

HỆ THỐNG CHUYỂN ĐỔI BÀI NHẠC TIẾNG VIỆT SANG PHIÊN BẢN TIẾNG ANH

Nguyễn Thiên Bảo¹

¹ Trường ĐH Công Nghệ Thông Tin,
ĐH Quốc Gia TPHCM

Mục tiêu

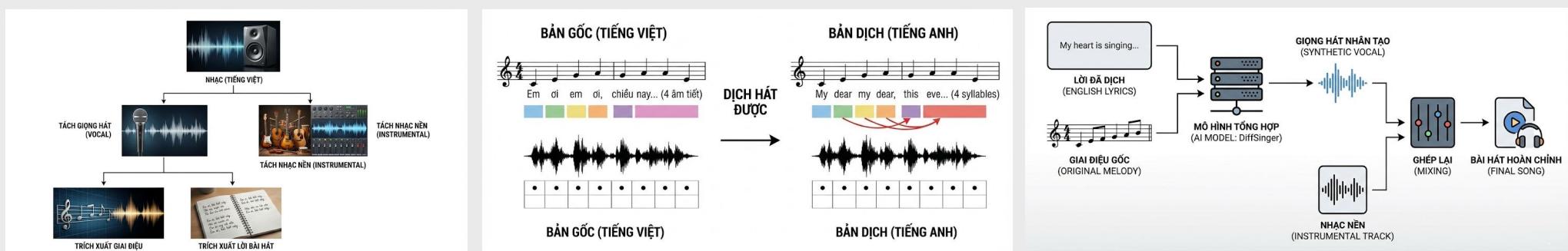
Chúng tôi đề xuất một hệ thống chuyển đổi bài nhạc tiếng việt sang phiên bản tiếng anh:

- Từ bản nhạc tiếng việt sẽ tiến hành tách nhạc, dịch lời ca cho đến tổng hợp giọng hát, tạo ra bản nhạc phiên bản tiếng anh.
- Là công cụ hỗ trợ đắc lực cho người học ngoại ngữ và những nhà sáng tạo nội dung

Lý do

- Âm nhạc là cầu nối văn hóa và là công cụ đắc lực trong việc học ngôn ngữ. Hiện nay, nhiều người học tiếng Anh có nhu cầu nghe lại các bài hát tiếng Việt yêu thích dưới phiên bản tiếng Anh để vừa giải trí, vừa học từ vựng và ngữ điệu.
- Ngôn ngữ tiếng việt chưa được chú trọng nghiên cứu rộng rãi, đặc biệt trong lĩnh vực dịch thuật âm nhạc.

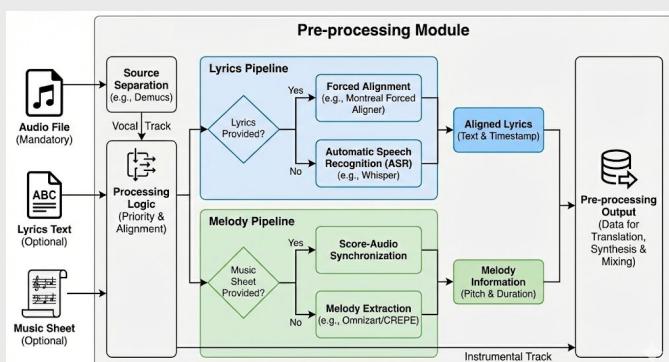
TỔNG QUAN



MÔ TẢ

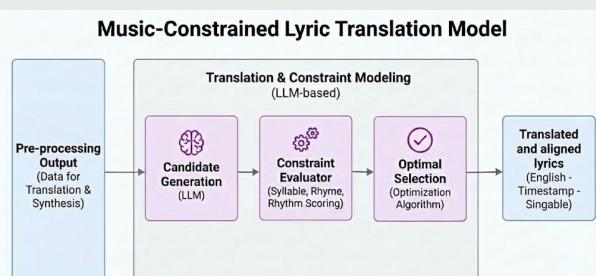
1. Hệ thống tiền xử lý đầu vào

- Sử dụng mô hình tách nguồn âm thanh (ví dụ: Demucs) để tách riêng phần Vocal (lời hát) và Instrumental (nhạc nền) từ file audio gốc.
- Áp dụng mô hình nhận dạng giọng nói (ví dụ: Whisper) để chuyển đổi Vocal thành văn bản (Lyrics) kèm thông tin thời gian.
- Sử dụng thuật toán trích xuất giai điệu (ví dụ: Omnidart hoặc CREPE) để lấy thông tin cao độ (pitch) và trường độ (duration).
- Phương pháp lai (Hybrid): Cho phép người dùng tải lên Music Sheet hoặc Lyrics chuẩn để tăng độ chính xác của dữ liệu đầu vào.



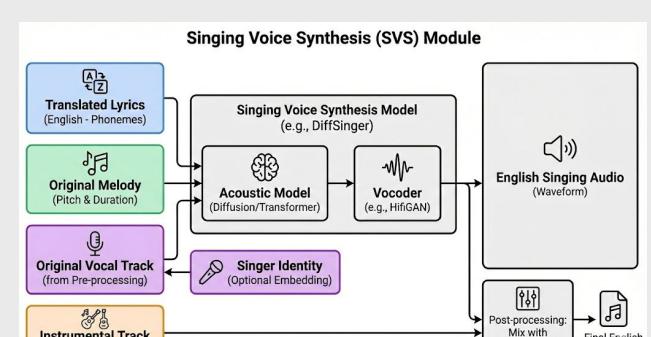
2. Dịch Lyrics có ràng buộc âm nhạc

- Dùng Mô hình Ngôn ngữ Lớn (LLM) để sinh ra nhiều ứng viên dịch thuật cho từng câu hát.
- Áp dụng thuật toán tối ưu hóa để lựa chọn phương án dịch tốt nhất dựa trên các hàm mục tiêu (Reward Functions) về:
 - Syllable count (số âm tiết)
 - Rhyme & Rhythm (vần & trọng âm)
 - Meaning (độ tương đồng ngữ nghĩa)



3. Mô hình tổng hợp giọng hát

- Xây dựng mô hình Acoustic Model (ví dụ: dựa trên kiến trúc Diffusion hoặc Transformer) để dự đoán phổ âm thanh (Mel-spectrogram) từ lời bài hát tiếng Anh và thông tin giai điệu gốc.
- Sử dụng Vocoder (ví dụ: HifiGAN) để chuyển đổi Mel-spectrogram thành dạng sóng âm thanh (Waveform).



4. Hậu xử lý & Tích hợp

- Ghép giọng hát tiếng Anh vừa tạo với phần nhạc nền (Instrumental) đã tách ở tiền xử lý để tạo ra sản phẩm cuối cùng. Tích hợp các bước xử lý thành một hệ thống hoàn chỉnh.