Automating music stimuli creation and analyses: A music synthesis algorithm for producing ground truth data

BEAT Lab

Maya B. Flannery[1], Lauren Fink[1]

[1] Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada

McMaster University

Institute for Music & The Mind

Introduction

We address two issues with modern-day music listening and psychological research.

Spotify
TIDAL
Apple Music

MusicGen
Jukebox
Soundraw

- Listeners are faced with **overwhelming choice** of musical content.
- Psychologists need objective, easy-to-use, **tools** to help understand music listening behaviour.

Past research. In both industry and academia, music is given labels (e.g., descriptors such as happy, fast, mellow, etc.) by experts in the field, which help us organize music.

Limitations. Manually labelling and transcribing music is time consuming—this process cannot scale to today's music collections—and is susceptible to subjective biases (Aucouturier and Pachet 2003).

Current research. There is promising development of computer software (i.e., Music Information Retrieval; MIR) that can automatically label music (Dong 2018). However:

1. Current ground-truth datasets, used to train models to predict music labels, require manual labelling (by experts) and are thus still potentially biased.
2. There has been little work investigating the validity of music labels
3. There is little known about *how to create* music with specific labels.

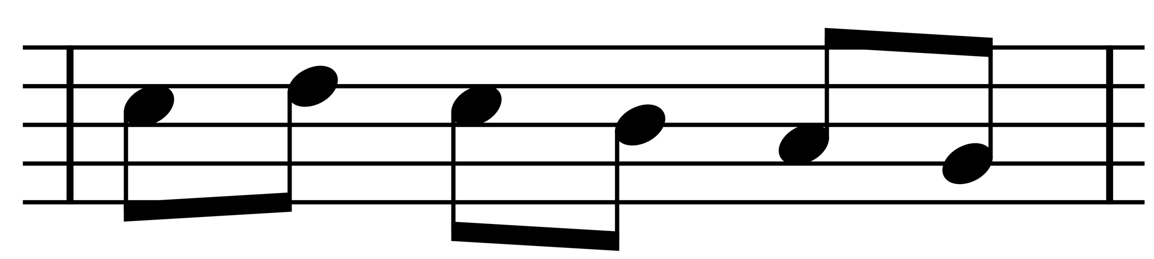
Objectives

The present research aims to **improve the way we use music labels** by defining compositional and performance parameters to automatically generate music datasets with corresponding labels (eliminating biased ground-truth datasets).

Additionally, we aim to use such generated datasets to train computer models to predict labels, and be validated (verified) by human listeners.

Methods

1. Seed Creation



A root note sequence must be specified:

- In International Pitch Notation (IPN)
  - e.g., [C5, D5, C5, ...]
- or Music Instrument Digital Interface (MIDI) notation
  - e.g., [72, 74, 72, ...]

2. Musical Parameter Definition

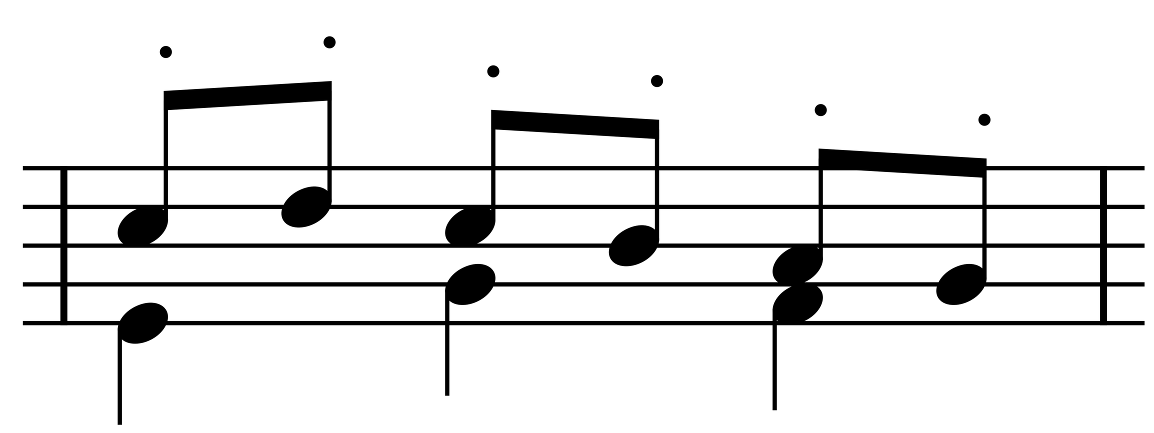
Many music features (e.g., register, dynamics, timbre, voices, note density, etc.) can be specified. They must be continuously manipulatable between two values, for example:

