ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Thesis!!!!!!!!!

Hehe

Research Title

A Comparative Evaluation Of Tonal Replication Techniques For Music Composition And Sound Design Through Tonal Synthesis

ACSML

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Chapter 1: Introduction

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Context

Context

- Music has constantly evolved through innovation, whether in sound design or composition
- One good example is the use of synthesizers starting in the 1960s
  - The use of synthesizers led to genres such as techno & house
- Even more recently, we've seen more advancements in music
  - Algorithmic composition, sample-based production, AI and ML-assisted music production
- This study aims to explore an intersection between AI and advancements in synthesizer technology: the replication of instrumental tones

ACSML

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Research Objectives

Research Objectives

The study aims to utilize various techniques in audio synthesis and replication, namely Tone Transfer, GANs, and DSPGAN, in order to replicate instrumental tones, comparing them in terms of accuracy and efficiency.

Sub-objective #1:

Determine how the quality and choice of input audio encoding and representation affect sound replication efficiency and accuracy.

ACSML

Research Objectives

The study aims to utilize various techniques in audio synthesis and replication, namely Tone Transfer, GANs, and DSPGANs, in order to replicate instrumental tones, comparing them in terms of accuracy and efficiency.

Sub-objective #2:

Determine the stability of the outlined methods in handling various factors in the audio input, such as noise, frequency masks, low audio quality, and fragmented audio inputs.

ACSML

Research Objectives

The study aims to utilize various techniques in audio synthesis and replication, namely Tone Transfer, GANs, and DSPGANs, in order to replicate instrumental tones, comparing them in terms of accuracy and efficiency.

Sub-objective #3:

Determine the perceptual quality of synthesized audio as compared to the original audio input.

ACSML

Research Objectives

The study aims to utilize various techniques in audio synthesis and replication, namely Tone Transfer, GANs, and DSPGANs, in order to replicate instrumental tones, comparing them in terms of accuracy and efficiency.

Sub-objective #4:

Determine the effect of contrasting feature selection and representation on the accuracy and efficiency of synthesized audio.

ACSML

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Research Questions

Research Questions

How do methods in audio synthesis and replication (Tone Transfer, GANs and DSPGAN) compare in their accuracy and efficiency in replicating instrumental tones?

Sub-question #1:

How does the quality and the choice of encoding and representation for the input audio impact the efficiency and accuracy of sound replication?

ACSML

Research Questions

How do methods in audio synthesis and replication (Tone Transfer, GANs and DSPGAN) compare in their accuracy and efficiency in replicating instrumental tones?

Sub-question #2:

How stable are different methods in audio synthesis and replication in handling noise and frequency masks in the audio input, low audio quality, and fragmented audio inputs?

ACSML

Research Questions

How do methods in audio synthesis and replication (Tone Transfer, GANs and DSPGAN) compare in their accuracy and efficiency in replicating instrumental tones?

Sub-question #3:

How does the synthesized output generated by the model compare to the original audio input in terms of perceptual quality?

ACSML

Research Questions

How do methods in audio synthesis and replication (Tone Transfer, GANs and DSPGAN) compare in their accuracy and efficiency in replicating instrumental tones?

Sub-question #4:

How do the contrasting feature selection and representation employed in different methods contribute to their accuracy and efficiency in replicating instrumental tones?

ACSML

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Scope and Limitations

Scope (Technology)

- We will namely be focused on specific implementations of:
  - Tone Transfer Architecture (Caspe et. al, 2023)
  - Generative Adversarial Network (Engel et. al, 2019)
  - DSPGAN (Song et. al, 2023)
- An architecture will be formulated from the comparison of these implementations. [TENTATIVE]
- The dataset used will be the NSynth dataset
  - 305,979 musical notes from 1,009 instruments

ACSML

Scope (Audio)

- Audio input is defined as any melodic recording of an instrument
  - Recordings will come from the following instruments:
    - Violin, Guitar, Flute, Trombone, Organ, Marimba
  - Quality of audio varies.
  - Note qualities also vary to determine which features of each architecture affect the audio output

ACSML

Limitations (Metrics)

- The metrics we will use are as follows:
  - Subjective:
    - Mean Opinion Score [Human Evaluation]
    - ABX Testing
  - Objective:
    - Frechet Inception Distance (FID)
    - Nearest Neighbor Comparison
    - Might Add More. [Tentative]
      - https://pypi.org/project/Audio-Similarity/

ACSML

Limitations (Metrics)

- There is no single metric that could effectively determine the accuracy of the models:
  - For human evaluation such as Mean Opinion Score and ABX, testing can be subjective, introduce personal biases and produce unexpected results
  - Objective evaluations like Frechet Inception / Audio Distance and Nearest Neighbor Comparison lack interpretability and perceptual relevance.

