnh bộ dữ liệu tham chiếu bằng một giọng nói hoàn toàn khác, giúp đánh giá khách quan khả năng chuyển đổi và độ tùy biến thực tế của mô hình.

Lưu ý: Hãy chắc chắn cách bạn cấu hình các tham số khi huấn luyện và tạo bộ tham chiếu giống nhau, nếu không có thể gây xảy ra sự cố.

1. Cấu Hình Trích Xuất F0

- **Huấn luyện cao độ:** Giúp mô hình hiểu tín hiệu F0. Đây là tùy chọn bắt buộc để có chất lượng tốt; việc tắt tính năng này chỉ dành cho các thử nghiệm cũ.
- **Chế độ F0 ONNX:** Sử dụng mô hình ONNX để tăng tốc trích xuất cao độ (thay vì Pytorch), nhưng có thể lỗi với tệp quá dài.
- **Phương pháp trích xuất:** Chọn thuật toán chính để trích xuất F0. Nếu chọn **Hybrid**, bạn có thể trộn nhiều phương pháp với nhau theo tỷ lệ tùy chỉnh.
- **Mở khóa tất cả:** Hiển thị toàn bộ các phương pháp trích xuất cao độ để nghiên cứu chuyên sâu.

- **Phương pháp trích xuất HYBRID:** Cho phép bạn kết hợp của hai hoặc nhiều loại trích xuất khác nhau. Ví dụ: hybrid[pm+dio].
- **Mức trộn của các phương thức:** Mức trộn các cao độ của các phương thức trích xuất với nhau.

2. Cấu Hình Nhúng

- **Phiên bản mô hình:**
 - **v1:** Sử dụng tầng (layer) 9, điều chỉnh xuống mức 256. Dù cộng đồng đánh giá có âm sắc hay nhưng do kiến trúc cũ và dễ bị rò rỉ âm sắc nên không còn được khuyến cáo.
 - **v2:** Sử dụng tầng 12, giảm thiểu rò rỉ âm sắc và hỗ trợ các mô hình mới tốt hơn.
- **Chế độ nhúng (Embedding):**
 - Fairseq: Dùng cho các mô hình pytorch truyền thống.
 - ONNX: Dùng cho các mô hình onnx để tối ưu tốc độ nhúng.
 - Transformers: Dùng cho các mô hình hiện đại và an toàn hơn.
 - Spin: Chất lượng cao nhất nhưng cần huấn luyện riêng.
 - Whisper: Đang trong giai đoạn thử nghiệm.
- **Mô hình nhúng:** Quyết định chất lượng giọng (thường dùng Hubert hoặc ContentVec). Nếu dùng mô hình khác Hubert và Contentvec thì bạn sẽ cần huấn luyện lại mô hình từ đầu để mô hình có thể phát âm bình thường.
- **Tên của mô hình:** Khi sử dụng mô hình nhúng tùy chỉnh sẽ được hiện thì để bạn nhập tên của mô hình nhúng, ví dụ: Nếu là Fairseq đầu vào nên là hubert_base.pt, nếu là ONNX đầu vào nên là hubert_base.onnx và nếu là Transformers đầu vào nên là hubert_base (Nơi chứa tệp mô hình và tệp cấu hình mô hình).
- **Hòa trộn tầng:** Trộn các tầng của mô hình nhúng để tạo độ "rò" âm thanh nhẹ để cải thiện chất lượng đầu ra.
- **Tầng hòa trộn:**
 - Tầng thấp (1-6): Chi tiết âm học.
 - Tầng giữa (7-9): Thông tin ngữ âm (RVC v1 thường dùng tầng 9).
 - Tầng cao (10-12): Nội dung ngôn ngữ, sạch hơn, tránh rò âm gốc (RVC v2 dùng tầng 12).
- **Tỉ lệ hòa trộn:** Mức độ rò âm sắc. Quá cao gây lạc điệu, quá thấp sẽ không có tác dụng.

