



기획서

프로젝트 소개 및 사전질문 B105 수정.pdf

개요

내 목소리로 여러 악기들을 연주하고, 이를 활용해서 루프스테이션 처럼 잼세션에서 활용하거나, 합성해서 완성된 음악으로 만들 수 있게 해주는 서비스

타겟층

1. 작곡을 해보고 싶지만 금액, 난이도 측면에서 진입장벽이 높아 시도해보지 못하는 작곡 입문자
2. 기타, 드럼 등 한가지 악기는 다룰 줄 알지만 여러 악기를 다루지 못해 혼자서는 곡 작업과 잼 세션을 하지 못하는 취미생
3. 곡 작업을 위해 간단한 샘플을 만들고 싶지만 연주자를 섭외할 비용이 부담스러운 작곡가

장점

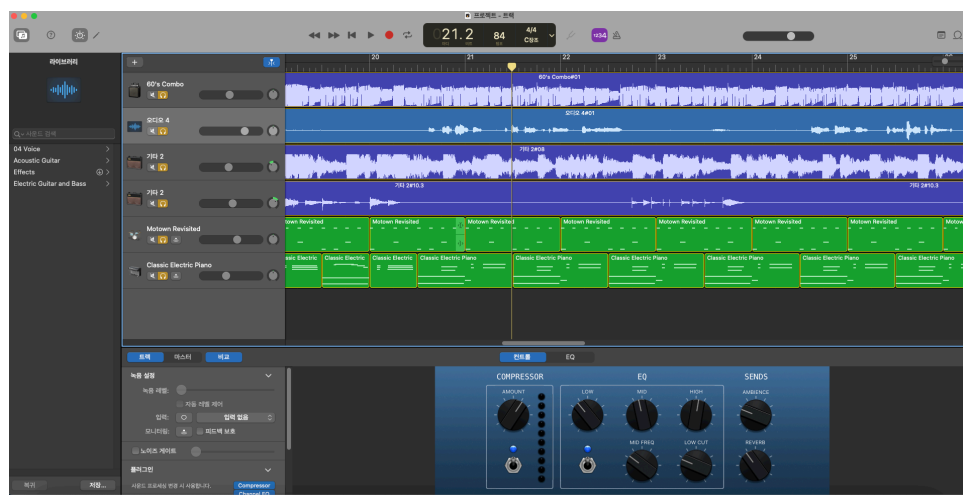
1. 초보자의 진입장벽을 낮추어 좀 더 작곡이라는 활동에 쉽게 다가갈 수 있음
2. 여러 악기를 다루지 못해도 간단하게 작곡 가능
3. 장비가 필요하지 않기 때문에 간단한 곡 스케치 혹은 아이디어를 메모할 때 편리하게 사용할 수 있음

주요 기능

음성 기반 악기 연주 및 곡 작업(MVP)

- 음성을 인공지능이 학습된 악기 소리로 변환
 - 사용자가 변환을 원하는 악기를 선택하고 음성(허밍)을 녹음한 뒤, 악기소리로 변환해주는 기능
 - 드럼
 - 바이올린
 - 색소폰
 - 트럼펫
 - MIDI는 디지털 신호이기 때문에 가상악기를 활용하면 소리의 변경이 쉽지만, 아날로그 신호인 음성을 다른 악기소리로 변환하기는 쉽지 않음. 또한 특정 악기만의 표현력을 재현하는 것에 한계가 있음
 - 때문에 인공지능의 사전학습을 기반으로 음성을 악기소리로 변환하는 기능을 개발할 필요가 있음
- 생성한 사운드 샘플을 활용한 작곡
 - 변환된 사운드 샘플들은 하나의 트랙으로 구성되어 사용자가 원하는 위치에 위치시킬 수 있음
 - 여러 트을 쌓아나가면서 하나의 음악을 만들 수 있음
 - 루프 스테이션 기능 제공하여 반복되는 패턴을 가진 샘플(드럼, 베이스라인 등)은 추가적으로 배치하지 않아도 간단하게 음악을 만들 수 있게 UI를 제공