- Articulation: percentage of note duration
  - from 25% to 100%
- Tempo: number of beats per minute
  - from 40bpm to 200bpm

3. Permutation and Combination

Parameters from **Step 2** are randomly sampled.

They are combined to form musical excerpts (as .midi files) based on the sequence in **Step 1**.



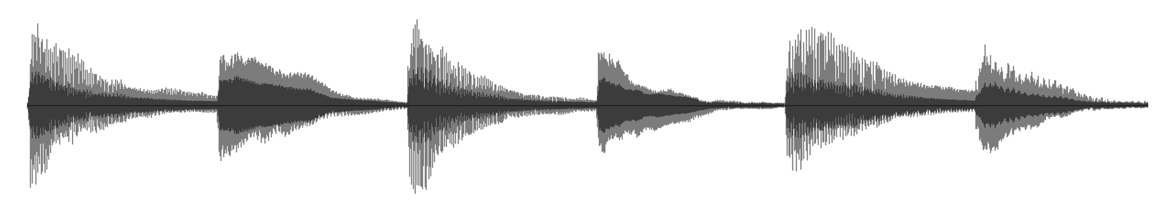
This process allows for quick generation of thousands of musical excerpts that represent a diverse range of music.

MIDI Database

4. Synthesis

MIDI excerpts are rendered as audio files with FluidSynth

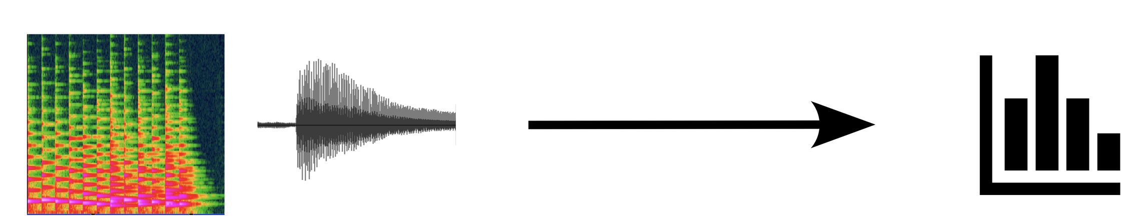
- These renderings use timbre-related parameters sampled from **Step 2**.



AUDIO Database

5. Feature Extraction

MIR tools such as Essentia, Librosa, and MIRtoolbox extract features from digital audio.



6. Identification

Models are trained and tested on features from **Step 6** to predict input parameters.

7. Verification

Human participants listen to excerpts and provide ratings for features.

- Ratings are compared to predictions in **Step 7** and musical parameters in **Step 2**.

We learn from this feedback which musical features are optimal for the generation of future datasets.

Conclusions

This algorithm systematically creates large well-labelled datasets of both symbolic and audio music.

- They can be used to verify accuracy of current MIR tools.
- They also provide ground-truth data for machine learning models.

Human feedback is used to validate the algorithms' input assumptions (i.e., used in Steps 1–4).

- This process can confirm that listeners perceive changes in stimuli as expected.

References

Aucouturier, Jean-Julien, and François Pachet. 2003. "Representing Musical Genre: A State of the Art." *Journal of New Music Research* 32 (1): 83–93. https://doi.org/10.1076/jnmr.32.1.83.16801.

Dong, Mingwen. 2018. "Convolutional Neural Network Achieves Human-Level Accuracy in Music Genre Classification." arXiv. http://arxiv.org/abs/1802.09697.