3. Tùy Chỉnh Khác

- **Huấn luyện năng lượng (Thử nghiệm - Không khuyên dùng):** Thêm lớp học năng lượng nhưng gần như không tăng chất lượng.
- **Tự động điều chỉnh:** Tự động nắn cao độ F0 theo các nốt nhạc nó hoạt động giống như Autotune.
- **Mức độ điều chỉnh:** (Chỉ hiện khi dùng Tự động điều chỉnh) Điều chỉnh mức độ can thiệp vào nốt nhạc của F0.
- **Lọc trung vị:** Kiểm soát nhiễu, làm mượt cao độ và độ tự tin của mô hình trích xuất.
- **Dịch chuyển cao độ:** Điều chỉnh giọng để khớp với mô hình (Ví dụ: Giọng nữ sang mô hình nam thì chỉnh -12 và ngược lại +12). Nếu cùng giới tính, có thể chỉnh nhẹ để tối ưu chất giọng.
- **Tự động đề xuất cao độ:** Tính toán để đưa cao độ về mức phù hợp nhất với mô hình giọng nói.
- **Ngưỡng tần số ước tính:** (Chỉ hiện khi dùng Tự động đề xuất) Thanh trượt để hệ thống tính toán mức độ kéo cao độ chính xác hơn, đề xuất sử dụng 155 trở xuống đối với mô hình giọng Nam và 255 trở lên đối với mô hình giọng nữ.

4. Thiết Lập Đầu Vào & Đầu Ra

- **Tải lên âm thanh:** Chọn tệp trực tiếp từ thiết bị (hỗ trợ .mp3, .wav, .flac...).
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử kết quả ngay sau khi xử lý xong mà không cần phải mở đến tệp để nghe thử.
- **Đường dẫn đầu vào âm thanh:** Lựa chọn tệp âm thanh đầu vào được sử dụng để tạo bộ tham chiếu, nên sử dụng tệp âm thanh ngắn dưới 30 giây.
- **Tên của bộ tham chiếu:** Tên của bộ tham chiếu khi được lưu, khi tạo bộ tham chiếu hoàn tất sẽ thêm một số thông tin vào tên của bộ tham chiếu.
- **Thông tin tạo bộ tham chiếu:** Gần như không hữu ích, vì quá trình tạo quá nhanh.

► CẤU HÌNH TẢI XUỐNG:

Mô tả: Dùng để tải xuống, tải lên mô hình giọng nói dùng cho chuyển đổi giọng nói hoặc là mô hình huấn luyện trước dành cho huấn luyện.

Lưu ý: Hãy chắc chắn rằng mô hình được hỗ trợ, đối với mô hình giọng nói phải thuộc RVC hoặc SVC (được chuyển đổi), đối với mô hình huấn luyện trước hãy chắc chắn rằng nó là mô hình D (Discriminator) và G (Generator). Và hiện tại tính năng chỉ hỗ trợ đường dẫn liên kết từ huggingface, google drive, mediafire, pixeldrain, meganz.

1. Tải Xuống Mô Hình Giọng Nói

- **Chọn cách tải mô hình:** Chọn cách thức tải mô hình.

- **Tải từ đường dẫn liên kết:** Tải xuống mô hình từ các nguồn đường dẫn liên kết.
- **Tải từ kho mô hình csv:** Chọn mô hình từ kho mô hình được đóng góp bởi cộng đồng.
- **Tìm kiếm mô hình:** Dùng để tìm kiếm và tải xuống mô hình thông qua Internet.
- **Tải lên:** Dùng để tải lên trực tiếp mô hình thay vì phải tải xuống.

- **Tải từ đường dẫn liên kết:**

- **Đường dẫn liên kết đến mô hình:** Đường dẫn liên kết đến mô hình giọng nói, hỗ trợ huggingface, google drive, mediafire, pixeldrain và meganz.
- **Tên của mô hình:** Tên của mô hình khi tải xuống về máy (Không bắt buộc).

- **Tải từ kho mô hình csv:**

- **Kho mô hình:** Các mô hình từ cộng đồng, chọn và nhấn `Nhận mô hình` để tải xuống mô hình.

- **Tìm kiếm mô hình:**

- **Tên để tìm kiếm:** Nhập tên để tìm kiếm mô hình, sau khi nhập nhấn `Tìm kiếm` để tìm kiếm.
- **Chọn mô hình đã được tìm kiếm:** (Sau khi tìm kiếm sẽ hiển thị) Nhấn vào và chọn mô hình để tải xuống về máy.

- **Tải lên:**

- **Thả mô hình vào đây:** Nhấn vào chọn mô hình từ máy để tải lên.

2. Tải Xuống Mô Hình Huấn Luyện Trước

- **Chọn cách tải mô hình:** Chọn cách thức tải mô hình.