▼ Apple의 GarageBand



- 완성곡 관련 부가 기능
 - 완성한 곡을 다운로드 또는 공유할 수 있는 기능
 - 어울리는 앨범 커버를 생성해주기

확장시 추가 가능한 기능

1. BPM 보정 & 음 보정

- a. 사용자의 허밍이 정확한 템포를 유지한다고 보장할 수 없기 때문에 BPM을 보정해주는 Quantization 기능이 필요할 수 있음
- b. 동일한 이유로 항상 정확한 음정을 짚는다고 보장할 수 없기 때문에 AutoTune 기능을 지원해야 할 수 있음

2. 이펙터

- a. 리버브, 필터링, EQ 등을 조정 가능하게 바꿔주기

3. 악기와 음성 추가

- a. 여러 악기 추가
- b. 다른 음성 추가(ex: 여자 목소리, 남자 목소리, 프레디 머큐리 목소리 ... etc)

4. 완성 곡 공유

- a. Youtube api 등을 활용해서 자기 채널에 업로드 하는 기능 등

5. 곡 이어 작곡하기

- a. 다른사람들이 만들어둔 미완성 곡을 가져다가 이어서 만들 수 있는 기능

5. 완성한 곡 악보 제공

- a. 완성한 곡에 대한 악보를 만들어 작곡한 곡을 진짜 악기로 연주할 수 있도록 제공해주기

4. 커뮤니티

- a. 사용자가 작곡한 음악을 업로드하고 피드백 받을 수 있는 커뮤니티

사용 모델

DDSP(Differentiable Digital Signal Processing)

DDSP는 Differentiable Digital Signal Processing (미분 가능한 디지털 신호 처리) 의 약자로, 기존의 디지털 신호 처리(DSP) 기술과 딥러닝을 결합하여 더 자연스럽게 효율적인 오디오 합성을 가능하게 하는 기술.

- 기존 오디오 합성 방식의 한계를 보완하기 위해 Google Research (Magenta Team) 에서 개발
- 클래식한 DSP 신디사이저 기법을 자동 미분(Auto-Differentiation)이 가능한 형태로 변환하여 딥러닝과 결합
- 소규모 데이터로도 고품질 오디오 합성 가능

생성 방식

신경망이 직접 "어떤 소리를 합성해야 하는지"를 학습합니다.

1. 신경망이 오디오 특성을 분석하고 학습

- 입력 데이터(예: 허밍, 악기 소리)를 받아 피치(f_0), 볼륨(loudness), 타임 스트레칭(Envelope) 등을 자동으로 추출.
- 딥러닝 모델이 이 데이터를 학습하여 특정 악기의 특성을 모델링.

2. 신경망이 DSP 파라미터를 예측

- 기존 신디사이저는 사람이 "이 정도 피치와 볼륨이면 이런 소리가 나와야 한다"고 직접 설정해야 함.
- DDSP는 신경망이 데이터에서 직접 "이런 피치와 볼륨을 가진 입력이라면, 이런 악기 소리가 나와야 한다"를 학습.
- 즉, 오실레이터, 필터, 리버브 등의 DSP 요소를 딥러닝 모델이 직접 컨트롤.

3. 신경망이 오디오를 생성

- 학습한 내용을 바탕으로 실제 소리를 생성하는 DSP 모듈의 매개변수를 신경망이 조정.
- 기존 딥러닝 모델(WaveNet 등)은 샘플 단위로 오디오를 예측하지만, DDSP는 DSP 요소를 사용하여 오디오를 생성하므로 더욱 효율적.

데이터셋

NSynth

제공: Google Magenta(DDSP 연구팀)

- 1,000개 이상의 악기로 연주된 30만 개 이상의 단일 노트 오디오 샘플 제공.
- 악기별(드럼, 기타, 베이스, 피아노 등)로 레이블이 지정된 데이터셋.

고려사항

1. 데스크톱 / 태블릿 / 모바일 UI 최적화

- a. 화면 하나에 많은 기능이 들어갈 가능성이 높음. 따라서 UI를 사용자에게 직관적이고, 초보자가 쉽게 활용할 수 있도록 구성해야 함

레퍼런스

<https://magenta.tensorflow.org/>

<https://openreview.net/forum?id=B1x1ma4tDr>