ACSML

Limitations (Other stuff)

- The study focuses on comparing specific architecture implementations of the concerned models for tonal replication
- Hence, the creation of new models and architecture is outside of the scope of this paper

ACSML

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Significance of the Study

Significance

- The study will...
  - offer comparative insights into the performance of the various tonal replication techniques
  - provide guidelines for choosing and optimizing audio encoding and representation strategies

ACSML

Significance

- The study will also...
  - assess the stability and robustness of these methods in various real-world scenarios
    - Example: Restoring audio
  - contribute to the understanding of feature selection and representation's role in audio synthesis effectiveness.
    - provide general guidelines for future architecture building

ACSML

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Chapter 2:
Review of Related Literature

Previous work

- WaveNet-style autoencoders (Engel et. al, 2017)
  - Learns temporal codes for consistency in long-term structure
  - Also introduced NSynth dataset (from sample libraries)
- NSynth dataset (Engel et. al, 2017)
  - 306,043 notes from 1006 instruments
  - Annotated with source, instrument family, qualities
- SING (Symbol-to-Instrument Neural Generator) (Défossez et. al, 2018)
  - Lightweight neural audio synthesizer
  - Improved perceptual quality compared to Wavenet-style

ACSML

Generative Adversarial Networks (GANs)

- Typically used for generative images
- Applications
- WaveGAN (Donahue et. al, 2019)
  - “First attempt” at using GANs for audio synthesis
  - Could synthesize 1 second slices of audio (speech, instrumental) with global coherence
- GANSynth (Engel et. al, 2020)
  - GAN capable of synthesizing locally coherent sound

ACSML

Tone Transfer and DDSP

- Differentiable Digital Signal Processing [DDSP] (Engel et. al, 2020)
  - Implemented a Differentiable version of Harmonic Plus Noise Model
  - Combined an Additive Synthesizer with a Filtered-Noise Synthesizer
  - Used 2 Datasets: NSynth and Solo Violin Performances
  - Has 2 Outputs: Harmonic Audio and Filtered Noise
- Tone Transfer (Caspe, McPherson & Sandler, 2023)
  - Utilizes the DDSP decoder to create a continuously controllable synthesizer
  - Includes interpretable parameters that allows
  - interventions for better results.

ACSML

Metrics (Choices)

- Human Evaluation (tentative)
  - Mean-opinion score (Défossez, 2018)
- Frechet Inception Distance (FID) (Kilgour et. al, 2019)
- Audio similarity methods (Python audio-similarity package)
  - e.g. Chroma Similarity, Energy Envelope Similarity, Spectral Contrast Similarity

ACSML

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

Chapter 3: Methodology

Very rough methodology

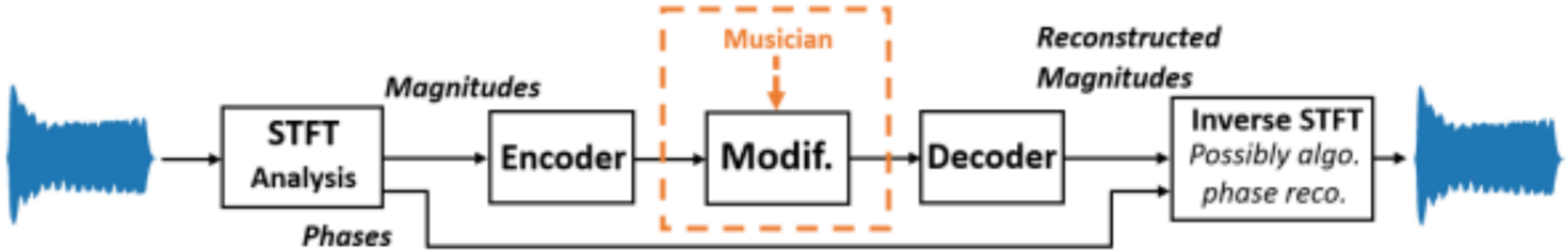

Figure 3.1: Global diagram of the analysis-transformation-synthesis process.

ACSML

Methodology (crying)

Mas maayos na methodology

1. Dataset ← **NSynth** → **Training Tuples**
    a. F0 ← Fundamental frequency using CREPE
    b. A-weighted Log Amplitude Loudness
    c. Getting “pseudo-envelope” through Hilbert transform (di ko pa sure to)
    d. Z-encoder: using CQT ← constant q transofmr

ACSML

Methodology (crying)

2. Data Processing
    a. multi-resolution STFT (MRSTFT) loss ← Guiding model learning
    b. GAN to refine pitch / F0 for micro-pitch variations???????
    c. Essentially same as DDSP

ACSML

Methodology (crying)

3. Metrics
    a. Subjective:
        i. MOS (Mean Opinion Score) ← Rating 1-5
        ii. ABX Testing
    b. Quantitative:
        i. Frechet Audio Distance
        ii. Ung nasa taas

ACSML

ATENEO COMPUTATIONAL
SOUND AND MUSIC
LABORATORY

THANK

you.