- **Tải từ đường dẫn liên kết:** Tải xuống mô hình từ các nguồn đường dẫn liên kết.
- **Danh sách mô hình:** Chọn mô hình từ kho lưu trữ của dự án, nhưng chỉ có phiên bản của bộ mã hóa mặc định và được huấn luyện trên hubert base.
- **Tải lên:** Dùng để tải lên trực tiếp mô hình thay vì phải tải xuống.

- **Tải từ đường dẫn liên kết:**

- **Đường dẫn liên kết đến mô hình huấn luyện trước D/G:** Nhập đường dẫn liên kết đến tệp D và G sau đó bấm `Tải Xuống`. Lưu ý đuôi mô hình phải là pth.

- **Danh sách mô hình:**

- **Chọn mô hình huấn luyện trước:** Chọn mô hình từ kho lưu trữ của dự án.
- **Tốc độ lấy mẫu của mô hình:** Chọn tốc độ lấy mẫu của mô hình.

- **Tải lên:**

- **Thả mô hình huấn luyện trước G, D vào đây:** Nhấn vào chọn mô hình huấn luyện trước từ máy để tải lên.

► CẤU HÌNH DUNG HỢP MÔ HÌNH:

Mô tả: Tính năng này cho phép bạn kết hợp hai mô hình giọng nói thành một mô hình duy nhất. Đây có thể là một cách thú vị để thử nghiệm và tạo ra giọng nói mới.

Lưu ý: Hãy chắc chắn rằng mô hình có cùng cấu trúc và cấu hình giống nhau khi dung hợp và mô hình phải là PTH (Không được dùng ONNX).

1. Cấu Hình Dung Hợp

- **Tên của mô hình:** Tên của mô hình sau khi dung hợp thành công.
- **Tải lên tệp mô hình 1 và 2:** Dùng để tải lên mô hình cần được dung hợp với nhau.
- **Tệp mô hình 1 và 2:** Tệp mô hình cần được dung hợp với nhau.
- **Tỉ lệ mô hình:** Tỉ lệ pha trộn của mô hình, giá trị 0.5 là tỉ lệ của 2 mô hình đồng đều với nhau, giá trị gần 0 sẽ giúp mô hình 1 chiếm tỉ lệ cao hơn trong việc pha trộn và ngược lại.
- **Đầu ra mô hình:** Sau khi dung hợp mô hình sẽ được hiển thị tại đây.

► CẤU HÌNH ĐỌC THÔNG TIN:

Mô tả: Tính năng này cho phép bạn truy cập và đọc các thông tin được ghi sâu bên trong mô hình.

Lưu ý: Đối với mô hình của từng dự án RVC khác nhau có thể có cách ghi thông tin khác nhau, có thể một số thông tin sẽ không thể đọc được.

1. Cấu Hình Đọc Thông Tin

- **Thả mô hình vào đây:** Dùng để tải lên mô hình cần được đọc thông tin.
- **Tệp mô hình:** Tệp mô hình cần được đọc thông tin.
- **Thông tin mô hình:** Thông tin đầu ra của mô hình sẽ được hiển thị tại đây.

► CẤU HÌNH CHUYỂN ĐỔI MÔ HÌNH ONNX:

Mô tả: Tính năng này cho phép bạn chuyển đổi mô hình RVC/SVC từ PYTORCH thuần sang ONNX để hỗ trợ tăng tốc suy luận.

Lưu ý: Quá trình xuất mô hình tùy theo phiên bản PYTORCH mà có cách xuất mô hình khác nhau và có thể xảy ra lỗi nếu phiên bản PYTORCH không phù hợp.

1. Cấu Hình Chuyển Đổi Mô Hình Sang ONNX

- **Thả mô hình vào đây:** Dùng để tải lên mô hình cần được chuyển đổi sang ONNX.
- **Tệp mô hình:** Tệp mô hình cần được chuyển đổi sang ONNX.
- **Đầu ra mô hình:** Sau khi quá trình xuất mô hình hoàn tất, mô hình sẽ được hiển thị tại đây.

► CẤU HÌNH CHUYỂN ĐỔI MÔ HÌNH SVC:

Mô tả: Tính năng này cho phép bạn chuyển đổi mô hình SVC sang định dạng mô hình mà dự án có thể đọc được. Nó giúp bạn tận dụng lại sức mạnh từ các mô hình SVC cũ.

Lưu ý: Hiện tại, hệ thống chỉ hỗ trợ tốt nhất cho mô hình SVC gốc (dùng HubertBase/ContentVec, không Diffusion). Nếu sử dụng các kiến trúc SVC khác, quá trình chuyển đổi có thể xảy ra lỗi hoặc thành công nhưng suy luận ra rác. Ngoài ra, pipeline của RVC và SVC khác nhau dẫn đến lệch thời gian, nên bên trong pipeline có thêm tính năng tự động co giãn kích thước để bù đắp sai số này.

1. Cấu Hình Chuyển Đổi Mô Hình SVC

- **Xóa các tệp gốc khi hoàn thành:** Dùng khi bạn muốn xóa các trọng số và tệp cấu hình SVC gốc sau khi xuất mô hình thành công.
- **Thả mô hình vào đây:** Dùng để tải lên mô hình cần được chuyển đổi.
- **Thả tệp vào đây:** Dùng để tải lên tệp cấu hình của mô hình cần được chuyển đổi.
- **Tệp mô hình:** Tệp mô hình SVC cần được chuyển đổi.
- **Đường dẫn tệp cấu hình SVC:** Đường dẫn đến tệp cấu hình của mô hình SVC.
- **Tên của mô hình:** Tên của mô hình sau khi chuyển đổi mô hình SVC thành công.
- **Đầu ra mô hình:** Sau khi quá trình xuất mô hình hoàn tất, mô hình sẽ được hiển thị tại đây.

► CẤU HÌNH TRÍCH XUẤT F0:

Mô tả: Tính năng này trích xuất cao độ F0 từ tệp gốc để áp dụng vào quá trình chuyển đổi giọng nói. Dù các thuật toán mới hiện nay đã tự động xử lý cao độ rất tốt, tính năng này vẫn được duy trì như một công cụ tùy chỉnh thủ công chuyên sâu cho những người dùng muốn kiểm soát chất lượng âm thanh chính xác nhất.

Lưu ý: Gần như không khuyến khích cho người dùng thông thường vì nó sẽ chẳng mang lại ích lợi gì cho việc suy luận nếu như không am hiểu về việc điều chỉnh thủ công lại đường cong F0.

1. Đường Dẫn Đầu Vào & Đầu Ra

- **Thả âm thanh vào đây:** Chọn tệp tin âm thanh trực tiếp từ thiết bị của bạn (hỗ trợ .mp3, .wav, .flac...).
- **Đường dẫn đầu vào:** Lựa chọn tệp âm thanh đầu vào được sử dụng để trích xuất F0, có thể sử dụng đường dẫn bên ngoài.
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử.
- **Đầu ra File & Image:** Sau khi quá trình trích xuất hoàn tất sẽ hiển thị tệp chứa cao độ và ảnh mô tả cao độ F0.

2. Cấu Hình Trích Xuất F0

- **Phương pháp trích xuất:** Chọn thuật toán chính để trích xuất F0. Nếu chọn **Hybrid**, bạn có thể trộn nhiều phương pháp với nhau theo tỷ lệ tùy chỉnh.
- **Mở khóa tất cả:** Hiển thị toàn bộ các phương pháp trích xuất cao độ để nghiên cứu chuyên sâu.
- **Lọc trung vị:** Kiểm soát nhiễu, làm mượt cao độ và độ tự tin của mô hình trích xuất.
- **Dịch chuyển cao độ:** Điều chỉnh giọng để khớp với mô hình (Ví dụ: Giọng nữ sang mô hình nam thì chỉnh -12 và ngược lại +12). Nếu cùng giới tính, có thể chỉnh nhẹ để tối ưu chất giọng.
- **Tự động đề xuất cao độ:** Tính toán để đưa cao độ về mức phù hợp nhất với mô hình giọng nói.
- **Ngưỡng tần số ước tính:** (Chỉ hiện khi dùng Tự động đề xuất) Thanh trượt để hệ thống tính toán mức độ kéo cao độ chính xác hơn, đề xuất sử dụng 155 trở xuống đối với mô hình giọng Nam và 255 trở lên đối với mô hình giọng nữ.
- **Tự động điều chỉnh:** Tự động nắn cao độ F0 theo các nốt nhạc nó hoạt động giống như Autotune.

- **Mức độ điều chỉnh:** (Chỉ hiện khi dùng Tự động điều chỉnh) Điều chỉnh mức độ can thiệp vào nốt nhạc của F0.
- **Chế độ F0 ONNX:** Sử dụng mô hình ONNX để tăng tốc trích xuất cao độ (thay vì Pytorch), phù hợp với máy yếu hoặc là cần độ trễ thấp.

► CẤU HÌNH TẠO TẬP SRT:

Mô tả: Tính năng này giúp tạo một tập SRT tương đối nhằm mục đích tạo tập lòng tiếng hoặc là dùng cho quá trình chuyển đổi văn bản thành âm thanh.

Lưu ý: Tính năng này sử dụng mô hình Whisper của OpenAI làm cơ sở nhận dạng tiếng nói nên có thể không thực sự chính xác hoàn toàn 100%.

1. Đường Dẫn Đầu Vào & Đầu Ra

- **Nội dung tập SRT:** Nội dung của tập SRT sẽ được hiển thị tại đây sau khi hoàn thành để hỗ trợ việc kiểm tra đầu ra của mô hình Whisper.
- **Đường dẫn đầu vào:** Lựa chọn tập âm thanh đầu vào được sử dụng để nhận dạng và tạo tập SRT, có thể sử dụng đường dẫn bên ngoài.
- **Đường dẫn tập SRT đầu ra:** Đường dẫn đầu ra của tập SRT.
- **Thả âm thanh vào đây:** Chọn tệp tin âm thanh trực tiếp từ thiết bị của bạn (hỗ trợ .mp3, .wav, .flac...).
- **Các phần hiển thị âm thanh:** Trình phát nhạc trực tiếp trên giao diện để nghe thử.
- **Tập SRT đầu ra:** Sau khi hoàn thành tập SRT sẽ được hiển thị tại đây để tải xuống.

2. Cấu Hình Mô Hình WHISPER

- **Kích thước mô hình Whisper:** Dùng để lựa chọn độ lớn của mô hình Whisper và chỉ tiếng anh hay không. Mô hình lớn có thể không cung cấp chính xác hoặc tạo ra đầu ra kì lạ, hãy chọn các mô hình cân bằng.
- **Ngôn ngữ đầu vào:** Dùng để chọn và ép mô hình biết bên trong tệp âm thanh là ngôn ngữ gì để tránh nhận dạng sai ngôn ngữ.

► TÙY CHỈNH THÊM:

Mô tả: Cung cấp khả năng điều chỉnh một số cấu hình của dự án.

Lưu ý: Cần lưu ý rằng có thể một số chủ đề không còn hoạt động và tính năng dừng tiến trình có thể hoạt động không đúng cách do có nhiều tác vụ đa luồng bên trong.

1. Cấu Hình Giao Diện

- **Đổi Chế Độ Sáng/Tối:** Dùng để đổi chế độ sáng và tối của giao diện nếu Theme có hỗ trợ.
- **Ngôn ngữ (Cần khởi động lại để áp dụng):** Dùng để đổi ngôn ngữ của toàn dự án, hiện tại chỉ hỗ trợ tiếng Anh và tiếng Việt. (Mặc dù dự án là Vietnamese-RVC nhưng phiên bản tiếng Anh được tạo ra do lâu lâu tác giả đọc không hiểu tiếng Việt của dự án 🤔).
- **Chủ đề (Cần khởi động lại để áp dụng):** Dùng để chọn và thay đổi Theme của giao diện.
- **Phông chữ (Cần khởi động lại để áp dụng):** Dùng để thay đổi phông chữ của giao diện Gradio. Sử dụng Phông Chữ từ Google để áp dụng.

2. Cấu Hình Nâng Cao

- **Sử dụng BF16 cho huấn luyện:** Kích hoạt huấn luyện trên định dạng BF16. Đây là định dạng tối ưu giúp duy trì sự ổn định của gradient trong khi vẫn tăng tốc độ tính toán và tiết kiệm bộ nhớ tương tự như FP16, nhưng có độ chính xác cao hơn đối với các mô hình lớn.
- **Sử dụng TF32 cho huấn luyện:** Định dạng toán học chuyên dụng trên các dòng GPU NVIDIA kiến trúc Ampere trở lên (RTX 30/40 series, A100...). TF32 cho phép tăng tốc độ xử lý các phép tính ma trận vượt trội mà vẫn giữ được độ chính xác tương đương với FP32.
- **Độ chính xác:** Dùng để lựa chọn độ chính xác và hiệu suất tính toán AI của toàn bộ dự án.
 - **FP16:** Sử dụng số thực 16-bit. Ưu điểm là tốc độ cực nhanh và giảm đáng kể dung lượng VRAM, nhưng đôi khi cần kỹ thuật Loss Scaling để tránh mất mát dữ liệu.
 - **FP32:** Sử dụng số thực 32-bit. Mang lại độ chính xác cao nhất và sự ổn định tuyệt đối cho mô hình, nhưng tiêu tốn nhiều tài nguyên phần cứng và thời gian huấn luyện lâu hơn.
- **Lệnh:** Một phần dùng để chạy một số lệnh debug thông qua Terminal, nó được tạo ra chỉ với mục đích thử nghiệm và debug.

3. Dừng Tiến Trình

- **Dừng Tách Nhạc/Chuyển Đổi/Tạo Dữ Liệu:** Dùng để dừng các tiến trình đơn giản không có nhiều tác vụ đa luồng bên trong.
- **Tên của mô hình:** Đối với quá trình huấn luyện mô hình, để dừng các bước trong quá trình huấn luyện cần nhập tên mô hình để truy xuất pid và ngừng tiến trình.
- **Dừng Xử Lí Dữ Liệu/Trích Xuất Dữ Liệu/Huấn Luyện:** Dùng để dừng các tiến trình bên trong có nhiều tác vụ đa luồng bên trong mã nhưng có thể không ngừng được.

Sử Dụng Tập Cấu Hình:

- **language:** Ngôn ngữ mặc định của toàn bộ dự án.
- **support_language:** Lưu số lượng ngôn ngữ được hỗ trợ (Lưu ý: Nếu thêm ngôn ngữ hãy chắc chắn rằng có tệp ngôn ngữ đó bên trong thư mục ngôn ngữ).
- **theme:** Chủ đề mặc định của giao diện sử dụng.
- **themes:** Các giao diện có trong danh sách hiển thị khi cần thay đổi trên giao diện. (Lưu ý: Hãy chắc chắn giao diện đó còn hoạt động trước khi thêm vào).
- **mdx_models:** Đây là bản định vị mô hình MDX-NET của tách nhạc, không nên thay đổi nó.
- **demucs_models:** Đây là bản định vị mô hình Demucs của tách nhạc, không nên thay đổi nó.
- **vr_models:** Đây là bản định vị mô hình VR-ARCH của tách nhạc, không nên thay đổi nó.
- **karaoke_models:** Đây là bản định vị mô hình của quá trình tách bè, không nên thay đổi nó.
- **reverb_models:** Đây là bản định vị mô hình của quá trình tách vang, không nên thay đổi nó.
- **denoise_models:** Đây là bản định vị mô hình của quá trình khử nhiễu khi tách nhạc với VR-ARCH, không nên thay đổi nó.
- **edge_tts:** Các giọng nói được hỗ trợ trong chuyển đổi văn bản thành giọng nói của dịch vụ Edge TTS.
- **google_tts_voice:** Các giọng nói được hỗ trợ trong chuyển đổi văn bản thành giọng nói của dịch vụ Google TTS.
- **whisper_languages:** Các ngôn ngữ được hỗ trợ của mô hình Whisper.
- **fp16:** Dùng để bật FP16 của dự án, chỉ nên bật khi GPU hỗ trợ FP16.
- **editing_tab:** Điều khiển hiển thị của tab Chỉnh Sửa.
- **inference_tab:** Điều khiển hiển thị của tab Suy Luận.
- **create_and_training_tab:** Điều khiển hiển thị của tab Huấn Luyện Mô Hình.
- **extra_tab:** Điều khiển hiển thị của tab Thêm.
- **separator_tab:** Điều khiển hiển thị của tab Tách Nhạc.
- **convert_tab:** Điều khiển hiển thị của tab Chuyển Đổi Âm Thanh.

- **convert_with_vad_tab:** Điều khiển hiển thị của tab Chuyển Đổi Âm Thanh Với Bộ Phát Hiện Giọng Nói.
- **tts_tab:** Điều khiển hiển thị của tab Chuyển Đổi Văn Bản.
- **effects_tab:** Điều khiển hiển thị của tab Hiệu Ứng Âm Thanh.
- **quirk_tab:** Điều khiển hiển thị của tab Hiệu Ứng Kỳ Quặc.
- **create_dataset_tab:** Điều khiển hiển thị của tab Tạo Dữ Liệu Huấn Luyện.
- **training_tab:** Điều khiển hiển thị của tab Huấn Luyện Mô Hình.
- **fushion_tab:** Điều khiển hiển thị của tab Dung Hợp Mô Hình.
- **read_tab:** Điều khiển hiển thị của tab Đọc Thông Tin.
- **onnx_tab:** Điều khiển hiển thị của tab Chuyển Đổi Mô Hình ONNX.
- **svc_tab:** Điều khiển hiển thị của tab Chuyển Đổi Mô Hình SVC.
- **downloads_tab:** Điều khiển hiển thị của tab Tải Xuống.
- **f0_extractor_tab:** Điều khiển hiển thị của tab Trích Xuất F0.
- **settings_tab:** Điều khiển hiển thị của tab Tùy Chỉnh.
- **create_srt_tab:** Điều khiển hiển thị của tab Tạo Tập SRT.
- **realtime_tab:** Điều khiển hiển thị của tab Thời Gian Thực.
- **create_reference_tab:** Điều khiển hiển thị của tab Tạo Bộ Tham Chiếu.
- **font:** Phong chữ mặc định của giao diện sử dụng.
- **app_port:** Cổng mạng của giao diện người dùng, nếu cổng đã bị chiếm nó sẽ tự chọn cổng khác.
- **tensorboard_port:** Cổng mạng của biểu đồ Tensorboard.
- **num_of_restart:** Số lần tìm lại cổng mạng khi cổng bị chiếm.
- **server_name:** Địa chỉ host của giao diện sử dụng, ví dụ như `0.0.0.0`, `127.0.0.1`.
- **app_show_error:** Kiểm soát việc hiển thị chi tiết lỗi trên giao diện sử dụng khi xảy ra lỗi trong dự án Gradio.
- **delete_exists_file:** Khi không sử dụng, nếu phát hiện tệp đã tồn tại thì sẽ tự động thêm hậu tố để tránh ghi đè. Khi sử dụng, nếu phát hiện tệp đã tồn tại thì sẽ ghi đè trực tiếp thay vì thêm hậu tố.
- **audio_effects_path:** Đường dẫn đến tệp mã xử lý phần hiệu ứng âm thanh.
- **convert_path:** Đường dẫn đến tệp mã xử lý phần chuyển đổi giọng nói (Không nên thay đổi nếu không có thể dẫn đến lỗi).
- **separate_path:** Đường dẫn đến tệp mã xử lý phần tách nhạc (Không nên thay đổi nếu không có thể dẫn đến lỗi).
- **create_dataset_path:** Đường dẫn đến mã xử lý phần tạo dữ liệu huấn luyện (Không nên thay đổi nếu không có thể dẫn đến lỗi).
- **preprocess_path:** Đường dẫn đến mã xử lý phần tiền xử lý dữ liệu huấn luyện (Không nên thay đổi nếu không có thể dẫn đến lỗi).
- **extract_path:** Đường dẫn đến mã xử lý phần trích xuất dữ liệu huấn luyện (Không nên thay đổi nếu không có thể dẫn đến lỗi).
- **create_index_path:** Đường dẫn đến mã xử lý phần tạo chỉ mục (Không nên thay đổi nếu không có thể dẫn đến lỗi).

- **train_path:** Đường dẫn đến mã xử lý phần huấn luyện (Không nên thay đổi nếu không có thể dẫn đến lỗi).
- **create_reference_path:** Đường dẫn đến mã xử lý phần tạo bộ tham chiếu huấn luyện (Không nên thay đổi nếu không có thể dẫn đến lỗi).
- **ico_path:** Đường dẫn đến tập icon của giao diện sử dụng.
- **csv_path:** Đường dẫn đến tập csv chứa kho mô hình của cộng đồng.
- **weights_path:** Đường dẫn đến thư mục chứa các trọng số huấn luyện.
- **logs_path:** Đường dẫn đến thư mục nhật ký nơi chứa tập nhật ký và phần dữ liệu huấn luyện D và G.
- **binary_path:** Đường dẫn đến thư mục chứa các tập nhị phân của dự án.
- **f0_path:** Đường dẫn đến thư mục chứa các tập trích xuất cao độ F0.
- **language_path:** Đường dẫn đến thư mục chứa các tập ngôn ngữ của dự án.
- **presets_path:** Đường dẫn đến thư mục chứa các tập cài đặt sẵn được lưu.
- **embedders_path:** Đường dẫn đến thư mục chứa các tập mô hình nhúng được tải xuống.
- **predictors_path:** Đường dẫn đến thư mục chứa các tập mô hình trích xuất cao độ được tải xuống.
- **pretrained_custom_path:** Đường dẫn đến thư mục chứa các tập mô hình huấn luyện trước tùy chỉnh.
- **pretrained_v1_path:** Đường dẫn đến thư mục chứa các tập mô hình huấn luyện trước v1.
- **pretrained_v2_path:** Đường dẫn đến thư mục chứa các tập mô hình huấn luyện trước v2.
- **speaker_diarization_path:** Đường dẫn đến thư mục chứa các tập mô hình nhận dạng giọng nói.
- **uvr5_path:** Đường dẫn đến thư mục chứa các tập mô hình tách nhạc.
- **audios_path:** Đường dẫn đến thư mục chứa các tập âm thanh.
- **reference_path:** đường dẫn thư mục chứa các tập của các bộ tham số.
- **demucs_segments_enable:** Dùng để bật/tắt chế độ phân đoạn âm thanh của mô hình Demucs, có thể làm tăng hoặc giảm tốc độ và tài nguyên tiêu thụ tùy theo từng trường hợp sử dụng.
- **demucs_cpu_mode:** Dùng để ép mô hình Demucs chạy trên CPU. Hữu ích khi thiết bị không đủ tài nguyên hoặc bị quá tải khi chạy Demucs trên GPU.
- **limit_f0:** Khi sử dụng tính năng đề xuất cao độ, giá trị tính toán có thể vượt ngoài mức thông thường; tùy chọn này dùng để giới hạn lại phạm vi giá trị đó.
- **debug_mode:** Bật chế độ gỡ lỗi của toàn bộ dự án, nó sẽ in ra tất cả cảnh báo ẩn các đoạn gỡ lỗi,...
- **pretrain_strict:** Khi không được sử dụng sẽ ép mô hình huấn luyện trước có thể hoạt động khi cấu trúc mô hình khác với cấu trúc huấn luyện hiện tại. Khi sử dụng sẽ yêu cầu cấu trúc phải khớp hoàn toàn.
- **cpu_mode:** Dùng để ép dự án chạy trên cpu thay vì bất kì thiết bị khác.

- **brain:** Dùng để bật BF16 để tăng tốc huấn luyện FP16. Chỉ nên bật nếu GPU của bạn hỗ trợ sử dụng BF16.
- **tf32:** Dùng để bật TF32 để tăng tốc huấn luyện FP32. Chỉ nên bật nếu GPU của bạn thuộc NVIDIA và hỗ trợ TF32.
- **discord_presence:** Khi sử dụng, nó sẽ hiện thị rpc việc bạn đang sử dụng dự án này trên discord.
- **f0_min:** Mức giới hạn thấp nhất được cho phép trên các mô hình trích xuất cao độ (Lưu ý: Đây là mức thiết kế gốc không tự ý chỉnh sửa có thể gây ảnh hưởng đến chất lượng đầu ra).
- **f0_max:** Mức giới hạn cao nhất được cho phép trên các mô hình trích xuất cao độ (Lưu ý: Đây là mức thiết kế gốc không tự ý chỉnh sửa có thể gây ảnh hưởng đến chất lượng đầu ra).
- **turn_on_resize_f0_for_all:** Hãy bật nó nếu muốn đảm bảo tất cả đầu ra F0 có thể có dân để phù hợp với độ dài âm thanh, có thể tăng thời gian xử lí.
- **allow_fp16_all_backend:** Cho phép về hình thức các Backend không hỗ trợ FP16 có thể chuyển đổi sang chế độ FP16, nhưng có thể nó sẽ tự chuyển về FP32 hoặc lỗi do không hỗ trợ.
- **asio_enabled:** Dùng để kích hoạt các thiết bị âm thanh ASIO để dự án có thể nhận diện.
- **ffmpeg_download:** Tự động tải FFmpeg nếu thiếu.
- **compile_all:** Bật/Tắt chế độ biên dịch mô hình Pytorch. Lưu ý cần Triton để hoạt động.
- **compile_mode:** Chế độ biên dịch mô hình.
- **compile_cache_dir:** Đường dẫn thư mục cache của quá trình biên dịch mô hình.
- **check_data:** Dùng để kiểm tra dữ liệu của phần huấn luyện trước khi chạy bất kỳ tiến trình nào liên quan đến huấn luyện.